University of Central Florida
STARS

Electronic Theses and Dissertations, 2020-

2022

An Exploratory Assessment Of Small Group Performance Leveraging Motion Dynamics With Optical Flow

Joshua DeSantiago
University of Central Florida

Part of the Operational Research Commons
Find similar works at: https://stars.library.ucf.edu/etd2020
University of Central Florida Libraries http://library.ucf.edu

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2020- by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

DeSantiago, Joshua, "An Exploratory Assessment Of Small Group Performance Leveraging Motion Dynamics With Optical Flow" (2022). *Electronic Theses and Dissertations, 2020-*. 1192.
https://stars.library.ucf.edu/etd2020/1192

University of Central Florida
STARS
Showcase of Text, Archives, Research & Scholarship

AN EXPLORATORY ASSESSMENT OF SMALL GROUP PERFORMANCE LEVERAGING MOTION DYNAMICS WITH OPTICAL FLOW

by

JOSHUA DESANTIAGO
B.S. University of Central Florida, 2020

A thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science
in the School of Modeling, Simulation, and Training
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, FL

Summer Term
2022

© 2022 Joshua DeSantiago

ABSTRACT

Understanding team behaviors and dynamics are important to better understand and foster better teamwork. The goal of this master's thesis was to contribute to understanding and assessing teamwork in small group research, by analyzing motion dynamics and team performance with non-contact sensing and computational assessment. This thesis's goal is to conduct an exploratory analysis of motion dynamics on teamwork data to understand current limitations in data gathering approaches and provide a methodology to automatically categorize, label, and code team metrics from multi-modal data. We created a coding schema that analyzed different teamwork datasets. We then produced a taxonomy of the metrics from the literature that classify teamwork behaviors and performance. These metrics were grouped on whether they measured communication dynamics or movement dynamics. The review showed movement dynamics in small group research is a potential area to apply more robust computational sensing and detection approaches.

To enhance and demonstrate the importance of motion dynamics, we analyzed video and transcript data on a publicly available multi-modal dataset. We determined areas for future study where movement dynamics are potentially correlated to team behaviors and performance. We processed the video data into movement dynamic time series data using an optical flow approach to track and measure motion from the data. Audio data was measured by speaking turns, words used, and keywords used, which were defined as our communication dynamics. Our exploratory analysis demonstrated a correlation between the group performance score using communication dynamics metrics, along with movement dynamics metrics. This assessment provided insights for sensing data capture strategies and computational analysis for future small group research studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Effective teams work together to solve some of society's most difficult problems through collaborative problem solving. Throughout history, *wicked problems* pose a challenge that a single person cannot solve alone due to the scale and complexity of the subject. The objectives are difficult and often contradictory in nature. Teams answer this challenge by steadily becoming more collaborative. Therefore it is important to better understand how to understand and measure team effectiveness and dynamics through ***Team Research***. Teams and groups of individuals make those impossible problems possible by having multiple minds working together to collaboratively solve a singular issue. With these societal challenges becoming more frequent, increasing in size, and growing in complexity, the need for individuals to work together is at an all-time high. Responding to this challenge, ***Team Research*** is an established field to study both models and methodology designed to empirically understand collaborative team-based research in the most unobtrusive way possible.

The use of teams has expanded to many problems and domains. For example, Figure 1.1(a) shows teamwork in education where students work collaboratively to tackle tougher problems in the classroom by cooperating with each other. This early introduction to teamwork prepares those students for future endeavors. In Figure 1.1(b) a corporate team combines multiple skills and abilities together to respond to a new competitive challenge. Teamwork has had a major effect on many workplaces, improving these companies' development cycle. Finally, in Figure 1.1(c) a cyber security incident response team respond to a cyber incident to determine threats and vulnerabilities in a particular piece of software. Cyber teams rapidly mobilize to effectively and efficiently respond to growing threats facing our nation. As the need for teams grows, so does the need to measure how effective these teams are in a variety of settings.

Figure 1.1: This figure shows three different example of teamwork in diverse domain areas. Figure 1.1(a)[84] demonstrates teamwork in a college background, showing teamwork's importance in educational situations. Figure 1.1(b)[12] demonstrates teamwork in a professional background, showing teamwork's importance in everyday issues.Figure 1.1(c)[75] demonstrates teamwork in a critical thinking background, where teamwork is necessary to tackle important issues.

Teamwork is an essential 21st-century skill, which is applied on different scales in the home, workforce, and the community. Many problems we face require team collaborative problem solving processes to understand team knowledge building processes, in conjunction with team members' individual knowledge building processes and internalized domain specific knowledge [25, 45]. The need to understand collaborative problem solving continues to grow an even more pressing problem in society today, where the complexity of socio-technological systems across the industry, the military, and academia is ever-increasing [27], but team problem solving processes is a skill not taught and often overlooked [17]. As a result, it becomes that much more important to better understand the definition of teamwork and effectively measure it unobtrusively, as it is crucial in terms of creating success in solving complex *wicked* problems [6, 24]. To fully develop a model of teamwork, the foundation of the word 'team' must be understood. A *team* is a collection of individuals who come together to solve a problem, to satisfy their own goals or the goal of an organization that individual is a part of. At a minimum, a team should be a cooperative functioning unit and, at its best, a team is a collaborative unit [59] functioning at a high level. Our goal is to develop an approach that can start to assess these team dynamics through contactless sensors that evaluate motion and communication.

Not only is understanding teamwork important, but it is also crucial to identify the fundamental characteristics of effective team performance and group dynamics. Researchers have studied methodological approaches to understand and classify team performance [49]. A strong and effective performing team can be defined as one that *communicates*, *coordinates*, and *collaborates* well, respects teammates' opinions, develops trust in each other, and can coordinate and cover for each other's weaknesses [85]. Teams containing a social loafer, an uncooperative unskilled member, a lack of communication and shared mental models, or a counterproductive alliance could lead to weak and ineffective teams[27]. These factors threaten the success of a strong team and successful group dynamics.

Past approaches have focused on understanding the group dynamics of a team to understand the team's ultimate behavior and performance. This can often be achieved through quantitative surveys [68] to measure the dynamics and performance of the team. The questionnaires measure team effectiveness through a variety of constructs which include understanding the team processes, team relationships and interactions, roles and responsibilities, goals, and problem solving abilities [77]. These survey instruments are often captured both during and after a team interaction session. Quantitative survey questionnaires suffer from two shortcomings, first interrupting the experience to conduct the measurements can often create systematic biases, and second introspective reflection on higher order cognitive processes can also potentially introduce systematic biases in the collected data [52]. This thesis studies potential objective computational sensing methods to automatically understand teamwork measures.

Recent research has attempted to collect qualitative data to measure the group dynamics of teams. Often these approaches analyze the communication dynamics of audio recordings and transcripts of the teamwork sessions, or visual recordings to investigate the motion dynamics.

Some datasets provide audio recording, transcripts of the recording, and visual recordings of non-interrupted teamwork posted publicly for other researchers to examine [8, 15, 63, 81]. While I think you can extract objective metrics, qualitative data normally involves much more subjectivity than much of the quantitative work. These metrics provide numerical measures of whether these teams' behavior and dynamics are effective or ineffective. Figure 1.2 presents an example of positive group dynamics where the group dynamics demonstrate effective collaboration through their motion dynamics. Some of these can be seen in Figure 1.2 are *Looking while Listening* (LWL), *Head Pose Angle*, *Attention Center* (ATC), and *Fraction of Mutual Gaze*.

Figure 1.2: The image above gives an example of some team metrics mentioned in this thesis. This image visualizes metrics such as *Attention Center*, *Looking While Listening*, and *Head Pose Angle*. Those are only a small number of potential metrics that can occur during collaborative problem solving. [58]

These metrics must follow certain criteria to be properly classified and labeled, for example, an LWL is the number of frames that a participant, i, spends focusing on another individual while i is not speaking [64]. Using this metric, it can be determined whether a person is fully participating in the team activity. These metrics provide a measurement that can score team behaviors and performance to determine whether a team portrays strong or weak attributes. This is one of many metrics that can be extracted from these datasets to further study the quality of teamwork and to help find problematic behaviors that can be fixed to stop further issues that arise in teamwork development. However, there is a need to computationally and automatically classify, code, and label these metrics. Having this task done computationally aid humans so they do not have to hand code events. This means metrics can scale to larger datasets and feedback can be provided in real-time to the team.

Information gathered from the data collected from teams and their respective teamwork behaviors and dynamics is integral for different fields since teams as used to solve many different complex problems. There are many approaches to team collaboration, including multidisciplinary teams (teams that contain individuals from different scientific fields that collaborate on a topic), interdisciplinary (results and expertise from two or more scientific fields are combined), or transdisciplinary (disciplinary boundaries are crossed to create a holistic approach) are expected to hold the key to success [10]. Therefore it is important to develop an approach to automatically assess teams to understand how to help advance collaboration, problem solving, and science in general.

This thesis's goal is to conduct an exploratory analysis of motion dynamics on a publicly available teamwork dataset to understand current limitations in data gathering approaches and provide a methodology to automatically categorize label, and code team metrics from multimodal data. I examined *The Unobtrusive Group Interaction (UGI) Corpus* [8] in this thesis. The technology used to collect multi-modal data includes video data from *Microsoft Kinects* and audio data from high-end microphones placed in the room. I will first attempt to understand how the shortcomings of data capture affect the usage of video data to understand the motion dynamics in our study of teamwork performance. With this information, I will then formulate a computational methodology to process the motion flow from this dataset and derive teamwork metrics. Lastly, I discuss how to correct data capturing shortcomings and provide more robust approaches for future human subject research studying teams.

Main Research Question: Do Motion Dynamics and Communication Dynamics impact the team's collaborative problem solving effectiveness?

Therefore, this thesis aims to advance the following sub-research questions (RQ):

- **RQ1:** What multi-modal approaches did previous literature use to understand *Communication Dynamics* and *Motion Dynamics* of the team's general *Group Dynamics*?
- **RQ2:** How are *Motion Dynamic* metrics automatically derived from video data?
- **RQ3:** Does the quantity of motion impact the team's collaborative problem solving effectiveness?
- **RQ4:** Does the amount of *Communication Dynamics* affect the *Motion Dynamics*?

1.1 Summary of Contributions

The work in this Master's thesis builds upon the motion dynamics in Team Research and ways to computationally measure metrics for the *Macrocognition in Teams Model* (MITM) [25, 45]. The major contributions are outlined below

- The creation of a dataset coding scheme. This schema included an assortment of all papers researched while investigating examples of different corpora. (Chapter 2).
- Categorizing the information found in research to include sections such as dataset, year,

author, title, publication, metrics, data channels, thesis, theory confirmation, and notes for any important or relevant information discovered. (Chapter 3).

- Identification of problems with current capture methods for the video data. (Chapter 3).
- A new method to extract the motion dynamics from a small team dataset. I propose using an *Optical Flow* which uses data to track differences between an image sequence to determine the sequence's motion. This was accomplished by taking the videos shared from the UGI corpus [8] and analyzing them for every participant in the video. From there, each participant will go through the *Lucas-Kanade Optical Flow* [56] algorithm to extract the amount of motion in the frames. (Chapters 4 and 5).

1.2 Organization of the Thesis

The thesis is divided into seven chapters. In Chapter 2, we introduce the fundamentals of small group teaming, their processes, and their measures. Chapter 3 examines four major research datasets that attempt to measure these small group dynamics. This leads to the identified research gap in the literature regarding how to analyze motion dynamics from video data captured in these datasets. Based on this foundation, Chapter 4 develops our methodology to measure quantitative and qualitative metrics, communication dynamics, and motion dynamics from multi-modal capture data. In Chapter 5, I describe in detail our findings from the study. Chapter 6 discusses how our findings could impact small group dynamics and the limitation encountered with the chosen dataset. In Chapter 7, I summarize our contribution and discuss any future work.

CHAPTER 2: BACKGROUND

The goal of *Team Research* is to understand team-based collaborative work. This chapter looks at how different researchers used multi-modal data to capture communication dynamics and motion dynamics of team behaviors and performance by different sensing channels. The data extracted is analyzed and separated into different teamwork metrics. The different metrics include measurements found in other studies, which are used to quantify different situations that take place in the teamwork data collected. These metrics are further grouped and evaluated for similarities and usage. The goal is to bring to light and improve any flaws that would hinder the proper collection of teamwork data in future human subject research.

I first created a list containing different research papers that had used teamwork datasets in their respective works to extract these metrics. That literature dataset contained basic information about each paper such as their title, year, author, publication, name of used teamwork dataset, and paper's thesis/research question(s). In addition, each paper was read to locate which metrics were used to assist with the paper's analysis. These metrics were classified depending on which data channel they examined and reported with as well as different groups that combined closely related metrics. These classifications and groupings can be seen in Table 2.1.

These metrics were necessary to understand how current teamwork data was being used in the literature. Our understanding of the current usage of these metrics will provide insight into different methods of analyzing teamwork data. From these metrics, I can also learn which section of the teamwork data could need future work by checking how much the metrics were found and how those metrics were used.

Table 2.1: This table shows summaries for both the audio and video metrics found during a literature search. In the Table, you can see how each metric was generally grouped as well as the search terms used to locate the metrics above.

Audio		
Search Term Group	**Search Terms**	**Results**
Speech Terms	Dialog Act	*Dialog Acts*, *Speaking Turns*, *Utterance*
	Speaking Turns	*Speaking Turn Duration*, *Average Turn Duration*
	Utterance	*Total Number of Turns*, *Start*, *End*
	Speaking Length	*Speaking Speed*, *Total of Speaking Length*
Sound Terms	Prosodic Features	*Prosodic Features*, *Pitch*, *Intensity*
	Interaction	*Energy Variation*, *Energy Spectral Flatness*
		Total Speaking Energy, *Overlap*
		Speaking Interruption, *Successful Interruption*
		Unsuccessful Interruption, *Backchanneling*
		Total Successful Interruption

Video		
Search Term Group	**Search Terms**	**Results**
Movement Terms	Body Movement	*Body Activity Length*, *Body Activity Turn*
	Head Movement	*Average Body Activity Turn Duration*
		Head Motion Length, *Head Motion Turn*
		Average Head Activity Turn Duration
Gaze Terms	Visual Focus of Attention	*Fraction of People Gaze*, *Fraction of Mutual Gaze*
	Gaze based on Speaking Turns	*Fraction of People Gaze*, *Fraction of Shared Gaze*
		Fraction of Convergent Gaze, *Gaze Skew*
		Looking While Speaking, *Looking While Listening*
		Being Looked at While Speaking
		Center of Attention While Speaking

2.1 Audio

The metrics are grouped into two major groupings (audio and video), with the first being audio. Metrics placed under the audio grouping include any metric that contains sound or noise collected by recording technology. From this grouping, the metrics are further separated into two subcategories. The speech group addresses the verbal portion of the sounds extracted from the data, and the sound group addresses the non-verbal sound that occurs throughout the meetings.

2.1.1 Speech

2.1.1.1 Dialog Act

Dialog Acts attempts to understand speaker's intentions. Dialog Acts are broken up into five major groups describing the speaker's intention:

- **Information** which is providing information;
- **Action** which is to give an offer or make a suggestion;
- **Commenting on the discussion** making comments regarding the speaker's intentions;
- **Segmentation** breaks up the conversation without adding to the content;
- **Other** is meant for utterances that do not fit into the other categories [29].

The studies that are marked with using the dialog act metric utilize all five of the groups in their studies. Dialog Acts were seen to impact many studies that dealt with recognition as the study's main focus. The most common recognition was speech-based, however some of this recognition went beyond speech where they also tracked the speaker's personality traits, social roles, or speech roles [29, 67, 76, 78]. Other studies used dialog acts to establish human-to-human interactions and the effect they could have on the small group dynamic [2, 35, 53].

2.1.1.2 Speaker Turns

Speaking turns is the number of turns counted throughout the meeting for each individual.

Speaking Turn Duration is the average turn duration for each individual throughout the meeting [66].

Average Turn Duration is the speaking duration for each individual divided by the individual's number of turns to discover the average length of turns in the session.

Total Number of Turns is the total number of speaking turns an individual spoke throughout the meeting [39].

These turn-taking metrics have been used to assist with studies dealing with behavioral patterns and personality trait factors affecting group performance [31, 39]. As well as other metrics under the speech data channel, they were used for research in speech recognition[21, 33, 79]. They were also applied to further their work in predicting those same personality traits and behavioral patterns that could change the outcome of a group performance score [3, 5, 22, 42]; specifically, studies predicting leadership and dominance were shown using these metrics [32, 63, 64, 66].

2.1.1.3 Utterance

Utterance is a spoken word, statement, or vocal sound. Utterances are used differently in different studies, in one study the paper separated utterances into two different types of utterances.

Subjective utterances which is a span of words where a mental or emotional state is being expressed either through the choice of words or prosody,

Objective Polar utterance which are statements or phrases that describe positive or negative factual information about something without conveying emotion.

These various versions of utterances were used in the paper to assist with recognizing participants' personal feelings being shown [60]. Other types of this metric can be seen in works dealing with prediction. In the prediction of dominance, the metric used was:

Total Speaker Turns without Short Utterance is the total number of turns the participant spoke for longer than one second [32].

Filled Pauses, which are specific utterances that use the words ‘uh’ and ‘um’, and basic utterances were also used to predict group performance [40, 51].

2.1.1.4 Speaking Length

Speaking Length is the total speaking time for each individual [64].

Start when the participant starts speaking.

End when the participant stops speaking and another participant starts.

Speaking speed is the quotient of an utterance by the length of the utterance. Then is averaged with all the utterances.

Total of speaking length is the total length of utterances observed in a session [53].

Studies that have these metrics lean towards different categories of predictions. The majority of categories are used to predict leadership and dominance in teams that could be seen in this type of group dynamic [32, 63, 64, 66, 22]. Other studies use the metric to predict the overall group’s performance in the given task [51]. These metrics have also been seen in studies to help recognition methods in speech recognition and social role recognition [82, 86].

2.1.2 Sound

2.1.2.1 Prosodic Features

Prosodic Features are the linguistic features found in voiced speech. [74]. The prosodic feature is a generic term that branches into different metrics, here are some of the metrics that were directly referenced in different studies:

Pitch is the tone of sound found from the participant.

Intensity is the captured loudness of the voice [53].

Energy Variation is the loudness perceived by the ear.

Energy Spectral Flatness is the measure that is used to determine whether the speaker's words are voiced or slurd [63].

Total Speaking Energy is the amount of speaking energy counted throughout the session [32].

Prosodic features are the most frequently used metric in all of the studies researched in this study. Many of the studies used this feature for improvement in their recognition studies or to continue advancement in their prediction models. The recognition studies varied from speech [28, 60, 67], social role [79, 82], or personality trait [78, 37]. The prediction studies seem to see more work with prediction personality traits [3, 22, 55], as well as some work in leadership [63], dominance [32], and group performance [51].

2.1.2.2 Interaction

Overlap is the number of frames when more than two or three participants talk at the same time.

Speaking interruption is when a participant interrupts the speaker [64].

Successful interruption is when a participant interrupts the speaker and the speaker finishes speaking before the participant finishes.

Unsuccessful interruption is when a participant interrupts the speaker and the participant finishes speaking before the speaker finishes.

Backchanneling is the total number of times that a participant speaks for less than 2 seconds while the speaker is talking [5].

Total Successful Interruption is the cumulative number of times a successful interruption occurs.

Since these metrics dealt with the interaction between participants, some studies were seen using the interaction metrics for research that were checking what factors heavily influenced the outcome of the small groups' performance score [4, 31]. The interaction metrics, much like many of the metrics in the audio data channel, saw work in studies dealing with recognition [79, 80] and prediction, because the studies that use the interaction metrics predict concepts that could be affected by the group. The concepts found while researching were group satisfaction [42], group performance [5, 40], dominance [32], and leadership [63, 64].

2.2 Video

Metrics are also categorized under the video category. Video metrics contain visual images collected by recording technology. From this grouping, the metrics are further separated into two subcategories. The first is the movement, which involves the changes in position the participant does with their body. The second grouping is gaze, which is the group with movement with respect to eye contact and attention to other participants.

2.2.1 Movement

2.2.1.1 Body movement

Body activity length is the total time during which the target participant moves.

Body activity turn is the number of turns the target participant's body goes from a 'moving' state to a 'not moving' state [22].

Average body activity turn duration is the average amount of time for a participant to stay in a moving state [63].

From the information gathered, it was shown that body movement metrics have been used mostly for studies focused on human interaction [5, 20, **?**, 73]. Other notable small group studies that body movement metrics have been used in include the improvement of computer vision practices[7], analyzing personality traits, and inferring impression scores[54, 63].

2.2.1.2 Head Movement

Head motion length is the number of frames during which the target participant's head moves.

Head motion turn is the number of times the target participant's head goes from a 'moving' state to a 'not moving' state [22].

Average head activity turn duration is the average amount of time for a participant's head to stay in a moving state [63].

These metrics were used in a number of ways, with popularity arising in its ability to understand and analyze social signals from group interactions [7, 22, 54, 69]. Another is to analyze the communication skills of individuals in small groups by measuring amounts of movement while someone is speaking [53, 63]. The metric could also be used to assist with the development of models that predict the performance of small group meetings [5, 51].

2.2.2 Gaze

2.2.2.1 Visual Focus of Attention

Fraction of People Gaze is the part of the interaction between two participants where we measure whether the participants are looking at each other or somewhere else.

Fraction of Convergent Gaze is the part of the interaction where we measure a participant being the center of attention.

Fraction of Mutual Gaze measures the fraction of gaze data in which two participants look at each other.

Fraction of Shared Gaze is the part of the interaction where two participants look at the same participant.

Gaze Skew measures the amount of time participants look at one another to see if participants are looking at each other equally or is one participant being focused on more.

These metrics were extracted from the paper "Predicting the Performance in Decision-Making Tasks: From Individual Cues to Group Interaction". Avci et al. [5] were addressing the problem of predicting the performance of decision-making groups. To assist in remedying this issue, the use of the above five group-level looking cues were used.

2.2.2.2 Gaze based on Speaking Turns

These metrics below were used to provide information about dominance and power when researching emergent leaders [64]:

Looking while speaking (LWS) is the amount of frames that a participant, i, spends focusing on another individual while i is speaking.

Looking while listening (LWL) is the amount of frames that a participant, i, spends focusing on another individual while i is not speaking.

Being looked at while speaking (BLWS) is the amount of frames that a participant, i, spends receiving attention from the other individuals while i is speaking.

Center of attention while speaking (CAWS) is the amount of frames that a participant, i, spends the focus of all other individual's attention while i is speaking.

CHAPTER 3: LITERATURE REVIEW

In this chapter, I reviewed the literature on four different publicly accessible teamwork multi-modal teamwork corpora: *AMI Meeting Corpus*, *ELEA Corpus*, *UGI Corpus*, and *GAPS Corpus*. I provide detailed breakdowns of the data corpus creation, technology used, channels captured, and what data they provided from their dataset. I also look at different metrics which were extracted from the different corpora that were classified and coded into our schema from Chapter 2. Finally, I discuss these corpora limitations and shortcomings I felt needed to be addressed in creating future publicly available multi-modal teamwork datasets.

The UGI corpus and GAPS corpus use are extremely limited but were chosen to detail the technological methodology. There were only a handful of papers found on these datasets. Their limited usage can be attributed to their newer releases, as AMI and ELEA were released to the academic world ten years before these corpora were created. This temporal gap explains the drastic difference in literature found. Despite their limited usage, these corpora have value by showing their new techniques and technology used to improve contact-less team data collection.

3.1 AMI Meeting Corpus

The *AMI Meeting Corpus* is a synchronized multi-modal dataset that contains data of 100 hours of meetings. This dataset was captured by 15-member multi-disciplinary researchers who wanted to understand how groups interact. The goal of the data capture was the development of a meeting dataset that could improve team behavior and effectiveness by providing a public resource for researchers to understand teamwork processes on the web [15]. The *AMI Meeting Corpus* provides recordings of participants engaging in natural conversation and playing fictitious roles participating in scenario-based meetings and non-scenario-based meetings. Participants could be assigned various roles in the meeting, such as the meeting are industrial designer, interface designer, marketing, or project manager. With these roles assigned, the participant must work on the given problem. The team takes on a project from scratch and completes it over a certain amount of time. The layout for each day starts with training and giving role assignments to each participant, followed by four meetings to discuss the project. Throughout the task, the participants will have different documents that will allow them to prepare for the day. All work completed through this time is recorded by the participants through web pages, email, text process, and slide presentations. The dataset consists of meeting recordings (audio and video) and transcription and annotation of all the meetings [15].

The *AMI Meeting Corpus* contains several unique attributes that have inspired a large amount of follow on research using the dataset from researchers across the globe. The first main feature is the corpus is licensed under a *Creative Commons Attribution ShareAlike License*. This license allows data to be freely shared and publicly distributed allowing researchers the ability to copy and use the data for non-commercial purposes. This was one of the first large-scale multimodal teamwork datasets to capture such robust data of teams. Another feature includes all the annotations provided are in the same format that represents how they relate structurally to the transcript, as well as to other annotations. For instance, two different types of annotation, topic segments, and dialogue acts are represented in the data. These annotation types are labeled with a timestamp attached as well as timed sequences of words. These summaries are divided into segments by not only time but by dialog acts. This structure was built and searched using an open-source *NITE XML Toolkit*. The *NITE XML Toolkit* contains tools and libraries with data containing annotations with complex structures equivalent to annotations found in the *AMI Meeting Corpus* [36]. This facet complements the license which permits the AMI corpus to not only be free to use, but easy to navigate and obtain data. The last feature this corpus possesses which makes this corpus unique is the experimental design used in the meeting structures. For most corpora, the main objective of recording the participants is to capture their natural, unrehearsed conversation during the meeting. AMI captured a varied series of meetings.

Figure 3.1: Example of the different meeting rooms used by researchers in the *AMI Meeting Corpus* capture dataset to capture data from different teamwork activities in varied meeting types [15].

The *AMI Meeting Corpus* was created in three instrumented meeting rooms as shown in Figure 3.1 that capture multi-modal data. The AMI corpus specifies that each meeting room was set up with microphones that record from both close and far away, a camera that records each individual as well as the room as a whole, a slide projector, and an electronic whiteboard [14]. The video was recorded using six cameras, four *Sony XC555* subminiature cameras which provide a close view of the participants, and two *Sony SSC-DC58AP* CCTV cameras which provide an angle to view all the participants in the room [14]. Six *Sony GV-D1000E* digital video recorders are used to record the video. The video data captured can be viewed regularly or in a higher resolution for video processing research if needed [14]. There were 24 microphones and cameras split between "16 *Sennheiser MK2E-P-C* miniature omnidirectional microphones placed between participants and at the ends of the tables, and 8 *Sennheiser EW300* Series radio microphones to record each of the participants" [14]. Along with the microphones, each participant wore a close-talking headsets condenser mic and a lapel mic [14].

Since the *AMI Meeting Corpus* is a speech-driven corpus, all the audio recordings of the individual speakers were turned into high quality, manually produced orthographic transcriptions. A look at the transcription can be seen in Figure 3.2 where all sound, whether verbal or not, is taken into account. The use of a speech recognizer was added to include word-level timing in the transcript, which tracks the amount of time each word takes. In addition to the transcript, the corpus also contained a wide range of other types of annotations not limited to solely the linguistic variety [13]. The audio annotations available in the dataset are; *dialogue acts* which is the act of classifying an utterance concerning the current dialogue; *topic segmentation* refers to how sentences are grouped in a conversation depending on the topic; *extractive summaries* which extract important information from the corpus and summarizes that information; *abstractive summaries* which is the summary of the corpus using important key points; and *named entities* as the proper name of real-world objects such as the name of the participants [13].

There is some focus on motion dynamics in this dataset. The AMI corpus team also manually annotated movement and behavioral data found in the meeting videos, which included different types of head gestures, hand gestures, gaze direction, movement around the room, emotional state, the focus of attention, and where the head is located in the video frame [13].

During the process of collecting and organizing the different papers found in our AMI literature search, I came across a pattern emerging from the data collected. In this section, I recorded the corpora's data channel in each paper and discovered two data channels that stood above the rest. Those data channels were the speech and sound.

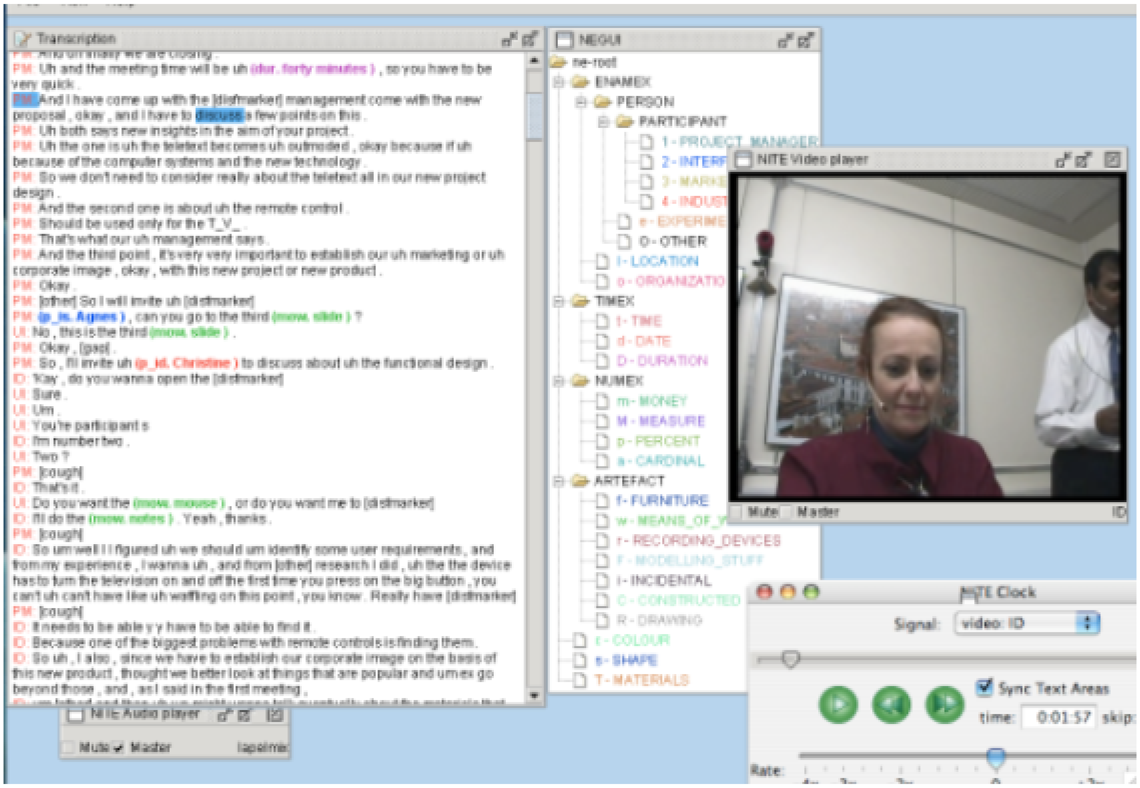


Figure 3.2: This figure shows an example of the type of video and audio provided by the AMI Meeting Corpus. Also seen is an example of the transcripts data that is used in many different research studies [15]

.

3.1.1 Speech

A large portion of the literature found using the *AMI Meeting Corpus* used speech as one of their main data channels. Our research noted that the papers that used speech as a data channel from the dataset used the transcriptions to formulate metrics for team behavior and dynamics. An example of this is demonstrated in Tur et al. [76] where the authors present the *Cognitive Assistant that Learns and Organizes* (CALO) meeting assistant which is a system used for speech recognition and understanding in a group setting. This paper presents the development of an automatic speech recognition system with the capabilities of understanding and extracting important information from automated transcripts. A system such as this requires a substantial amount of data to both train and test the accuracy [76]. The *AMI Meeting Corpus*, with 100 meeting hours and a small-group conference meeting format, was the ideal candidate for this study [15].

CALO-MA encountered an issue of uncertainty in the data, which was with the usage of the word "you". The word "you" with text alone is ambiguous because without proper context the word can refer to several different entities in the room. To tackle this issue, CALO-MA teams used 1000 utterances of the word "you" from the corpus as test data to support solutions that were created. The *AMI Meeting Corpora* has *Focus of Attention* annotations that can assist with uncovering who the "you" is meant for in the meeting context [76]. The dataset often sees usage beyond their annotation and transcriptions. Lai and Murray [42] present a method of predicting group satisfaction automatically using non-verbal acoustic features. This information is taken from the post-meeting questionnaire, where participants are asked to subjectively rate various aspects of the meetings. These ratings and features described by Lai and Murray [42] helped create an algorithm to predict the satisfaction of groups [42].

3.1.2 Sound

Sound is another common data channel often used by researchers studying teams, where the noises the participant makes during the meeting recording are used to obtain metrics. AMI is a unique dataset since the sound data is augmented with annotations of sound data to provide more robust information to researchers [13]. Annotated sound data is used in several ways and an example can be shown in Leamy et al.'s study "Re-annotation of cough events in the AMI corpus" [44]. They showed the usefulness of coughing sounds to develop new detection algorithms [44]. Leamy et al. demonstrated this by extracting the data from the *AMI Meeting Corpus* and explaining how it could be enhanced. The data received from the AMI, by Leamy et al. needed to be enhanced to better represent the cough events. For example, start and end times for cough events were missed and labeled by being the wrong length. Some cough events were labeled as too short because the smaller increment of time was not being used to represent the time allotted to the cough events. Similarly, some cough events were marked as too long because multiple coughs took place in succession and were grouped as a single event. These issues were fixed by using non-annotated audio data given by the AMI Meeting Corpus to assist with the re-annotation of cough events through the meeting events providing new data of 1369 cough events [44].

Figure 3.3: This figure shows the position of the surrounding cameras, on the left is camera positioned on the corners of the room looking down, and on the right is the view from the camera directly above the meeting table looking down [16].

AMI Meeting Corpus main contribution is a large publicly available speech corpus that has been pedantically transcribed and annotated. Besides the annotation, the rest of the raw data shared through the AMI corpus interface is sometimes problematic to process. For example, the video recordings given in the dataset are not of the highest resolution or quality, despite having an option for increasing the quality of the video. The video contains too much background movement, noise, and grain during the individual videos for any movement detection program to properly process the video for movement dynamics. This can be seen in Figure 3.4 where another participant is shown in the shot of the individual camera of a participant. The different from the camera also proved troublesome to process as captured. Figure 3.3 shows where the first image is presented at an angle to allow the view of the whole space, however, when processed, the bodies and movement of the participants are morphed because of the positioning of the camera's angle. Figure 3.3 does not provide any useful data due to both low resolution and the way the camera is positioned the only useful data available is upper-body tracking which can be skewed by certain movements where parts can disappear depending on the sitting position of the participant. This view also is grainy and noisy due to the low resolution. The low resolution is likely to save on storage space as the dataset is very large. Generally, a top down view is useful when properly combined with additional video angles to capture a full picture of the movement in the space. Additionally, data should be provided to help unwarp the fish-eye camera distortions from the lens.

Figure 3.4: Example of the single person camera view that was captured. [15].

3.2 ELEA Corpus

The *Emergent LEader Analysis Corpus* (ELEA) was gathered to study leadership that emerged from newly created teams [65]. The ELEA dataset includes annotations of perceived self-reports, two questionnaires, and a survival task activity [65]. The first questionnaire is taken before the task where participants answer questions to measure their personality and dominance. The survival task that the ELEA Corpus uses is the *Winter Survival Task* scenario [63], with participants first completing a ranking task individually and then performing the same ranking task jointly as a group attempting to produce a single ranked list from the group of all the items.

The *Winter Survival Task* conducted by the ELEA group is a group task that consisted of a plane crash scenario, in which participants need to rank 12 items ranked what was most and least important in their scenario [63]. Once they complete ranking the items individually, which takes approximately five minutes, they then rank the item as a group, which then takes an additional fifteen minutes [63]. After the survival task, the participants took a post-task questionnaire of 17 questions that captured how they perceived themselves as well as their teammates [63]. The purpose of these questions were to measure leadership found in the group, participant's power over the other participants, overall participants intelligence, likability of participants, and the ranking of dominance from the group [65]. The ELEA Corpus consists of 40 meetings, 27 meetings were recorded using a portable setup, and 10 meetings were recorded with a static setup. In 3 of the recorded meetings using the portable setup, the video recordings were discarded because of the meeting being unsuccessfully recorded. From the 40 meetings recorded there were 28 meetings that contained a four-person team and 12 meetings that contained a three-person team. Each team was a newly formed group [63].

The ELEA study took place around a table, with one to two people on either side of the table. The room where the meeting took place was equipped with audio and video recording devices. Figure 3.5 it is shown how the audio and video recording devices were placed on the table for each of the group meetings. For the audio recording, the *Microcone* (microphone array) was placed in the center of the table and was used to record the conversation that took place during the meeting [65]. The *Microcone* allows the segmentation of each speaker from the meetings audio recording to aid future data analysis [65]. The video recording was captured by two different setups, one had six cameras (four close-ups, two side-views, and one center-view) [65] and the one center-view can be seen in Figure 3.6. The other setup was portable with two *Logitech Webcam Pro 9000*, creating a scenario where they could capture the meeting in more realistic conditions moving away from the in-lab approach and towards a more natural in-field approach [63].

3.2.1 Sound

The ELEA sound data channel extracted the noise participants made during the meeting in several different research studies. Many papers using ELEA data as their main corpus use sound metrics in their work. Avci and Aran [5] address the problem of predicting the performance of decision-making groups. Their paper uses sound metrics such as interruptions, backchannels,

Figure 3.5: This figure shows an example of the different capture devices used to track both audio and video data for the ELEA corpus dataset [65].

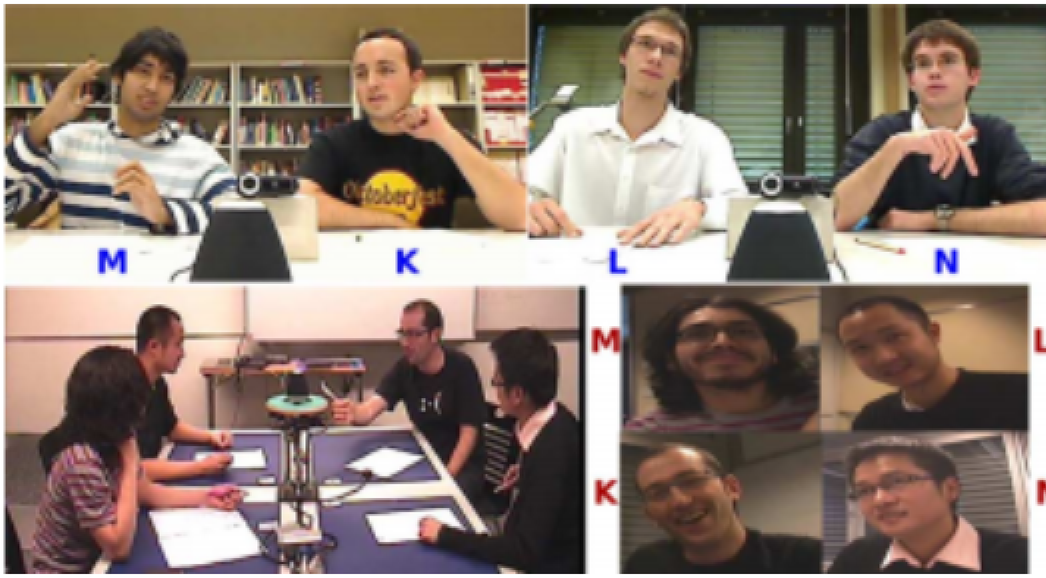


Figure 3.6: This figure shows the different camera views provided in the ELEA dataset [63].

group speaking cues, overlapped speech, and silence to assist with the analysis and prediction of non-verbal audio taken throughout the meeting interruption measures whether participants break the continuity of the speech coming from the participant that is currently speaking. The interruption is measured whether the interruption is successful or not, meaning if the interruption is successful the interrupter has taken the speaking role for the original speaker and vice-versa. Backchannel, in this case, is the total number of times that a secondary conversation takes place at the same time as the main conversation during the meeting. Finally, the speaking cues that took place throughout the whole meeting are also analyzed to classify the type and number of interruptions [5]. Avci and Aran [5] looked at successful and unsuccessful interruptions to understand how this affected group dynamics.

Other group speaking cues calculated were total speaking length and total speaking turns. Overlapped speech and silence were separated in fractions to only include the portion when no one was speaking or more than one person was speaking. Features used from these metrics were *Frac-*

tion of Silence (FoS), *Fraction of Non-overlapped Speech* (FoNS), *Fraction of Overlapped Speech* for 2 People (FoOS2), and *Fraction of Overlapped Speech* for 3 People (FoOS3) [5]. Similar to AMI, the papers that use the ELEA also frequently use *Prosodic Cue Extraction*. Sanchez-Cortes et al. [63] address the problem of inferring leadership from newly formed groups using non-verbal features from recorded data. Several non-verbal metrics were taken and used in the paper, many of those metrics being these prosodic features. Those metrics used, not including the prosodic features, are *Total Speaking Length*, *Total Speaking Turns*, *Average Speaking Turn Duration*, *Total Successful Interruption*, and *Speaking Turn Matrix*. The prosodic feature metrics used are *energy variation*, which is the loudness perceived by the ear, and *pitch variation* which measures pitch variability [63]. Multiple papers using the ELEA Corpus use a structure similar to this for *Audio Nonverbal Feature*, which is a combination of speaking turn feature and prosodic nonverbal cues.

3.2.2 Movement

The movement data channel is also used in many papers dealing with the ELEA corpus. Movement is data taken when a participant is in motion with any part of their body. Okada et al. [54] propose a new framework to extract features for finding personality traits, leadership, and communication skills. To create this framework they extracted two different visual features to assist in creating this framework. The first is binary activities status, those metrics from the binary activity status are checking the activity of the participants' head movement. The other set of features tracks the participants' continuous movement, which is the time-series change of motion activities during the experiment. Those features are made using optical flow to track the participants' upper body region and head. Besides the basic head and body movement metrics found in many of these papers, some use unique metrics extracted from the ELEA Corpus. Umut Avi et al. [4] studies the relationship between group members, and their compositions, against team performance. The authors' used metrics extracted from the ELEA Corpus consisting of nonverbal audio-verbal behavioral features to assist with searching for a relationship between the group and their team performance. Some of these metrics include head and body motion status, but also contain unique metrics such as *Unsuccessful Interruptions* for head and body movement. Relating to the types of interruption mentioned before, an interruption is when a participant breaks the continuity of another participant. In this case, the interruption takes place when the participant starts moving while another participant is also moving. The study only takes into account the unsuccessful attempt, meaning the interrupter stopped moving before the other participant did, and measures both head and body movement [4].

The ELEA Corpus is useful for filling holes where the AMI Meeting Corpus lacked with these speech and motion annotations. However, the ELEA Corpus is equipped with its own set of problems that need to be addressed. First, the ELEA Corpus is presented with data kept behind many walls. This proves frustrating when trying to decide which corpus is wanted in our research study because besides the preliminary study conducted, they do not provide well enough documentation to explain the data that comes with the dataset. Another issue with the corpus is seen in Figure 3.7 where there are obstacles in the cameras' view completing video processing programs that could lose track of certain body parts if moved behind one of these obstacles. Also, the corpus does not come complete with an official transcript of the group meetings.

Figure 3.7: This figure shows an example of the flaw in the video data provided by the ELEA corpus where the microphone in the middle of the video frame potentially blocking movement data [63]

.

3.3 UGI Corpus

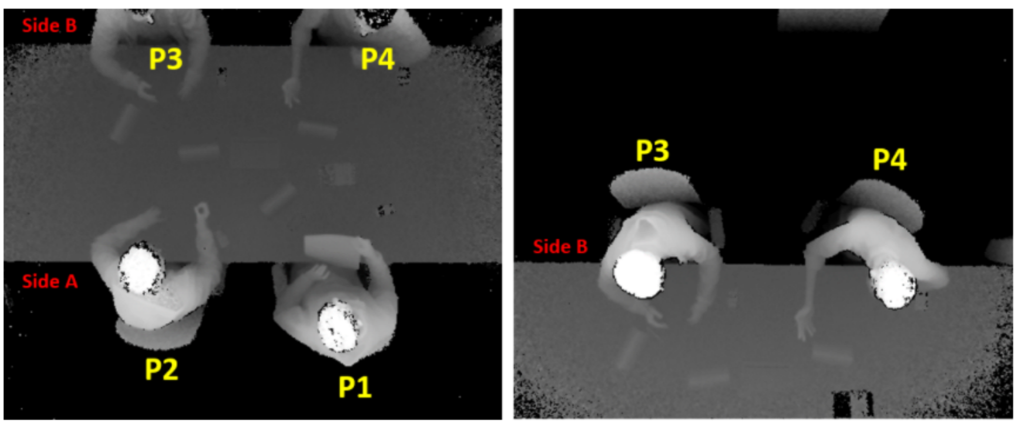
The *Unobstructive Group Interaction* (UGI) corpus presents a new multimodal dataset studying group dynamics with the use of ceiling-mounted depth sensors [8]. This corpus contains a combination of video data of participants' heads and bodies as well as audio data with time-stamped meeting transcripts. The content recorded the *Lunar Survival Task*. A small group must rank the value of 15 supplies from most to least important depending on how that item might aid them to continue living on the moon. This task, much like the *Winter Survival Task*, was broken up into two stages: the first had each person rank their items' value, and once complete the second task was then to share their ranking with the group and attempt to create a group ranking for each item. After the end of the group task, all the participants completed a post-task questionnaire. The post-task questionnaire contained questions to retrieve information about each participant as well as follow-up questions regarding the task [8]. The participants answered these questions

Figure 3.8: This image shows the room layout for UGI corpus data capture [8]

.

using a 5-point scale. The dataset comprises 86 participants, those participants were made up of undergraduate and graduate students that range from the ages 18 to 29 years old. From those participants, 22 small groups, consisting of 3-5 members, were formed and recorded doing this task [8].

The UGI study room is shown in Figure 3.8. The UGI team configured the room with microphones placed at different angles to pick up audio data from around the room, as well as *Microsoft Kinects* to record the interaction and movement data between team members [8]. Figure 3.9 visualizes the video data captured from the *Kinect* looks like for each group meeting. Each group meeting has two videos, one for each *Kinect*, showing a clear video of every participant on either side of the table. The video data is also in black and white where the whitest part is the top of the depth map captured. This is caused by the Kinect's depth sensor which causes the things closer to the sensor the be lighter pixels and farther pixels darker. The audio data was recorded using lapel microphones attached to each participant. The speech data is preprocessed to reduce any noise that messes with audio taken from all the participants [8]. The data is also transcribed to text using a speech-to-text API. Once the audio was transcribed, it was then hand-checked for any error processed by the API. An example of the transcript can be found in Figure 3.9; where the transcript is broken up into which person in the group is speaking followed by what they said. This is in addition to a timestamp on each line to allow synchronization between the audio and video data [8]. Along with the video and audio data, the dataset also provides the *Lunar Survival Task* information sheet that was given to all participants, the post-task questionnaire given after completing the *Lunar Survival Task*, the answer to the post-task questionnaire, and rankings of each group, as well as the rankings of all the participants [8].



An example transcript snippet.

```
40.64:>>Person2>>So then the second thing I had was water.
42.99:>>Person3>>Yeah, I put water too.
43.14:>>Person1>>I had the five gallons of water as well.
45.14:>>Person4>>I had magnetic compass, but I guess I'm in agreement with water too. Water was anyway pretty high, in my, it was in three.
53.48:>>Person1>>As your third one? Okay, what was the reasoning for the, putting the magnetic compass at two?
54.45:>>Person3>>Okay, yeah.
57.81:>>Person4>>I was like, we're two hundred miles away, right? So, if we don't know which direction to move, then we can't even start moving towards the
mother ship.
```

Figure 3.9: Here two examples of the different data provided by the UGI corpus. The upper part of the figure shows the Kincet video capture and the other shows a sample transcript snippet [8]

.

Zhang et al. [87] presented a method to study team performance using non-verbal communication leveraging the UGI data. The paper uses several metrics from different data channels with the majority coming from the movement data channels. They plan to accomplish this by creating an algorithm that computes a visual focus of attention of each participant and then uses prosodic features for non-verbal speech analysis. The algorithm for calculating the visual focus of attention uses a metric extracted from the UGI corpus data. These metrics were head pose angles, general visual metrics, and eye gaze angles. The UGI corpus is similar to the ELEA corpus which captures video and sound data channels. The UGI corpus is a recent corpus but still has interesting literature using its data. Lin and Lee [47] use both the UGI and GAPS corpus to build upon a previous study using the ELEA corpus. Lin and Lee proposed a model that predicts team performance using face-to-face behavioral interaction [47]. UGI data usage in the study augments the data previously taken from the ELEA dataset, to assist with the stabilization and improvement of their predictor's accuracy. This is only possible because the three corpora used similar tasks when creating their corpora.

The UGI corpus had numerous limitations. The main problems come from the video data capture is less than ideal given they only provide the depth map video and not the RGB video. The depth map would greatly augment the standard RGB data, however by itself is problematic to process as much of the video has background noise. The background noise creates issues when having to run the video through different detection programs because noise can be mistaken as movement and skew the data. The actual taping of participants was not choreographed well either, some of the participants can be found leaning out of the frame of the camera view or entering a different participant's space. This makes it difficult to process the video data collected for more advanced metrics. Figure 3.10 shows an example of a participant from one group leaning back in the chair and completely leaving the frame. These issues can be solved by processing the data through filters to cancel out some of the noise and a border could be created in the space so there's

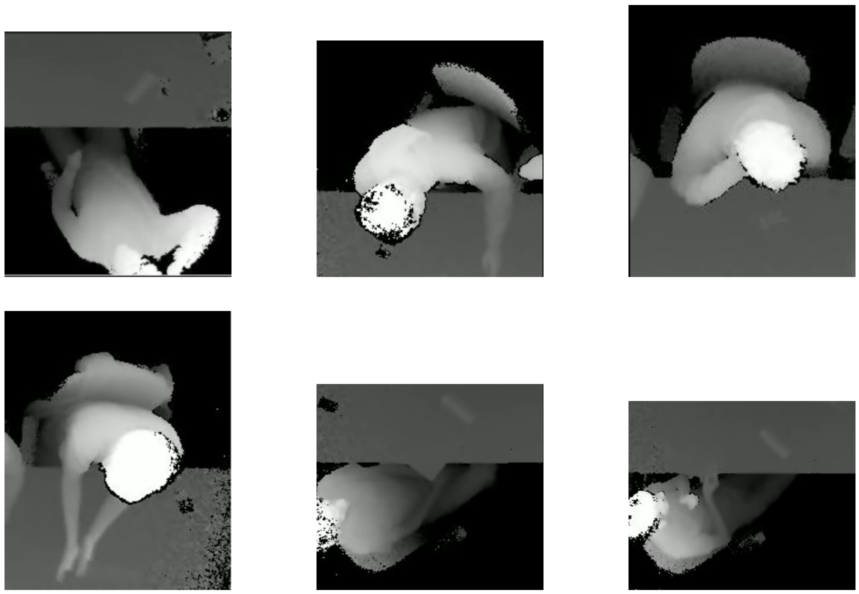


Figure 3.10: Some participant video frames that contain error such as moving out frame of the camera view or having other participant enter their frame

a clear line where the video stops picking up the participants. Another issue is the transcripts and video data do not sync properly. In some cases, the video recording would run longer than the final time found in the transcripts making it difficult to figure out the proper timing between the two sets of data.

3.4 GAPS Corpus

The Group Affect and Performance (GAP) corpus, collected by the team at the University of the Fraser Valley, presents a dataset using social signals and natural language processing to fill the gap found in team research. *Social signal processing* (SSP) is the process of using technology to model and analyze social interaction displayed by groups [81]. *Natural language processing* (NLP) is the process of teaching computers to understand written and spoken words [30]. These two fields, SSP and NLP, benefit from the nonverbal and verbal cues assessed by the recordings of small group dynamics. The GAP corpus contains audio recordings along with the transcript of the recordings, annotations of the meeting, and a questionnaire surveying satisfaction, demographic, and influence [11]. GAP uses the *Winter Survival Task* in the study, with participants first completing a ranking task individually and then performing the same ranking task jointly as a group. Once the task was completed, each participant completed a post-task questionnaire [11]. The questionnaire asked the participants for general information about themselves, then they were able to ask questions using a 5-point Likert scale with their opinions about certain aspects of the meeting. Figure 3.11 gives a glimpse of the types of questions that were asked on the questionnaire. The post-task questionnaire motive was to calculate the overall satisfaction of the groups as well as highlight any leadership qualities found thorough the process of the group meetings. The dataset comprises 37 participants, the participants were recruited from the university's online sign-up recruitment system. From those participants, 28 small groups, consisting of 2-4 participants, were formed and recorded doing the task [11].

Label	**Item**
Time Expectation	(1) "This task took longer than expected to complete."
Worked Well Together	(2) "Our group worked well together."
Time Management	(3) "Our group used its time wisely."
Efficiency	(4) "Our group struggled to work together efficiently on this task."
Overall Quality of Work	(5) "Overall, our group did a good job on this task."
Overall Satisfaction	Items one to five combined and averaged.
Leadership	(6) "I helped lead the group during this task."

Figure 3.11: This is an example of the post-task questionnaire provided to participants to gauge the group's satisfaction and determination of leadership [11]

.

The meeting audio was recorded using the *Zoom H1 Handy Recorder* which is a portable high-quality audio-recording device. This device was placed in the center of the group members. Each participant was also recorded using a webcam, which was placed directly in front of each participant [11]. This corpus does not contain any video data, all video records are taken with the audio data and turned into transcripts. The GAP Corpus first segmented the audio data where there were clear topics for each sentence spoken by the participants. Once those sentences were segmented, they were then transcribed word for word. The GAP Corpus did not manually fix any errors, false starts, stutters, or pauses [11]. They were all left in the transcript and marked differently depending on the error. The transcript is separated by when a participant spoke, start and end time, followed by the sentence that was said in the allotted time. There was also an organizational coding scheme implemented into the transcript. These different codes served different purposes whether it was to keep track of group member data, identification code for each group member, or code in text files keeping track of members, and several utterances. The corpus also includes two different annotations, one called the *sentiment annotation* which referred to the underlying emotion of the speech segment. The *sentiment annotation* is separated between two sentiment values, these values are positive sentiment or negative sentiment. Sentiment values were determined depending on certain phrases, either positive or negative, participants used during the meeting [11]. The second annotation was based on the groups' decision-making, they measured the amount of decision-making seen during the four phases of decision-making. The decision-making model included a proposal, agreement or disagreement, and confirmation [11]. Other data that is offered from the GAP corpus is scoring for both the task performance and the post-task questionnaire. The *Winter Survival Task* scoring is derived from the scores of the absolute group score, absolute individual score, and absolute individual influence. The post-task questionnaire only scored the questions that dealt with the actual task, and the score was based on the overall satisfaction of the task which was calculated for both an individual and group level.

The GAP corpus is another relatively newer corpus, which leads to few papers published at the moment of this thesis. Kubasova et al. [40] is an example of a paper using the GAP corpus data as a predictor for team performance using verbal and nonverbal features. From the GAP corpus, several features were extracted for the use of this paper both on the verbal and non-verbal sides. For the non-verbal features, a large number of audio features were extracted from the audio channel. The verbal features that were extracted from the transcripts of the meetings included *Dependency Parse Features*, *Part-of-Speech Tags*, *Filled Pauses*, *Psycho-linguistic*, *Sentiment, GloVe Vectors*, *Lexical Cohesion*, and *Sentence and Document Length*. These features are extracted to assist with the finding of prediction for group task performance. The GAP Corpus augmented data from the ELEA Corpus to assist with the prediction's creation [40]. Various papers in the literature used these datasets together. Another example of this is Kubasova et al. [41] combined the data of three corpora, the GAP Corpus, the UGI Corpus, and the ELEA Corpus. With these three corpora, Kubasova et al. wanted to see if they could predict team performance solely with the data from the start of the meeting. Features were extracted from the first minute of all the conversations that took place during the meetings. From the features extracted, they fell into two groups, one being the *linguistic features* that had been seen already used in previous works; these included filled pauses, psycho-linguistic, image-ability, the typical age of acquisition, familiarity, as well as other metrics that were mentioned before. The other features are listed with *Graph-based Conversation Interaction Features*, which are features that assist to create a graphical model to represent the group interactions that took place during the conversation [41]. The model is equipped with all participants found in each meeting along with sentences they spoke and the topic for each sentence. Then running through the model tracks, the number of interactions in a meeting, which person influenced the conversation the most, and how different topics were spread throughout the meeting [41].

The GAP corpus limitations suffer from some issues that keep it from being used more often. The lack of video data is the main issue, which limits the study analyses. Therefore, any study that needs the video data on top of the existing audio data cannot use this dataset alone. The lack of video data would have not been as important if it was paired with a more in-depth transcript. The transcript presented by GAP corpus does not present any major issue but is lackluster compared to the AIM transcripts. The GAP Corpus has less to offer in small meeting research besides the addition of emotion sentiment in their annotation section and additional data.

CHAPTER 4: METHODOLOGY

This chapter discusses our general methodology. The aim of this work is to explore how to understand team performance by leveraging motion dynamics metrics extracted from contactless unobtrusive sensor data. When determining which corpora to use, I first ensured the corpora would provide useful data under these constraints for both the audio and video channels. As a result, the UGI dataset was chosen to test our preliminary team performance analysis. This dataset contained 22 small groups, each with two videos showing participants from either side of the table. In addition to the video data, the UGI corpus also contained a transcript for all 86 participants from the meeting, as well as score and ranking quantitative data for both the teams and participants from the task at hand during the meeting. In this section, I explain the process used for the quantitative data, video data, and audio data for the team performance analysis.

4.1 Quantitatively Analyzing Team Performance from Ranking Scores

4.1.1 Score Analysis

The UGI corpus contained both each individual member's rankings and team final ranking data from the *Lunar Survival Task*. Along with these rankings, I have subjective survey data containing the scores of the participant's familiarity with group members, whether the participant acted as a leader, and how the participant acted with other group members. To gain more insight from the rankings, I found three comparisons to test to understand individual and team performance: (1) comparing the individual ranking to the expert's ranking to understand how each team member performed, (2) comparing the group ranking to the expert's ranking to understand how the group performed, and (3) comparing the individual's ranking to their group's ranking to understand the change during the group session. I derived a series of (performance) scores based on these three categories.

4.1.1.1 Individual Score vs. Expert Rubric (Difference)

Table 4.1: This table shows an example comparison between the expert score and a random participant's score received from the dataset. As well as the absolute value of the differences between the two rankings.

Items	**Expert Score**	**Individual Score**	**Difference**
Matches	15	14	1
Food	4	4	0
Rope	6	6	0
Parachute	8	10	2
Heating	13	7	6
Pistols	11	13	2
Milk	12	11	1
Oxygen	1	1	0
Map	3	5	2
Raft	9	8	1
Compass	14	15	1
Water	2	2	0
Flares	10	12	2
FirstAid	7	9	2
FM	5	3	2

To compare the individual's ranking and the expert ranking, I first gather the set of individual rankings $\boldsymbol{E}$ and the set of $\boldsymbol{I}$ expert rankings. Each set contained the ranking of the fifteen items (i) each individual team member needed to rank. There is only one set of expert rankings, and there are 86 participant sets. Table 4.1 shows the provided expert score and a random participant's score. These two sets go through a variety of comparisons to determine the performance scores described in the equations below.

$$\boldsymbol{E} = \{\ E_1 = \{i_1\},\ E_2 = \{i_2\},\ \dots\ , E_{15} = \{i_{15}\}\} \tag{4.1}$$

$$\boldsymbol{I} = \{\ I_1 = \{i_1\},\ I_2 = \{i_2\},\ \dots\ , I_{15} = \{i_{15}\}\} \tag{4.2}$$

$$\boldsymbol{S_1}^I = \sum_{i=1}^{15} \mid E - I \mid \tag{4.3}$$

Where $\boldsymbol{E}$ is the expert ranking for all the items and $\boldsymbol{I}$ is the individual ranking for all the items; with the i representing the specific item. Equation 4.3 subtracts the two different rankings for each item calculating how far the individual ranking for that item was from the expert ranking. An example of the differences between those two rankings can be seen in Table 4.1. Once all the differences have been calculated, they are then added together. The number calculated from this equation will be a new score, the difference between the expert's ranking and the individual's ranking, $\boldsymbol{S_1}^I$. $\boldsymbol{S_1}^I$ is a score calculated using the equation presented by the *Lunar Survival Task*. Alongside the new score, $\boldsymbol{S_1}^I$, I also use Equation 4.4 which calculates the average gap between the expert ranking and individual ranking.

$$\boldsymbol{S_2}^I = \frac{\boldsymbol{S_1}^I}{n} \tag{4.4}$$

Where n is the number of items on the list which is 15. This process will be done for all 86 participants.

Another comparison tested between the individual ranking and expert ranking was reversing the order of the expert ranking to change the penalty when an item was incorrectly ranked.

$\overline{\boldsymbol{E}}$ is added to create a more sensitive indicator when locating errors in the rankings. This changes the equation by having the last item of the expert ranking subtract itself from the first item on the individual ranking.

$$\overline{\boldsymbol{E}} = \{\ E_1 = \{i_{15}\},\ E_2 = \{i_{14}\},\ \dots\ , E_{15} = \{i_1\}\} \tag{4.5}$$

$$\boldsymbol{S_3}^I = \sum_{i=1}^{15} \mid \overline{E} - I \mid \tag{4.6}$$

Where $\overline{\boldsymbol{E}}$ is the reverse ordering of the expert ranking of all the items, $\boldsymbol{I}$ is the unchanged individual ranking of all the items; with i being the specific item in the ranking. The difference found in Equation 4.6 will follow $\boldsymbol{S_1}^I$ score equation to create a new scoring by adding all the differences. $\boldsymbol{S_3}^I$ is the summation of all the differences between the two rankings. I will also use Equation 4.7 to calculate the average gap between the reverse order expert ranking and individual ranking by dividing $\boldsymbol{S_3}^I$ by the number of items on the ranking.

$$\boldsymbol{S_4}^I = \frac{\boldsymbol{S_3}^I}{n} \tag{4.7}$$

Where n is the number of items (which is 15). These scores are calculated for all 86 participants in the dataset.

4.1.1.2 Individual Score vs. Expert Rubric (RBO)

I also wanted to compare these to rankings while also penalizing the participants for incorrect ranking for the more important items, such as oxygen, and less so for items such as the compass or matches. The method I used was a *Rank Biased Overlap* (**RBO**) which applies weight to each rank position. The equation on how **RBO** applies the weight is:

$$\boldsymbol{RBO}(S, T, p) = (1 - p) \sum p^{d-1} \cdot A_d \tag{4.8}$$

Where d is the depth of the ranking being provided, X_d is the size of the overlap of S and T up to depth d, and A_d is the agreement between S and T given by the proportion of the size of the overlap up to depth d. The parameters are S and T, which are the two rankings being compared, and p, which is a parameter between the range $(0, 1)$ that is used to determine the weight top d ranks will have compared to the other ranking. This equation will provide a value in the range of $[0, 1]$, where 0 determines the lists are completely different and 1 determines the lists are identical [34]. For this test, I selected three p-values, which were 1, 0.9, and 0.8. With the test, I also check if the participant ranking had any item that contains two items with the same rank. If this was the case then the **RBO** score for that participant would be set to null for all p values and the testing would continue to the next participant.

4.1.1.3 Team Scores vs Expert Scores

The team scores were calculated by subtracting the ranking of each item in the team ranking and the expert ranking. For the team ranking data provided, there are cases of a null value in the place of a ranking. These null values are placed there to indicate that the group could not come to a consensus on that item ranking. This was the basis for calculating several team performance scores to handle this lack of consensus in various ways. This can be seen in $\boldsymbol{T^+}$ which goes up until the non-null item value, n^+, in the ranking; and $\boldsymbol{T^-}$ which goes up until the last null value, n^-, in the ranking. When subtracting a null ranking from another ranking, that null ranking will act as zero as shown in Equation 4.13. Similar to the $\boldsymbol{S_1}^I$ scoring, each team's rankings go through the equation for each item:

$$\boldsymbol{T} = \{\, T_1 = \{i_1\},\ T_2 = \{i_2\},\ \dots\ , T_{15} = \{i_{15}\}\} \tag{4.9}$$

$$\boldsymbol{T} = T^+\ +\ T^- \tag{4.10}$$

$$\boldsymbol{T^+} = \{\, T_1^+ = \{i_1\},\ T_2^+ = \{i_2\},\ \dots\ , T_{n^+}^+ = \{i_{n^+}\}\} \tag{4.11}$$

$$\boldsymbol{T^-} = \{\, T_1^- = \{null\},\ T_2^- = \{null\},\ \dots\ , T_{n^-}^+ = \{i_{n^-}\}\} \tag{4.12}$$

$$\boldsymbol{T_0}^- = \sum_{i=1}^{n^-} R[i] = 0 \tag{4.13}$$

$$\boldsymbol{T_0} = T^+ + T_0^- \tag{4.14}$$

$$\boldsymbol{S_1}^T = \sum_{i=1}^{15} \mid E - T_0 \mid \tag{4.15}$$

Where $\boldsymbol{E}$ is the expert ranking for all the items and $\boldsymbol{T_0}$ is the team ranking for all the items with the null values changed to 0 as seen in Equation 4.14; with the i representing the specific item.

Similar to ${\boldsymbol{S_1}}^I$, Equation 4.15 shows the two different rankings being subtracted to calculate the difference between the team ranking against the expert ranking. Once all the differences have been calculated then all the differences will be added together. The number calculated from this equation will be the score, the difference between the expert ranking and the team's ranking, ${\boldsymbol{S_1}}^T$. ${\boldsymbol{S_1}}^T$ is a score calculated using the equation presented by the NASA Lunar Survival Task but using the team ranking rather than the individual ranking. In ${\boldsymbol{S_1}}^T$, I use Equation 4.16 to calculate the average gap between the expert ranking and team ranking.

$$
{\boldsymbol{S_2}}^T = \frac{{\boldsymbol{S_1}}^T}{n} \tag{4.16}
$$

Where n is the number of items on the list which is 15. This process will be done for all 22 groups.

Similar to the comparison between the individual ranking and expert ranking, I also did the reverse ordered expert ranking using Equation 4.5 against the team ranking.

$$
{\boldsymbol{S_3}}^T = \sum_{i=1}^{15} \mid \overline{E} - T_0 \mid \tag{4.17}
$$

Where $\overline{\boldsymbol{E}}$ is the reverse ordering of the expert ranking, $\boldsymbol{T_0}$ is the team ranking with the null values replaced with 0, and i is the specific item in the ranking. The difference found in the Equation 4.17 will follow ${\boldsymbol{S_1}}^T$ score equation to create a new scoring ${\boldsymbol{S_3}}^T$ which only changes the order of the expert ranking. ${\boldsymbol{S_3}}^T$ is the summation of all the differences between the reverse ordered expert ranking and team rankings. Equation 4.18 will also calculate the average gap between the reverse order expert ranking and individual ranking by dividing ${\boldsymbol{S_3}}^T$ by the number of items on the ranking.

$$
{\boldsymbol{S_4}}^T = \frac{{\boldsymbol{S_3}}^T}{n} \tag{4.18}
$$

Where n is the number of items (which is 15). These two scores are determined for all 22 groups in the dataset.

4.1.1.4 Team Scores vs Expert Scores (Non-consensus Penalty)

I plan to calculate the lack of items that lack consensus into two different types of penalties that would affect their overall team scores in different ways. The first penalty adds fifteen points to the group's team score for each null value that the group contains. The fifteen-point penalty was chosen for the fifteen different items in the list. Before adding the penalties, I first found the score between the team ranking and expert ranking but excluded the null values. The null values, in this case, would be ignored, and only if both lists had a numeric ranking for that item would the subtraction between them be calculated.

$$\boldsymbol{S_5}^T = \sum_{i=1}^{n^+} \mid E[i] - T^+[i] \mid \tag{4.19}$$

Where $\boldsymbol{T^+}$ being the non-null rankings from the team ranking which is one part that makes the full team ranking shown in Equation 4.10. With n^+ being the non-null values in the team ranking, $\boldsymbol{E}$ being the expert ranking, and i being the items on the specific ranking. Once differences in Equation 4.19 between the non-null values were found between the expert ranking and the team ranking, the numbers would be added together to give a new team score, $\boldsymbol{S_5}^T$. Once a score containing solely the non-null ranking was calculated, I can now calculate the first penalty.

$$\boldsymbol{S_6}^T = (15 * d) + \boldsymbol{S_5}^T \tag{4.20}$$

With d being the number of times the team did not come to a consensus in the team ranking.

Equation 4.20 takes the score found from Equation 4.19 and adds 15 points for each non-consensus found within the team ranking, penalizing teams that were unable to come to a group consensus before the end of their meeting. This process is also done with the reversed order expert ranking.

$$\boldsymbol{S_7}^T = \sum_{i=1}^{n^+} \mid \overline{E[i]} - T^+[i] \mid \tag{4.21}$$

Where $\boldsymbol{T^+}$ being the non-null rankings from the team ranking, n^+ being the non-null values in the team ranking, $\overline{\boldsymbol{E}}$ being the reverse ordered expert ranking, and i being the items on the specific ranking. Similar to Equation 4.20, I are going to add the 15 point non-consensus penalty to $S_7{}^T$ to create Equation 4.22 with d being number of non-consensuses found in the team ranking.

$$\boldsymbol{S_8}^T = (15 * d) + \boldsymbol{S_7}^T \tag{4.22}$$

Another penalty used was adding the average score of items that were missed in the team ranking. To calculate this penalty I need to add the score of only non-null values with the average of the null values.

$$\boldsymbol{S_9}^T = \boldsymbol{S_5}^T + \frac{\sum_{i=1}^{n^-} \mid E[i] - T_0^-[i] \mid}{n^-} \tag{4.23}$$

Where $\boldsymbol{S_5}^T$ provides the scores for the non-null values; $\boldsymbol{E}$ is the expert ranking; T_0^- is the team ranking only with the null values, which all null values are now 0; and n^- which is the number of the null values in the team ranking. Equation 4.23 uses a similar method by finding the difference between expert ranking and null values of the team ranking that were turned into 0, then that difference is divided by the number of null values in the team ranking to provide the average difference. The score, $\boldsymbol{S_7}^T$, is then made of the average difference that will serve as the penalty which is added to the score containing only the non-null values from the team ranking. I also calculate the average penalty for the reversed order expert ranking as well.

$$\boldsymbol{S_{10}}^T = \boldsymbol{S_7}^T + \frac{\sum_{i=1}^{n^-} \mid \overline{E[i]} - T_0^-[i] \mid}{n^-} \tag{4.24}$$

Where $\boldsymbol{S_7}^T$ provides the scores for the non-null values with the reverse order expert ranking; $\overline{\boldsymbol{E}}$ is the reversed ordered expert ranking; T_0^- is the team ranking only with the null values; and n^- which is the number of the null values in the team ranking. Similar to $\boldsymbol{S_9}^T$, Equation 4.24 finds the sum of each difference between the reverse ordered expert ranking and the team ranking including only the null values, which have been turned to 0. Then adds the quotient of the summation and the number of times the group did not reach consensus with $\boldsymbol{S_7}^T$ to provide the penalty score $\boldsymbol{S_{10}}^T$.

The final penalty introduces a penalty where mistakes created earlier in the ranking of the items would be penalized more. For this penalty, the difference between the group ranking and expert ranking would be subtracted, the non-consensus would be replaced with an automatic score of fifteen in place of the number while subtracting the two lists.

$$\boldsymbol{S_{11}}^T = \sum_{i=1}^{15} \left| E[i] - T_p[i] \right| \cdot (15 - (x[i] - 1)) \tag{4.25}$$

Where $\boldsymbol{T_p}$ is the penalty team ranking that if team ranking produced a null value then the calculation from $\left| E[i] - T_p[i] \right|$ will equal 15 automatically and x is the ranking number for that item. Equation 4.25 shows the difference between the expert ranking and the new penalty team ranking. Once the difference for all the items has been calculated, then each ranking difference will be multiplied by the inverse of their expert ranking, then added together.

4.1.1.5 Team Scores vs Expert Scores (Accuracy)

The scores also accounted for the accuracy of the team ranking when comparing it to the expert ranking. Equation 4.26 calculates how many items in the team ranking matched with the items on the expert ranking. To calculate this, a conditional value variable $T^{'}_{ans}$ for all 15 items is added, i in the specific items in the ranking, E which is the expert ranking, and T_0 which is the team ranking that substitute null values for 0. $T^{'}_{ans}$ is a conditional value that is determined by whether the difference between $\mid E - T_0 \mid$ is 0 or greater than 0. If the difference is 0 then $T^{'}_{ans}$ for this item is 1 noting the two ranked items are the same, otherwise, $T^{'}_{ans}$ will be 0 noting they are different.

$$\boldsymbol{S_{ans}}^T = \sum_{i=1}^{15} T^{'}_{ans} \tag{4.26}$$

$$T^{'}_{ans} = \begin{cases} 1, & \text{if} \mid E - T_0 \mid = 0 \\ 0, & \text{if} \mid E - T_0 \mid \neq 0 \end{cases}$$

To calculate the accuracy of the two rankings while allowing some room for error between the rankings, Equation 4.27 calculates how many items in the team ranking match with either being exact or one spot away from the expert ranking. The conditional value for this equation, $T^{'}_{alm}$, which is the difference between $\mid E - T_0 \mid$ is less and equal 1 or greater than 1. If the difference is less and equal 1 then $T^{'}_{alm}$ for this item is 1 noting the two ranked items are either the same or one away from the expert ranking, otherwise $T^{'}_{alm}$ will be 0 noting they difference is higher than one. Other factors to calculate Equation 4.27 include i in the specific items in the ranking, E which is the expert ranking, and T_0 which is the team ranking that substitutes null values for 0.

$$\boldsymbol{S_{alm}}^T = \sum_{i=1}^{15} T^{'}_{alm} \tag{4.27}$$

$$T'_{alm} = \begin{cases} 1, & \text{if} \mid E - T_0 \mid \leq |1| \\ 0, & \text{if} \mid E - T_0 \mid > |1| \end{cases}$$

I also wanted to calculate the averages for the times there was non-consensus in the team ranking.

$$\boldsymbol{S_{miss}}^T = \frac{\sum_{i=1}^{n^-} \mid E[i] - T_0^-[i] \mid}{n^-} \tag{4.28}$$

where i in the specific items in the ranking. Equation 4.28 sums the difference between the expert ranking, E, and team ranking's non-consensuses, T_0, where all the null values are replaced with 0. Then divides the difference by the number of non-consensus, n^-, the team ranking had to produce the score, $\boldsymbol{S_{miss}}^T$. In addition, I also calculated the average for the lack of consensus using the reverse ordered expert ranking. Equation 4.29 follows the same path to produce $\boldsymbol{S_{miss}}^T$, but instead it sums the difference between the reversed ordered expert ranking, $\overline{E}$, and team ranking's non-consensuses, T_0, where all the null values are replaced with 0. The rest of the process is the same as Equation 4.28.

$$\boldsymbol{S_{\overline{miss}}}^T = \frac{\sum_{i=1}^{n^-} \mid \overline{E}[i] - T_0^-[i] \mid}{n^-} \tag{4.29}$$

4.1.1.6 Individual Score vs Team Scores (Differences)

The final comparison measured the individual participants ranking against their own group's ranking. To measure this I subtracted the ranking from both the individual ranking and the group's ranking. Similar to the other rankings previously calculated, two rankings go through the equation for each item on the equation:

$$\boldsymbol{S_1}^C = \sum_{i=1}^{15} \mid I - T_0 \mid \tag{4.30}$$

With the i represent the specific item being subtracted, $\boldsymbol{I}$ being the individual ranking, T_0 being the individual's team ranking but with the null values substitute for 0. As seen in Equation 4.30, once all the differences have been calculated then they will be added together. The number calculated from this equation will be the new score noting the change between the individual's own ranking with the ranking the individual produced with the team. This process will be done for all 86 participant and their corresponding groups. Equation 4.31, I calculated the average difference between these two rankings.

$$\boldsymbol{S_2}^C = \frac{S_1^C}{n} \tag{4.31}$$

Where n is the number of item in the ranking which is 15.

To complement the scores found from the individual ranking and their group ranking, I subtracted the score gained from the individual and expert ranking, against the score from their group and expert ranking.

$$\boldsymbol{S_3}^C = \boldsymbol{S_1}^I - \boldsymbol{S_1}^T \tag{4.32}$$

Equation 4.32 was to see if the individuals score, $\boldsymbol{S_1}^I$, improved while in their groups, $\boldsymbol{S_1}^T$, or worsened as a result. The team ranking did use the score that changes the null values with 0. Similar to Equation 4.31, I also calculated the average difference between the two scores.

$$\boldsymbol{S_4}^C = \frac{S_3^C}{n} \tag{4.33}$$

Where n is the number of item in the ranking which is 15.

The final set of calculations to find differences between the individual and their groups are Equation 4.34 and Equation 4.35.

$$\boldsymbol{S_5}^C = \sum_{i=1}^{n^+} \mid I[i] - T^+[i] \mid \tag{4.34}$$

Where $\boldsymbol{T^+}$ being the non-null rankings from the team ranking which is one part that makes the full team ranking shown in Equation 4.10. With n^+ being the non-null values in the team ranking, $\boldsymbol{I}$ being the individual ranking, and i being the items on the specific ranking. Equation 4.34 subtracts the individual and team rankings but ignores any disagreement, or null values, that their group ranking may have. Then those differences will be added together to create the individual vs team score excluding any non-consensus. Equation 4.35 will show us the average difference between the individual ranking and their group ranking without non-consensuses.

$$\boldsymbol{S_6}^C = \frac{S_5^C}{n^+} \tag{4.35}$$

Where n^+ is the number of non-null items in the ranking which is 15.

4.1.1.7 Individual Score vs Team Scores (Non-consensus Penalty)

Next, I must calculate scores penalizing the teams for containing values where there were no consensuses. These penalty scores will be similar to the score derived from penalizing the score between the expert ranking and the team ranking where we first calculate a score between the individual ranking and their group's ranking that contain zero non-consensuses. In Equation 4.35 I calculate that score and it becomes our base which I will add penalties to reflect the non-consensuses found in each team's rankings.

Resembling the penalties calculated before Equation 4.36 will add 15 point on to base score, which is ${\boldsymbol{S_5}}^C$, for each penalty the team suffered during their meeting.

$$ {\boldsymbol{S_7}}^C = (15 * d) + {\boldsymbol{S_5}}^C \tag{4.36} $$

With d being number of non consensuses found in the team ranking and ${\boldsymbol{S_7}}^C$ is the base score that provides the scores for the non-null values. Next penalty will be Equation 4.37 that will add the average of missed items in team ranking to the base score.

$$ {\boldsymbol{S_8}}^C = {\boldsymbol{S_5}}^C + \frac{\sum_{i=1}^{n^-} \mid I[i] - T_0^-[i] \mid}{n^-} \tag{4.37} $$

Where ${\boldsymbol{S_5}}^C$ is the base score that provides the scores for the non-null values; $\boldsymbol{I}$ is the individual ranking; T_0^- is the team ranking only with the null values, which all null values are now 0; and n^- which is the number of the null values in the team ranking. Equation 4.38 will use the final penalty method demonstrated in Equation 4.25, this equation will find the difference between the individual ranking and a penalized team ranking which changes the null values into 15. That difference is calculated and then multiplied by the reverse ranking of that specific item. Once that product is found for all 15 items then it is added to the score to create a new score, ${\boldsymbol{S_9}}^C$.

$$ {\boldsymbol{S_9}}^C = \sum_{i=1}^{15} \left| I[i] - T_p[i] \right| \cdot (15 - (x[i] - 1)) \tag{4.38} $$

Where $\boldsymbol{T_p}$ is the penalty team ranking that if team ranking produces a null values than the calculation from $\left| E[i] - T_p[i] \right|$ will equal 15 automatically and x is the ranking number for that item.

4.1.1.8 Individual Score vs Team Scores (Accuracy)

Similar to the accuracy scores associated with the expert rankings, I also wanted to account for the accuracy between the individual ranking and the expert ranking. The first equation will calculate the number of exact matches between the individual ranking and their group's ranking.

$$\boldsymbol{S_{ans}}^C = \sum_{i=1}^{15} C'_{ans} \tag{4.39}$$

$$C'_{ans} = \begin{cases} 1, & \text{if} \mid I - T_0 \mid = 0 \\ 0, & \text{if} \mid I - T_0 \mid \neq 0 \end{cases}$$

In Equation 4.39, I add all the values calculated from conditional variable C'_{ans} for all 15 items, i, in the ranking. C'_{ans} is a conditional variable that produces a 0 or 1 depending on the answer found for the equation $\mid I - T_0 \mid$, where I is the individual ranking; T_0 is their group ranking with the null values equal to 0. If the result of the ranking is equal to 0 then C'_{ans} equals 1, otherwise C'_{ans} will equal 0. I also calculated the amount of ranking almost matching with the range being $[-1, 1]$.

$$\boldsymbol{S_{alm}}^C = \sum_{i=1}^{15} C'_{alm} \tag{4.40}$$

$$C'_{alm} = \begin{cases} 1, & \text{if} \mid I - T_0 \mid \leq |1| \\ 0, & \text{if} \mid I - T_0 \mid > |1| \end{cases}$$

Similar to Equation 4.39, Equation 4.40 adds all the values for a conditional values, the conditional value being C'_{alm} for all 15 items, i, in the ranking. C'_{alm} is a conditional variable that produces a 0 or 1 depending on the answer found for the equation $\mid I - T_0 \mid$, where I is the individual ranking; T_0 is their group ranking with the null values equal to 0. If the result of the ranking is less than or equal to 1 then C'_{alm} equals 1, otherwise C'_{alm} will equal 0.

4.1.2 Other Quantitative Measures

Apart from the ranking, there were also other subjective measures that were collected in the UGI dataset. The measure consisted of the participants ranking the contribution each group member had to the discussion and which participants acted most like the team's leader. These measures were originally ranked on a 5 point Likert scale, where 1 is where the participant showed a lower aptitude, and 5 is the participant showed a higher aptitude. Included were five different subjective measures indicating whether the participant thought the conversation in their group was either:

- Boring or Engaging: 1 being extremely boring and 5 being extremely engaging
- Cold or Warm: 1 being extremely cold and 5 being extremely warm
- Awkward or Comfortable: 1 being extremely awkward and 5 being extremely comfortable
- Dull or Interesting: 1 being extremely dull and 5 being extremely interesting
- Detached or Friendly: 1 being extremely detached and 5 being extremely friendly

Two scores had a score from each other participant including the participant's own ranking. From the ranking the participant received, I added those numbers and divided them by the number of people in the group to get their average score for that measure. The five different conversational subjective data were not altered.

4.1.3 Binary Scores

After receiving the data from the subjective score analysis above, a binary version of the analysis was made classifying the teams into two groups. This version would assign a 0 or a 1 depending on if the score passed a mean threshold or grouping category of the expert's scores. Each column in the document has its own threshold based on its data's mean. For columns containing the score for the original three comparisons, which were individual ranking vs expert ranking, team ranking vs expert ranking, and individual ranking vs expert ranking, the threshold was any score under 46 received a 1, while all other scores received a zero. For the number of non-consensus columns, if there is at least one non-consensus recorded the score will become 1. For the rest of the columns mentioned in the score analysis section, the threshold is the median score for each individual column. If the score is over the median, the score will become 1 and all other scores below the median score will receive 0. For the columns containing the information from the other measures, the threshold is determined by finding the mean score for each column and comparing it to all the scores. If any of the original scores are greater than the mean then that score will receive a 1 and score under the mean score will receive a 0.

4.2 Motion Analysis

4.2.1 Video Processing

The first step to analyze the UGI video data is to preprocess the data. The objective of video processing was to eliminate as much noise from the video as possible without losing or warping the movement of each individual participant. Then the noise in the video is the irregular fluctuation found that is not supposed to be a part of the video. For most participants, the video movements and the noise issue were found in the presence of another participant. As seen in Figure 4.1 many videos have multiple participants within them, making it difficult to track the movement of an individual participant without another participant interfering. To solve this issue, code was created using the *OpenCV* library, which is a library of programming functions mainly aimed at obtaining real-time information from digital images or videos. This code selected a specific region of interest, an individual participant, and created a separate video solely containing that participant. In addition to the creation of a separate video for each participant, the code also created a *Comma Separated Values* (CSV) file which contained the position of the top-left pixel of the region of interest taken, as well as the height and width of the region. After separating the participants into separate videos, I found there was still a noise issue that could cause possible errors. *OpenCV* library allowed us to use different filters to help fix these many of these distortions.

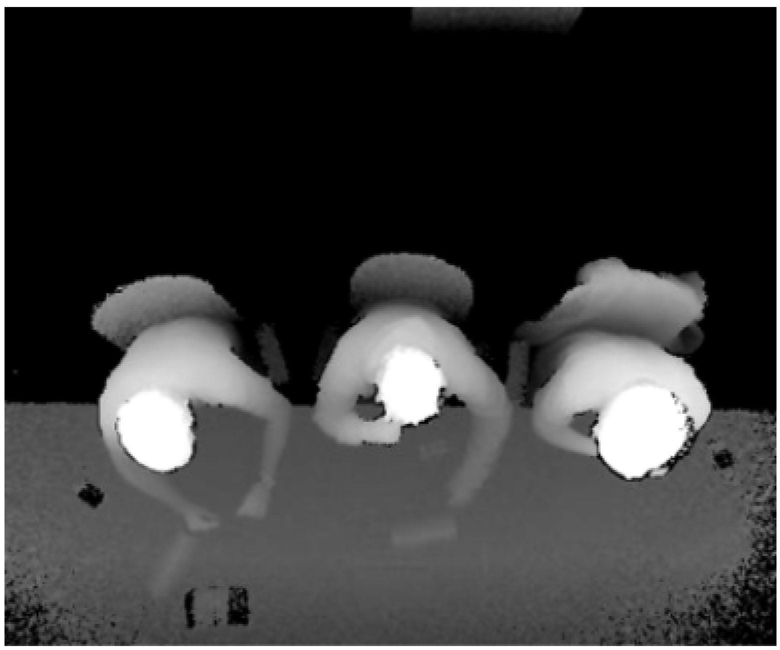


Figure 4.1: This figure shows an example of the raw video data given by the UGI dataset. This example shows the problem with the dataset's video data of having multiple people in frame.

4.2.2 Movement Tracking

To locate movement synchrony from the participant individual videos, I first extracted features from each frame from the participant videos. The extracted features from these videos allowed us to track the difference between each individual frame. This allowed us to track and record the movement of the participant along with the video.

4.2.2.1 OpenPose (attempt)

The first algorithm I used to attempt extracting motion features from the videos was OpenPose. The OpenPose code was a pre-made code developed by Gines Hidalgo and Yaadhav Raaj. The approach is the first real-time multi-person system that detects the whole human body including hands, feet, and facial feature. In total 135 key points can be seen on a single image. The technique works by reading the videos and outputs the video, including key points on top of the moving participant showing their movement as shown in Figure 4.2.2.1A. This technique was not used for the final product because of the poor video quality and position of the camera, it was difficult to get a proper reading to use.

4.2.2.2 Optical Flow

Optical flow analyzes the pixel to pixel difference of feature points between two consecutive frames caused by the movement. The images are seen as 2D vector field, showing the movement of these points between the first frame and second frame [56]. Lets say a point can be found in pixel $I(x, y, t)$ in the first frame. The points moves a certain distance, which can be represented as (dx, dy). The distance taken between the two frames takes up time which represented by dt.

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{4.41}$$

We remove terms and divide by dt to get:

$$\frac{dI}{dx}\delta x + \frac{dI}{dy}\delta y + \frac{dI}{dt}\delta t = 0 \tag{4.42}$$

On dividing by δt, I obtain the equation for *Optical Flow*, :

$$\frac{dI}{dx}u + \frac{dI}{dy}v + \frac{dI}{dt} = 0 \tag{4.43}$$

where, $u = dx/dt$ and $v = dy/dt$. The directional changes found were dI/dx, from horizontial axis; dI/dx, from the vertical axis; and dI/dt is along the time. That leaves two unknown variables magnitude and direction that require different methods to acquire [26].

4.2.2.2.1 Farneback

The first approach to extracting the magnitude and direction for each frame from individual participant videos was made centered around the dense optical flow method. The dense optical flow method uses all features found in the per-frame and detects pixel intensity changes between these frames. The dense optical flow method is calculated using Gunnar Farneback's algorithm, which is a function implemented in the *OpenCV* library [23]. The calculation of the Gunnar Farneback algorithm was done by an *OpenCV* function called *calcOpticalFlowFarneback*.

The method would take in the previous and current frame and output a 2-channel array with optical flow vectors. These vectors would be fed into another function called the *cartToPolar*, which read and converted the vectors into magnitude and direction between the two frames. Figure 4.2 shows three participants' first five frames of their meeting video going through the Farneback algorthim. This method was not used in the final product because of errors found in the videos. These errors ranged from too much noise found in the video to the participants being removed from the frame of the camera. I was unable to use the data collected due to these errors, and the dense method only producing a singular magnitude and direction per frame.

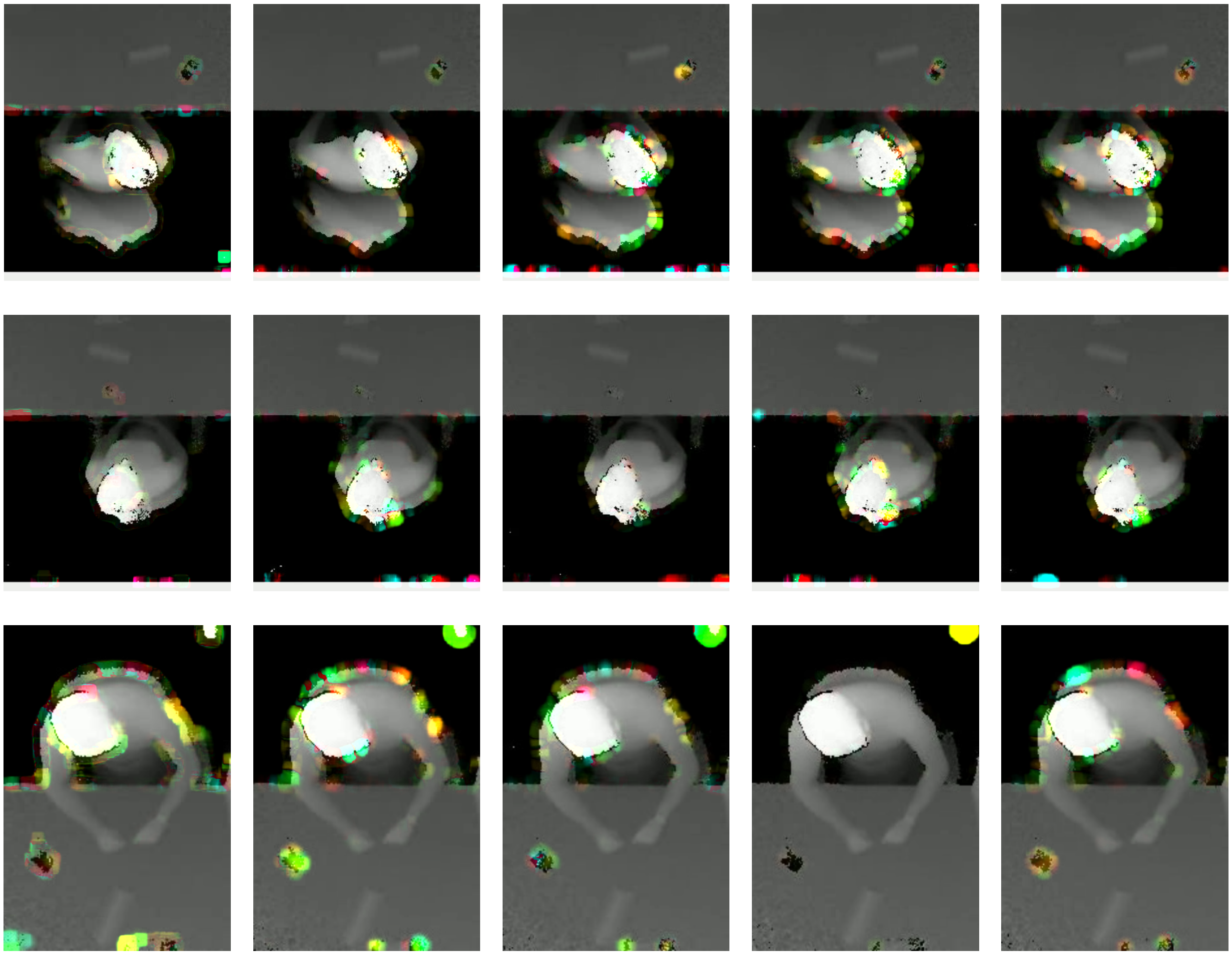


Figure 4.2: Example of the Farneback optical flow process. The above images shows three different participants and their first five frame going through the algorithm.

4.2.2.2.2 Lucas Kande Optical Flow Method

The final approach for extracting the magnitude and direction for each frame from the individual participant's videos was the sparse optical flow method. The sparse optical flow method, unlike the dense optical flow method, does not use all the features from the frame, however, only a select few of the best features are found on the frame. The selection of these features uses the

Shi-Tomasi Corner Detection formula. The Shi-Tomasi Corner Detection algorithm is:

$$f(X,Y) = \sum (I(Xk,Yk) - I(Xk + \Delta X, Yk + \Delta Y))^2 \quad \text{where} \quad (Xk,Yk)\,\epsilon\, W \tag{4.44}$$

where for a given window (W) at position (X,Y) with $I(X,Y)$ pixel intensity.

Calculation of $f(X,Y)$ will be lengthy. The Taylor expansion is then used on the function simplify it to,

$$R = min(\lambda 1, \lambda 2) \tag{4.45}$$

where $\lambda 1$, $\lambda 2$ are eigenvalues of the resultant matrix. This process is taken care of by an *OpenCV* function called *goodFeaturesToTrack* which takes in the frame, the number of corners wanted, quality level, and distance apart. As a result, the function output is the optimum number of corners. These points are selected from the first frame and do not change, allowing the movement to be properly tracked throughout the video [56].

The second part of the sparse optical flow method is calculating the magnitude and direction of the selected points. The sparse optical flow method utilized is the Lucas-Kanade Equation:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i {f_{x_i}}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i {f_{y_i}}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{bmatrix} \tag{4.46}$$

From receiving the selected points from the *goodFeaturesToTrack* function, another function from the *OpenCV* library called the *calcOpticalFlowPyrLK* function runs the equation for us. The function requires the previous frame, the current frame, and the points selected from the first frame; with the output being the position of the new points. From collecting the new position and having the position from the previous points you can use a standard distance formula,

$$AB = \sqrt{(x_2 - x_1)^2 \ + \ (y_2 - y_1)^2} \tag{4.47}$$

to calculate magnitude as well as,

$$\tan\theta = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{where } (x_1, y_1) \text{ is the initial point and } (x_2, y_2) \text{ is the terminal point.} \tag{4.48}$$

to calculate direction. This process is done for every selected point for every frame in each individual participant's video. Figure 4.3 shows three participants' first five frame of their meeting video going through the Lucas Kande method. All participants receive a CSV file containing the magnitude of points for all frames present in their video [56].

4.2.3 Error Removal

The data received from extracting features from the participant's video was affected by the multiple errors found in the videos. This included a selected point that would stop tracking movement and freeze in place. This error was generally caused by participants moving out of the frame of the video, causing the selected points to be stuck on the edges or disappear. In the data collected, a point that would freeze would not produce any more outputs and would default to 0 from the frame it froze to the final frame. Those selected points that froze would be sifted out of the data to not affect the overall.

Another error found is caused by a sudden movement from the participant, this created a situation in which the selected point moved sporadically before readjusting either to the correct next position or an entirely new position. These selected points would not freeze and continue to produce information after. In the data, these sporadic movements would be shown as numbers

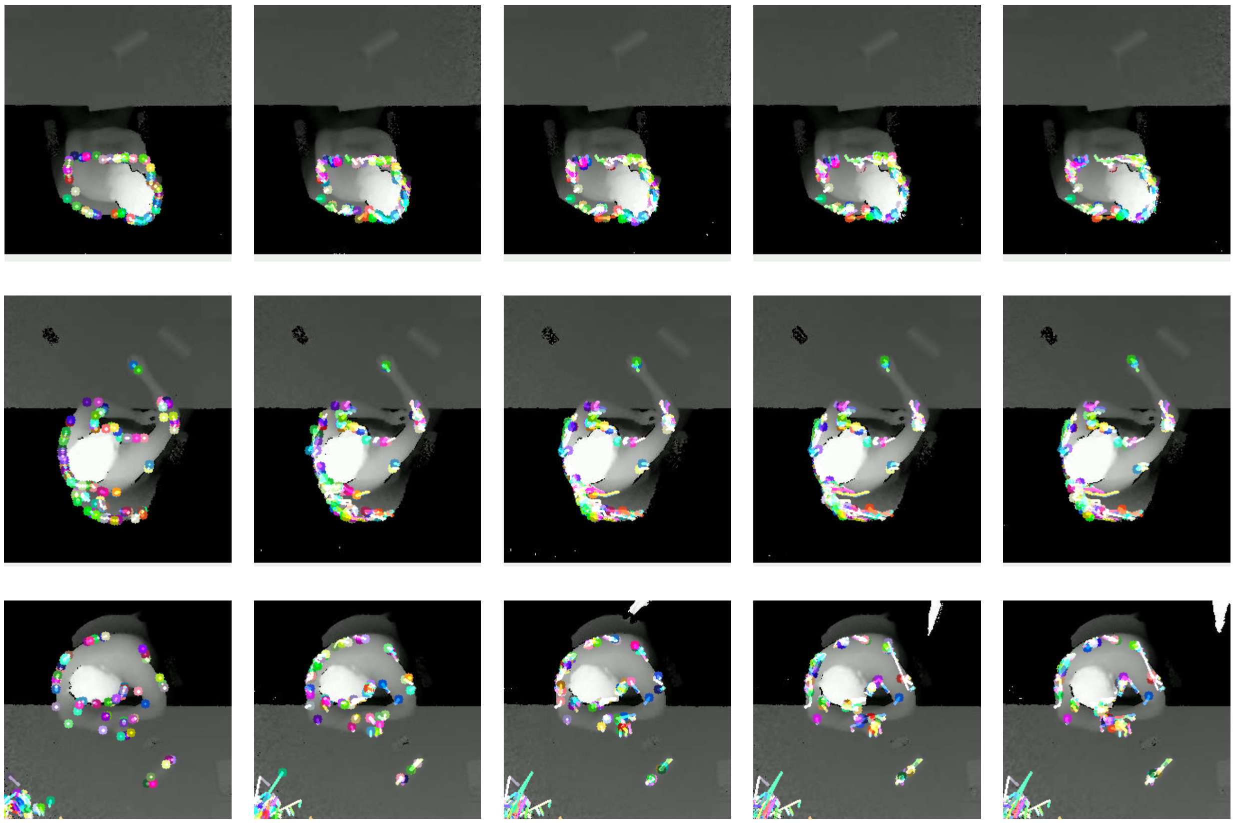


Figure 4.3: Example of the process for the optical flow process. The above images shows three different participants and their first five frame going through the algorithm.

greater than the average reading gathered from other points. To solve this issue, I stated that any movement past a value of 3.8 would be considered a sporadic movement not associated with the participant's movement. The selected point for that singular frame would be ignored, however not discarded, as it could be possibly pertinent information after the incident. I then organized the data by using the min-max normalization method,

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{4.49}$$

where x is the original value and x' is the normalized value. The new range of all the data would be $[0, 1]$ [38]. The error found before I used a global minimum, which would be 0, and a global maximum, which would be 3.8. This allowed all further data to be organized under the same range.

4.2.4 Filtering Out

After removing the error and normalizing the data, I then decreased the noise found in the data. The decrease in the noise will allow us to interpret the movement in the graph more easily. The data is run through two different noise filters, the first is a smoothing function that uses convolution. The convolution is the integral that expresses the amount of overlap of one function v as it is shifted over another function a. It, therefore, "blends" one function with another.

$$(a \cdot v)[n] = \sum_{m=-\infty}^{\infty} a[m]v[n-m] \tag{4.50}$$

where n is the variable. Once the data is run through this filter, I tried another method for filtering the data. This method was the low-pass Butterworth filter.

$$B_n(s) = \prod_{k=1}^{\frac{n}{2}} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\pi\right) + 1 \right] \quad n = \; even \tag{4.51}$$

$$B_n(s) = (s+1) \prod_{k=1}^{\frac{n-1}{2}} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\pi\right) + 1 \right] \quad n = \; odd \tag{4.52}$$

where n is the filter order. A low-pass filter is a filter that contains a selected cutoff frequency that limits signals based on whether their frequency is lower than the cutoff. This filter was implemented using a python library called *Scipy*. The parameter of this function was to choose a filter order cutoff frequency and a bandpass type. I set the filter order to be 2, the cutoff frequency to be $.02$, and the bandpass type to be a low-pass filter.

4.2.5 Calculating Statistics

The cleaned data was statistically analyzed to check for any other possible factors relating to movement found in the video data. Analysis ran on the data was the calculation of the mean and variance of the data.

$$\text{Mean} \qquad \mu = \frac{\sum_{i=1}^{N} X_i}{N} \tag{4.53}$$

$$\text{Variance} \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N} \tag{4.54}$$

where N is the number of frames. Besides the overall mean and variance, I also separate each participant's data into halves and quarters. With those halves and quarters, I calculated the mean and variance for the data between those new time frames. Additionally, I calculated the mean and variance for each of the 22 groups.

4.3 Communication Analysis from the UGI Transcripts

Additionally, the transcript data provided by the UGI corpus was used to determine communication patterns. The use of transcript data allowed the extracting of new factors which would be used to find comparison and/or correlation between any other factors found in previous mention analyses. The UGI corpus provides transcripts for each of the 22 groups, those transcripts hold the time code, which participants in the group spoke, as well as what they said. For our analysis, code was created that allowed the separation of each participant's transcript information into a separate file. This gave us the opportunity to run each of the following tests on each of the participant's transcript data without fear of interference from another.

4.3.1 Talking Frequency

Talking Frequency, or *Speaking Turns*, is the total time a participant spoke during the small group meeting. Each participant has a separate transcript that contains the time codes, their number in the group, and the words they spoke. The number of speaking turns was calculated by counting the number of time codes found in their transcript. In addition to calculating *Speaking Turns*, I also calculated each participant's *Speaking Turns* percentage within their groups. I found this number by adding up each individual participant's *Speaking Turns* and dividing each participant's *Speaking Turns* number by their group's total *Speaking Turns*. Finally, the original *Speaking Turns* data in a single range was normalized data to fit in the range $[0, 1]$, using the min-max method.

4.3.2 Word Frequency

The transcript data contains the words spoken by the participant. While calculating for *Talking Frequency*, I came across a number of time codes with a low count number of words. To reinforce or locate any contrast between *Talking Frequency* and *Word Frequency*. The *Word Frequency*, or *Utterances*, is a spoken word, statement, or vocal sound created by the participants. To find the number of *Utterances* for each participant, I read the participant's speaking turn and separated them into individual words, and counted the words found. Similar to *Speaking Turns*, I also calculated the percentage of *Utterances* for each participant in their own groups. To calculate the percentage of words spoken, I added all *Utterances* from each participant in their group and divided the individual's *Utterances* by their group's *Utterances*. Finally, the original *Utterances* count is also organized using the min-max method, and normalized the data to fit the range $[0, 1]$.

Table 4.2: Keywords from the *Lunar Survival Task*.

Keywords
Matches
Nylon
Rope
Parachute
Silk
Heating
Pistol
Milk
Oxygen
Stellar
Map
Raft
Compass
Water
Flares
Aid
Kit
Needle
FM

4.3.3 Keyword Frequency

To check for topic-related words, a set of keywords was constructed that consisted of names of items found in the *Lunar Survival Task*. These keywords can be seen in Table 4.2. To check if the keyword appeared in any of the participants' *Speaking Turns*, I read the words found when searching for *Word Frequency* and I filtered through them only counting when I found a keyword. Once the number of keywords spoken by each participant was determined, I then calculated the percentage of keywords spoken by each participant per their group. The percentage was calculated by adding up each group's keyword count and dividing the participant's keyword count by the group's overall keyword count. Similar to the other frequency checks, this data was also organized using the min-max method and normalizing the data to fit the range $[0, 1]$.

4.3.4 Sentiment and Emotion

This analysis finds a way to track the participant's attitudes and emotions while going through the meeting task. *Sentiment analysis* is the process of identifying whether the speaker's emotion fall closer to positive, negative, or neutral. This is done by categorizing the speaker's opinions found in text written by the speaker. To run this analysis, I run three different sentiment analysis libraries to determine the participant's attitude.

Two of the libraries are rule-based sentiment analysis which means this approach is without training and uses machine learning on the model. The model has a set of rules based on predetermined labeling on whether the word is positive, negative, or neutral [9]. The libraries that use this approach are *Textblob* and *VADER*. *Textblob* reads in the words and outputs the polarity and the subjectivity. The polarity is in the range $[-1, 1]$ where -1 is negative sentiment and 1 is positive sentiment and subjectivity is in the range $[0, 1]$ checking how much personal opinion and emotion are found. *Textblob* also had a second function I used that would output the positive and negative polarity numbers. These two numbers would be in a range between $[0, 1]$ where the addition of each sentiment would equal 1. With each of the polarities, there would be a classification to determine if those were positive or negative in nature. *VADER* library works similarly, where it reads in the word and outputs the percentage for how positive, negative, or neutral the words are. These three probabilities will add up to 100%, and the probability is a final output called a compound. The compound is the sum of each polarity score for every individual word, then normalized to the range $[-1, 1]$ where -1 is negative and 1 is positive. The final library uses an embedding based model for its sentiment analysis. Text embedding is a type of word representation that pair different words with similar meaning to the same representation. This allows words to be judged depending on their similarity to other positive, negative, or neutral words. The library that uses this model is the Flair library. Flair is a pre-trained sentiment analysis model that reads in the words and outputs whether the words were positive or negative, as well as a score in the range $[.5, 1]$.

With the sentiment analysis libraries, I also added an emotion analysis library. The library was called *text2emotion*, which is a library that finds the emotion embedded in the text provided. The library outputs the score in 5 relative emotional categories, including anger, fear, happiness, sadness, and surprise. The scores for all the emotions range $[0, 1]$ with the total of the five scores equaling 1. These scores represent how much of the five emotions influence the text.

CHAPTER 5: RESULTS

This chapter summarizes the results of this thesis's main study. I first look at the team performance from the score metrics that I calculated, then look at the results of our motion analysis and transcript communication analysis. Lastly, I show some potential correlations between these metrics that show areas for future investigation.

5.1 Results on Team Performance Analysis from Scores

5.1.1 Results of Individual Scores

Table 5.5 shows the results of all four different tests taken to measure individual participants' scores against the Expert Rubric score. Shown in the first three columns of the table is information related to properly identifying the participant. This includes the subject's number, which team they were in, and the size of their group. The following column, S_1^I, displays the difference between the participant's individual ranking and the Expert Rubric ranking. The scores gathered from the participant were put on a list to determine whether that participant is a high-performing or low-performing individual. The threshold between high and low participants or teams came directly from the *Lunar Survival Task* rank groupings and was set as a score of 45 is the highest score a participant/team can have before falling below average. Those high-performing individuals gained a score of 45 or lower and were highlighted in blue on the table, and any number higher than 45 was a low-performing individual and was highlighted in red on the table. The next column, S_2^I, is the average difference between ranks against the two previously mentioned rankings. S_3^I is the same as S_1^I except instead of using the expert ranking in the difference calculation this score uses the reverse ordered expert ranking. S_4^I is the average difference between ranks

Table 5.1: The table shows an example of the basic information about each participant including both their individual and group ID number as well as their groups size. Showing the results from four different Individual Ranking vs Expert Ranking difference calculation and the results three different Individual Ranking vs. Expert Ranking RBO calculation.

ID (1-86)	ID (1-22)	Size	S_1^I	S_2^I	S_3^I	S_4^I	(RBO)	(RBO-0.9)	(RBO-0.8)
1	1	4	22.000	1.467	84.000	5.600	0.898	0.710	0.866
2	1	4	56.000	3.733	86.000	5.733	0.656	0.448	0.460
3	1	4	38.000	2.533	90.000	6.000	0.818	0.639	0.777
4	1	4	54.000	3.600	72.000	4.800	0.769	0.616	0.775
5	2	4	50.000	3.333	80.000	5.333	0.789	0.633	0.795
6	2	4	66.000	4.400	70.000	4.667	0.638	0.450	0.478
7	2	4	54.000	3.600	88.000	5.867	0.747	0.581	0.708
8	2	4	50.000	3.333	86.000	5.733	0.677	0.459	0.458
9	3	4	48.000	3.200	90.000	6.000			
10	3	4	40.000	2.667	80.000	5.333	0.850	0.691	0.869
11	3	4	44.000	2.933	82.000	5.467	0.806	0.640	0.798
12	3	4	58.000	3.867	78.000	5.200	0.711	0.540	0.645
13	4	4	50.000	3.333	66.000	4.400			
14	4	4	60.000	4.000	86.000	5.733	0.681	0.506	0.603
15	4	4	48.000	3.200	82.000	5.467	0.753	0.573	0.685
16	4	4	48.000	3.200	82.000	5.467	0.786	0.624	0.779
17	5	4	56.000	3.733	66.000	4.400	0.742	0.585	0.734
18	5	4	42.000	2.800	80.000	5.333	0.835	0.675	0.845
19	5	4	38.000	2.533	84.000	5.600	0.819	0.637	0.764
20	5	4	36.000	2.400	68.000	4.533	0.840	0.665	0.821
21	6	5	24.000	1.600	84.000	5.600	0.884	0.698	0.860
22	6	5	56.000	3.733	70.000	4.667	0.755	0.603	0.761
23	6	5	60.000	4.000	58.000	3.867	0.743	0.596	0.756
24	6	5	44.000	2.933	72.000	4.800			
25	6	5	66.000	4.400	64.000	4.267	0.609	0.408	0.409
26	7	3	30.000	2.000	76.000	5.067	0.873	0.699	0.868
27	7	3	48.000	3.200	70.000	4.667	0.793	0.633	0.794
28	7	3	48.000	3.200	76.000	5.067	0.783	0.620	0.774

against the team ranking and the reverse ordered expert ranking. The last three columns demonstrate the usage of the RBO method. The RBO method read in the individual and expert list and presented a number between 0, which has no similarities, and 1, which is for complete similarity. The default RBO contributes no weight, the RBO method returns the average overlap. The other two contain different RBO weights; the RBO-0.9 gave a higher weight to the top 13.5 items in the list and the RBO-0.8 gave a higher weight to the top 12 items in the list. In these columns some participants have no score, those with no score indicate that the participant had the same ranking for two different items so a proper score could not be issued.

5.1.2 Results of Team Scores

Table 5.2 shows the results of all the different tests taken to measure the group's scores against the Expert Rubric score. Shown in the first column of the table is the team identification number. The next column shows the number of disagreements found in each group. The following column, S_1^T, displays the difference between the team's group ranking and the Expert Rubric ranking. The scores gathered from the different groups were put on a list to determine whether that team is a high-performing or low-performing team. Those high-performing teams gained a score of 45 or lower and were highlighted in blue on the table, and any number higher than 45 was a low-performing team and was highlighted in red on the table. Besides the column with the difference score, S_2^T presents the average difference between the two different rankings. S_3^T is the same as S_1^T except the order for expert ranking was reversed when calculating the difference and S_4^T is the average difference between those rankings. S_5^T is the difference between the expert ranking and team ranking excluding the disagreement found in the team ranking, those expert ranking items that corresponded with the team ranking disagreement were ignored for this calculation. S_6^T implements the first penalty mentioned before where I added 15 onto the score for the number of non-consensus the team suffered during the meeting. S_7^T followed a similar process except the expert ranking was reversed when doing the calculation and S_8^T is the penalty added to the reverse ordered score.

S_9^T was another penalty that added the score without null values, which was S_5^T, to the average score between the ranking that consisted solely of the disagreements, S^T_{miss}. S^T_{10} is follows the same formula with the score that contained the reversed ordered expert ranking instead of the original expert ranking. The score that contains the score without the null values for S^T_{10} was S^T_7 and the average score with just the disagreement was $S^T_{\overline{miss}}$. S^T_9 and S^T_{10} contain empty spaces because of S^T_{miss} and $S^T_{\overline{miss}}$ respectively. The equation for all four of these scores contain division by the number of disagreement and some of the team rankings do not contain disagreement which creates a division by zero scenarios in the equation. This means all empty space regarding these scores signify that those team had no disagreement and came to a consensus which meant no penalty in the score. The final penalty score, S^T_{11}, replaced the disagreements in the team ranking with 15 before calculating the difference, then multiplying with their reverse ordered ranking. The last four columns dealt with the accuracy of the team ranking against the expert ranking. S^T_{ans} shows the number of exact matches between the team's ranking and the expert's ranking and S^T_{alm} shows the number of matches between the team's ranking and the expert's ranking with an error margin of ± 1. S^T_{miss} and $S^T_{\overline{miss}}$ mentioned before calculating the average score for solely the disagreement of the team ranking against their respective expert ranking or reverse ordered expert ranking.

5.1.3 Result of Individual Score vs Team Score

Table 5.3 shows the results of all the different tests taken to measure the individual participant's score and their group's score. Shown in the first two columns of the table are the participant's subject and team identification number. The following two column displays the difference between the individual participant's ranking and the team's group ranking this score is $\boldsymbol{S_1}^C$. The second column is the average ranking error between the fifteen items this score being under $\boldsymbol{S_2}^C$. The next two column shows the difference between the individual's original score and the team's group's original score. This score, unlike the other, presented the score either with a positive or negative number. This is to show when the participant went to the team portion of the task, their list

Table 5.2: The table shows the basic information about each participant including their group ID number and number of non-consensuses. This also shows the results from the five different Team Ranking vs Expert Ranking difference calculation, the results from five different Team Ranking vs Expert Ranking penalty calculation, and the results of four different Team Ranking vs Expert Ranking accuracy calculation.

ID (1-22)	n^-	$(S_1)^T$	$(S_2)^T$	$(S_3)^T$	$(S_4)^T$	$(S_5)^T$	$(S_6)^T$	$(S_7)^T$	$(S_8)^T$	$(S_9)^T$	$(S_{10})^T$	$(S_{11})^T$	$(S_{ans})^T$	$(S_{alm})^T$	$(S_{miss})^T$	$(S_{\overline{miss}})^T$
1	6	69	4.600	83	5.533	19	109	43	133	27.333	49.667	823	2	4	8.333	6.667
2	0	28	1.867	84	5.600	28	28	84	84			156	4	9		
3	0	42	2.800	70	4.667	42	42	70	70			251	2	4		
4	7	94	6.267	92	6.133	21	126	42	147	31.429	49.143	717	1	6	10.429	7.143
5	4	40	2.667	92	6.133	18	78	56	116	23.500	65.000	718	2	7	5.500	9.000
6	0	46	3.067	74	4.933	46	46	74	74			238	3	6		
7	0	48	3.200	74	4.933	48	48	74	74			345	3	6		
8	5	65	4.333	91	6.067	24	99	55	130	32.200	62.200	746	4	5	8.200	7.200
9	3	44	2.933	94	6.267	19	64	67	112	27.333	76.000	442	3	8	8.333	9.000
10	0	28	1.867	78	5.200	28	28	78	78			195	4	8		
11	0	48	3.200	76	5.067	48	48	76	76			288	2	5		
12	0	48	3.200	84	5.600	48	48	84	84			284	2	6		
13	0	36	2.400	76	5.067	36	36	76	76			223	3	6		
14	0	34	2.267	86	5.733	34	34	86	86			178	3	10		
15	0	24	1.600	78	5.200	24	24	78	78			161	4	11		
16	2	45	3.000	71	4.733	25	55	66	96	35.000	68.500	329	4	7	10.000	2.500
17	2	51	3.400	83	5.533	30	60	67	97	40.500	75.000	335	3	7	10.500	8.000
18	0	34	2.267	70	4.667	34	34	70	70			238	3	8		
19	3	62	4.133	94	6.267	22	67	67	112	35.333	76.000	275	3	5	13.333	9.000
20	0	40	2.667	82	5.467	40	40	82	82			259	4	7		
21	6	69	4.600	87	5.800	17	107	51	141	25.667	57.000	762	4	5	8.667	6.000
22	7	94	6.267	88	5.867	31	136	43	148	40.000	49.429	885	2	3	9.000	6.429

ranking worsened as a result. This score will be under $\boldsymbol{S_3}^C$ with the second column displaying the average ranking error between the fifteen items under $\boldsymbol{S_4}^C$. Those with positive scores in this column had their ranked list improved throughout the study. The following two column displays the difference between the individual participant's ranking and the team's group ranking not including the disagreement. The difference is under $\boldsymbol{S_5}^C$, the second of these two columns demonstrates the average ranking error between the items that were not disagreements under $\boldsymbol{S_6}^C$. Score $\boldsymbol{S_5}^C$ would serve as the base score for the next three scores which show the three penalties that were calculated for these rankings. Score $\boldsymbol{S_7}^C$, $\boldsymbol{S_8}^C$, and $\boldsymbol{S_9}^C$ are three different penalty score described before. $\boldsymbol{S_7}^C$ is the penalty score adding 15 points for each disagreement to the base score. $\boldsymbol{S_8}^C$ is the penalty score that adds the average missed ranking to the base score. The final penalty score does not use the base score but finds the difference between individual ranking and their group's ranking and multiplies the difference by the inverse ranking of the specific item. The change is that if a null value is found in the team ranking then the difference between the two rankings will equal 15. Once all items from the ranking go through this process then each item is added together to give the score $\boldsymbol{S_9}^C$. The final score shows the similarities between the two rankings; the first score $\boldsymbol{S_{ans}}^C$ shows the number of exact rankings between the two rankings and the last score $\boldsymbol{S_{alm}}^C$ shows the number of rankings that in the range $[-1, 1]$.

5.1.4 Correlation Results

With the team performance scores calculated I investigated which of the different variables from the movement analysis and text analysis have an effect on the outcome of the individual team performance scores. Our investigation will consist of locating any correlation between all scores and variables calculated through the experiment, more in-depth analysis and display of the correlation found, as well as whether the relationship found (if any) between the score and variable could be statistically relevant. While the investigation did produce some results it must be understood that this paper is exploratory in nature signifying that less intensive restrictions were placed on the scores when searching for positive correlations. If I were not simply looking for correlations for the smart room then more rigorous implantation, such as the Bonferroni correction, would have been used.

We did not use the Bonferroni correction because of the exploratory nature of the thesis. Therefore to find the correlation coefficient for the 86 participants in the dataset, I used a table found on the website "Statistics Solution" [70]. In this table gives an estimate for the correlation coefficients. I did not find an exact correlation coefficient for 86 participants, but the table provided correlation coefficients for 80 participants which was $.283$ and the coefficient for 90 participants which was $.267$. For our 86 participants, I choose our correlation coefficient by finding the average

Table 5.3: The table shows an example of the basic information about each participant including their individual and group ID numbers. Showing the results from six different Individual Ranking vs Team Ranking difference calculation, the results from three different Individual Ranking vs Team Ranking penalty calculation, and the results of two different Individual Ranking vs Team Ranking accuracy calculation.

SID(1-86)	T(1-22)	${S_1}^C$	${S_2}^C$	${S_3}^C$	${S_4}^C$	${S_5}^C$	${S_6}^C$	${S_7}^C$	${S_8}^C$	${S_9}^C$	${S_{ans}}^C$	${S_{alm}}^C$
1	1	65	4.333	-47	-3.133	18	2	108	25.833	811	1	6
2	1	71	4.733	-13	-0.867	16	1.778	106	25.167	785	2	7
3	1	77	5.133	-31	-2.067	20	2.222	110	29.5	788	2	4
4	1	77	5.133	-15	-1	29	3.222	119	37	879	2	4
5	2	28	1.867	22	1.467	28	1.867	28		184	7	9
6	2	54	3.6	38	2.533	54	3.6	54		391	0	5
7	2	40	2.667	26	1.733	40	2.667	40		283	3	6
8	2	36	2.4	22	1.467	36	2.4	36		290	2	6
9	3	34	2.267	6	0.4	34	2.267	34		243	3	6
10	3	24	1.6	-2	-0.133	24	1.6	24		183	6	8
11	3	44	2.933	2	0.133	44	2.933	44		284	3	7
12	3	38	2.533	16	1.067	38	2.533	38		307	3	8
13	4	76	5.067	-44	-2.933	20	2.5	125	28	753	2	6
14	4	88	5.867	-34	-2.267	13	1.625	118	23.714	716	2	6
15	4	100	6.667	-46	-3.067	26	3.25	131	36.571	802	2	3
16	4	94	6.267	-46	-3.067	12	1.5	117	23.714	694	1	3
17	5	68	4.533	16	1.067	39	3.545	99	46.25	909	1	5
18	5	52	3.467	2	0.133	26	2.364	86	32.5	751	2	5
19	5	42	2.8	-2	-0.133	18	1.636	78	24	722	3	5
20	5	52	3.467	-4	-0.267	31	2.818	91	36.25	820	1	5
21	6	44	2.933	-22	-1.467	44	2.933	44		250	3	4
22	6	24	1.6	10	0.667	24	1.6	24		173	4	8
23	6	20	1.333	14	0.933	20	1.333	20		158	8	11
24	6	38	2.533	-2	-0.133	38	2.533	38		198	3	5
25	6	48	3.2	20	1.333	48	3.2	48		369	3	7
26	7	44	2.933	-18	-1.2	44	2.933	44		349	3	5
27	7	14	0.933	0	0	14	0.933	14		82	5	12
28	7	28	1.867	0	0	28	1.867	28		213	5	6

between the 80 participants coefficient and the 90 participant coefficient which was calculated to be $.27$. Knowing our correlation coefficient, I then calculated all the coefficients for each score and variable. Figure 5.1 shows a heatmap of all the correlations found for each score or variable against the others in the data. The data in the figure is broken into three categories. Those categories are the scores found from the individual team performance data, the movement data found from the videos, and the text data found from the transcripts. Because of these groupings, there are sections of the correlation found that will be ignored because the focus is on the scores being affected by either movement or text data.

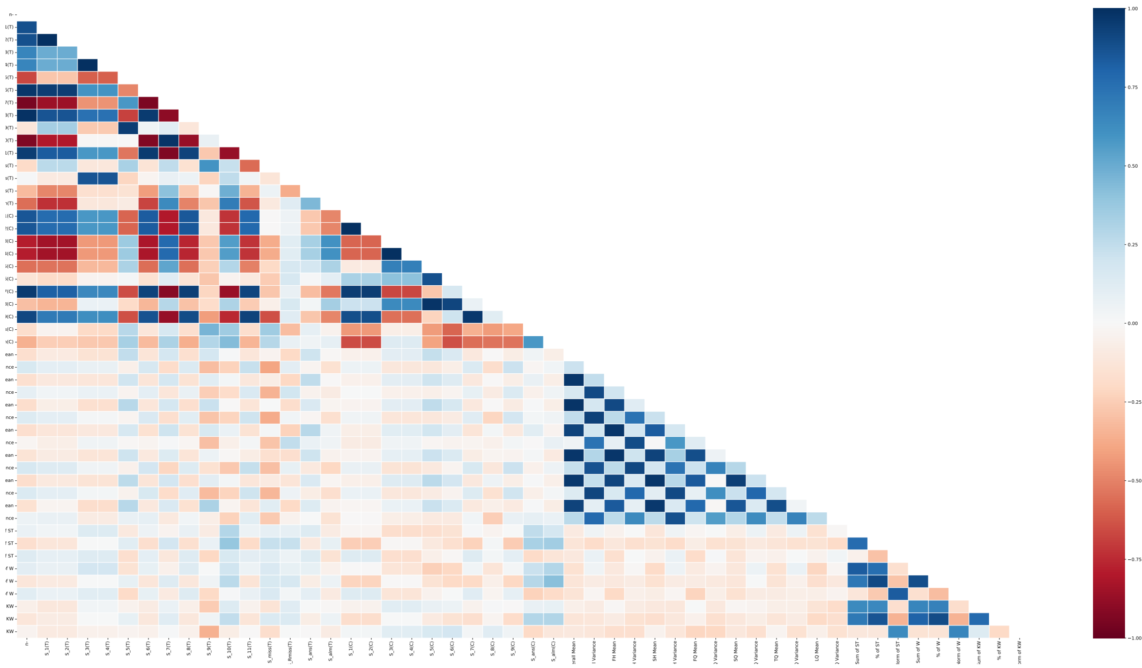


Figure 5.1: The figure above shows a heatmap displaying the correlations found between the performance scores, movement data, and text, data. The scale is $[-1, 1]$ where the close to -1 you get the more blue the square becomes ad the closer to 1 you get the more red the square becomes.

I also displayed the data with correlation plots. Figure 5.2 is an example of the correlation plot to clarify the analysis between the scores and the data. Each plot contains a single score with both movement and text data also the figure is in shorthand to be able to visibly read some of the variables. The movement data will have one data being a mean of the overall, half, quartered movement and the other data being a variance of the overall, half, quartered movement. The text data will have three different calculations from the groupings of *Speaking Turns* (ST), *Words* (W), or *Keywords* (KW); each of the groupings will be measured by finding the total number (Sum), the percentage (%), and the normalization (Norm) of each of the groupings. As you can see the plot is six-by-six containing the correlation coefficient of each pair, a scatter plot showing the relationship of each pair, and a bar graph showing the distribution for each individual score or data. The focus of this plot is mainly in the top row which gives the coefficient of the score against each data type and the first column which determines the type of relationship the score has with that data type. Each score was checked for correlation and out of the 27 scores calculated only 7 of them were found with at least 1 correlation coefficient that was above the threshold ($coef \geq 0.27$).

Figure 5.3(A) and Figure 5.3(B) are two graphs that demonstrate which score found a correlation, showing the number of correlations found, as well as which 15 data types correlated with each score that contains a correlation. The x-axis provides the number of correlations found between a score and the different variables found between the movement and text data. The different variables either from the movement data or text data are represented by a color which can be seen on the legend on the bottom right side of each graph. From Figure 5.3(A) I see S_5^T, S_9^T, S_{10}^T, S_{miss}^T, and S_5^C all found correlation with at least one movement variable with S_{miss}^T having the most correlations. The movement variables are split between mean and variance, and as seen in the graph a majority of the correlation found is with the variance of the movement data. This shows the difference in movement whether positive or negative may have an effect on these scores. For Figure 5.3(B) I see that only four scores, S_9^T, S_{10}^T, S_{ans}^C, and S_{alm}^C, have at least one correlation with S_{alm}^C having the most with four correlations. I see in these correlations a more even spread between the different text variables, specifically between the three groupings of Speaking Turns, Words, and Keywords. Out of the 7 scores, only two of the scores contained correlation with both the movement data and test data; those scores were S_9^T and S_{10}^T.

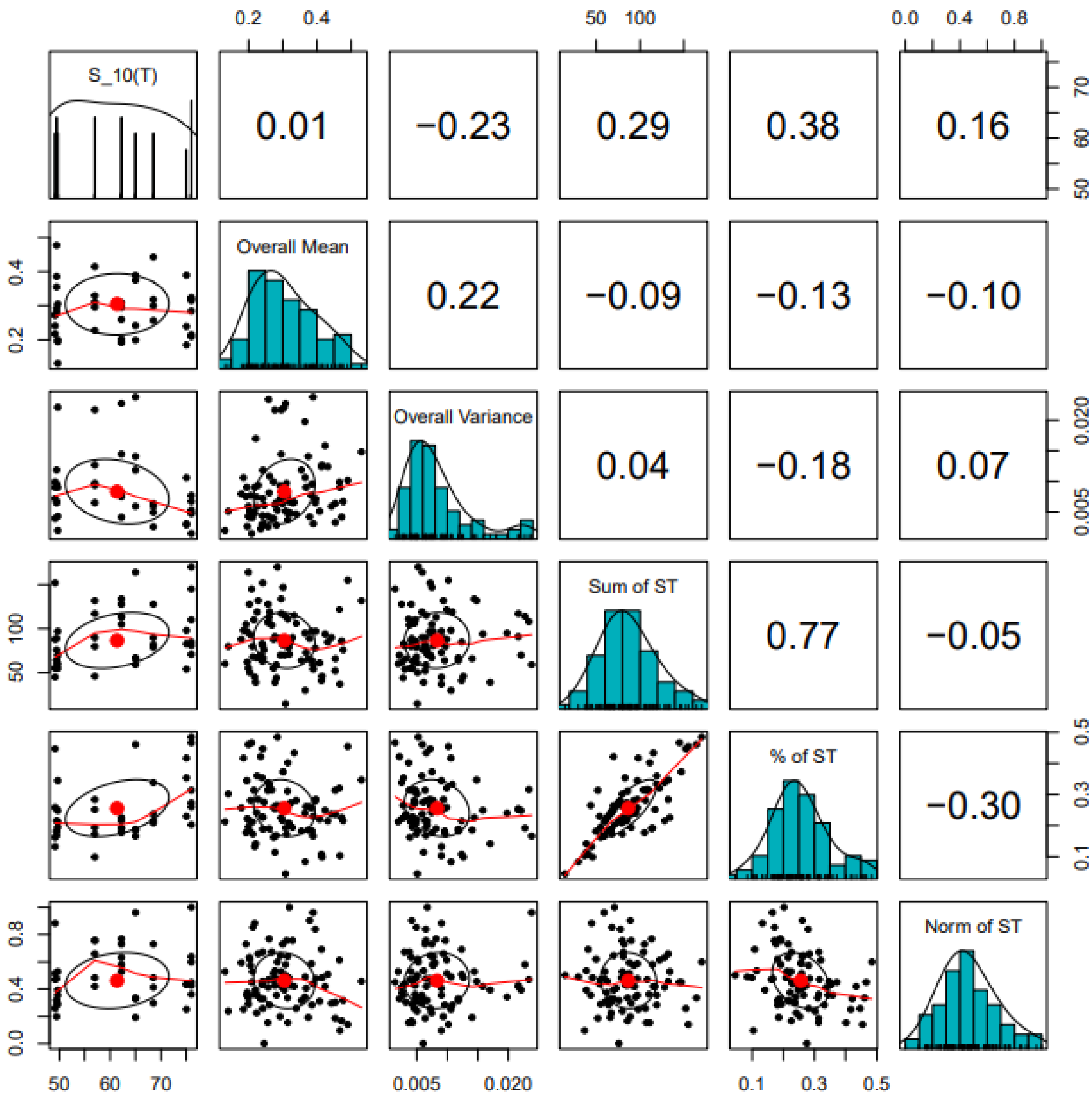


Figure 5.2: The figure above shows an example of the correlation matrix graphs found. This correlation matrix demonstrates the distribution for each variable using the bar graph as well as a graphical representation of the correlation between the two variables. Alongside the two graphs is a number which is the correlation coefficient for each pair.

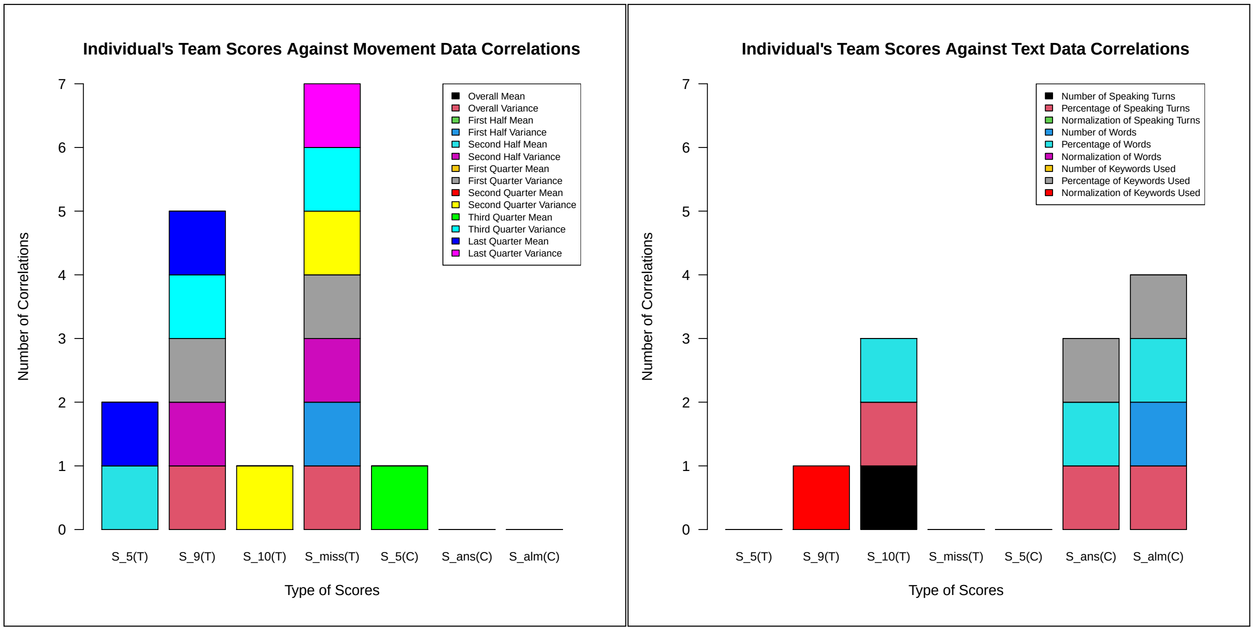


Figure 5.3: The figure above shows two graphs, these graphs demonstrate how many correlations each performance score had if they had any. Each variable added on is color coordinated to also show which variable had the correlation with the specific score. There is a legend on the bottom right of each graph showing each variables' respective color.

5.1.5 Results of Other Quantitative Measures

Table 5.4 shows the result from other measures mentioned. The first two columns display the participants' individual identification numbers and team identification numbers. The following two measures are the averages calculated from all the members of their group's ranking of that participant. The last five columns are the rankings the original document had given each participant, these rankings were not altered.

Table 5.4: The table shows and example of the basic information about each participant including their group ID number and participant number. Showing the results of the other measurements found from the questions. The questions were about the other participant in their group and how they were perceived throughout the meeting.

SID(1-86)	TID(1-22)	Talkative?(1-5)	Group leader?(1-5)	Boring/ Engaging(1-5)	Warm/ Cold(1-5)	Comfortable/ Awkward(1-5)	Interesting/Dull(1-5)	Friendly/ Detached (1-5)
1	1	3	3.25	3	3	5	4	4
2	1	4	3.75	4	3	4	4	4
3	1	3.25	2.75	5	4	4	5	4
4	1	3.5	3	5	4	4	5	5
5	2	4.5	4	5	4	5	5	5
6	2	4	3	4	4	4	4	5
7	2	3.75	2.667	5	5	5	5	5
8	2	4.25	3.25	4	4	4	3	4
9	3	4.25	3.25	5	5	4	5	5
10	3	4	3.5	4	3	5	4	5
11	3	4	3	4	3	4	4	4
12	3	3.667	2.667	5	4	5	5	5
13	4	4.333	3.667	5	5	5	5	5
14	4	4.25	2.5	5	5	5	5	5
15	4	4.5	4.25	5	5	5	5	5
16	4	4.25	3	4	4	4	5	4
17	5	3.667	3	1	2	4	1	1
18	5	3.25	2.25	3	2	2	2	3
19	5	5	4.5	4	3	3	4	2
20	5	2.667	1.5	5	3	4	4	4
21	6	4.6	4.2	5	5		5	5
22	6	3.4	3.2	4	4	5	4	4
23	6	4.2	3.6	5	5	5	5	5
24	6	2	1.6	3	2	2	2	1
25	6	2.6	1.8	4	5	3	5	4
26	7	4	3	4	3	4	3	4
27	7	4	2.667	4	3	4	5	5
28	7	4	2.667	4	4	4	4	5

5.2 Results of Motion Analysis

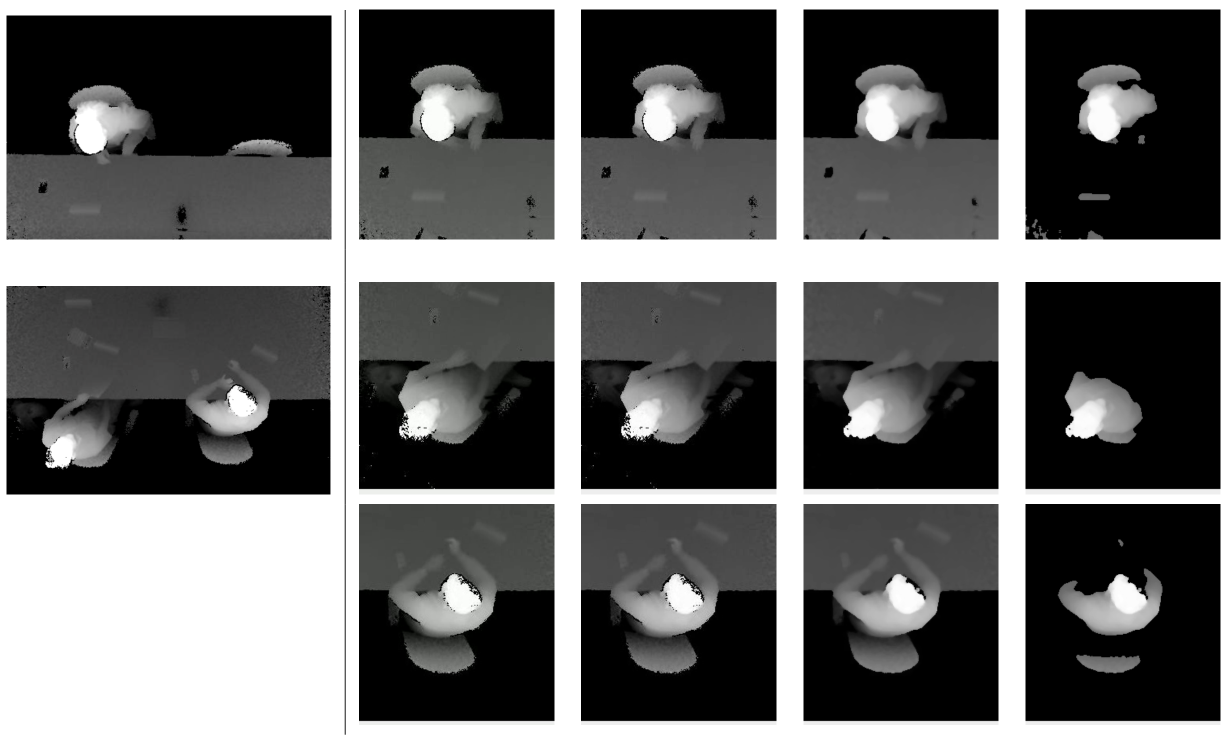
5.2.1 Video Processing

Each group consisted of two videos, containing the record of two to three participants on either side of the table. Figure 5.5 shows two participants from the same side separated by the video separation algorithm; the top set of images is the first participant and the bottom is the second participant. Along with the separation Figure 5.5 shows the two different participants and the stages the video went through when it is being processed. The first image contains the image of the first frame before being processed. Each picture adds an image enhancer, from left to right you can see the image turning into the final product. All 86 participants received individual videos and images of their first frame for all other participant processed images can be found in **Appendix A:**

PROCESSED IMAGES.

Table 5.5: A visualization of the prepossessing approach to **(A)** separates the participants into different videos; **(B)** and cleans the noise.

Original	A	B	C	D



5.2.2 Optical Flow

Appendix B: CONVOLUTION GRAPHS contains the result of Lucas Kanade dense optical flow algorithm ran on the participants' video data. All participants' data tables were set up the same way. Where each of the participants' first column had the recorded frame number. Following the frame number, each participant had 100 selected points, those 100 points had their magnitude tracked and recorded for each of the following columns. This pattern can be seen in these tables where the frames increase so does the number of zeros in place of their magnitude. This is caused by the selected point having an error. This error happens cause of noise or confusion from the movement the participant does. This reason is why I have 100 selected points per participant in case of these errors.

5.2.3 Filtering Out

The table below is from the error removal as well as the filtering of the data received from the Lucas Kanade algorithm. These tables keep frame numbers as indicators of where in the video these movement magnitudes correspond to. The frame numbers are the overall magnitudes of the selected point, that overall magnitude had the errors mentioned before removed only keep the selected point that stayed in play and did not jump over 3 in magnitude. The following two columns contain the magnitude of the two filters mentioned before. The first contains solely the filter containing the convolution filter and the second contains the low band-pass Butterworth filter. All three of the columns containing the different versions of the magnitude were normalized.

5.2.4 Calculating Statistics

Below contains two tables containing the results of the mean and variance calculation done on filtered data. Table 5.7 contains the mean and variance of all the individual participants' and Table 5.6 contains the mean and variance of all 22 groups. Each table has the mean and variance for the overall meeting, the meeting halved, and the meeting quartered. I only used data received from the convolution method to calculate the mean and variance. I came to the decision by looking at the graphical representation of the two sets of data. I saw that both sets of data produce a clear indication of the path the participants' magnitude took over the duration of the meetings. Although the convolution method path contains more character I believe would provide

Table 5.6: The table below shows the groups participant's movement's mean and variance. Each participant contains an ID and movement score separated by overall (O), first half(FH), second half(SH), first quarter(FQ), second quarter(SQ), third quarter(TQ), last quarter(LQ) movement scores.

TID(1-22)	O Mean	O Variance	FH Mean	FH Variance	SH Mean	SH Variance	FQ Mean	FQ Variance	SQ Mean	SQ Variance	TQ Mean	TQ Variance	LQ Mean	LQ Variance
1	0.208	0.009	0.221	0.010	0.194	0.007	0.229	0.008	0.214	0.011	0.184	0.007	0.205	0.007
2	0.340	0.008	0.343	0.009	0.337	0.006	0.359	0.010	0.326	0.007	0.334	0.005	0.339	0.007
3	0.328	0.011	0.326	0.009	0.329	0.013	0.317	0.007	0.335	0.010	0.327	0.012	0.332	0.015
4	0.251	0.008	0.239	0.007	0.264	0.008	0.236	0.005	0.243	0.007	0.250	0.007	0.277	0.007
5	0.302	0.014	0.298	0.012	0.305	0.014	0.269	0.009	0.327	0.011	0.344	0.013	0.266	0.010
6	0.403	0.014	0.397	0.013	0.410	0.014	0.380	0.013	0.413	0.010	0.409	0.013	0.410	0.015
7	0.422	0.007	0.398	0.006	0.445	0.007	0.379	0.007	0.418	0.005	0.432	0.005	0.459	0.008
8	0.315	0.014	0.337	0.014	0.292	0.012	0.350	0.013	0.325	0.014	0.292	0.012	0.292	0.012
9	0.282	0.005	0.285	0.005	0.279	0.005	0.293	0.005	0.277	0.003	0.264	0.003	0.295	0.005
10	0.354	0.005	0.352	0.005	0.356	0.005	0.341	0.004	0.362	0.005	0.353	0.004	0.358	0.005
11	0.323	0.004	0.320	0.004	0.326	0.003	0.328	0.004	0.313	0.004	0.324	0.002	0.329	0.003
12	0.256	0.006	0.241	0.004	0.271	0.007	0.233	0.003	0.248	0.005	0.279	0.007	0.262	0.005
13	0.252	0.006	0.251	0.005	0.253	0.006	0.256	0.005	0.246	0.005	0.255	0.006	0.251	0.007
14	0.246	0.008	0.255	0.008	0.238	0.008	0.273	0.009	0.237	0.005	0.232	0.007	0.243	0.008
15	0.297	0.005	0.299	0.005	0.294	0.006	0.309	0.005	0.289	0.004	0.284	0.005	0.304	0.007
16	0.329	0.005	0.330	0.004	0.328	0.005	0.338	0.003	0.323	0.005	0.327	0.004	0.328	0.006
17	0.272	0.006	0.262	0.004	0.281	0.006	0.256	0.003	0.269	0.005	0.265	0.003	0.297	0.008
18	0.243	0.010	0.234	0.010	0.252	0.011	0.229	0.010	0.238	0.009	0.246	0.010	0.257	0.011
19	0.273	0.006	0.270	0.006	0.276	0.005	0.284	0.007	0.256	0.005	0.275	0.005	0.277	0.004
20	0.470	0.010	0.476	0.009	0.465	0.011	0.469	0.008	0.482	0.009	0.491	0.007	0.439	0.013
21	0.395	0.015	0.399	0.009	0.391	0.018	0.381	0.008	0.417	0.009	0.394	0.015	0.388	0.015
22	0.448	0.011	0.430	0.009	0.465	0.012	0.418	0.006	0.441	0.011	0.450	0.009	0.481	0.014

more correlation with other factors found throughout the study. The graphical representation of convolution and low band-pass Butterworth filter can be seen in **Appendix B: CONVOLUTION GRAPHS** and **Appendix C: LOW BAND PASS BUTTERWORTH** respectively.

5.3 Results of Text Analysis

5.3.1 Result of Talking Frequency

Table 5.8 shows the results for the number of times a single participant spoke during the small group activity. Each participant is accompanied by the participant's ID number and the number of times they spoke overall, regardless of the amount spoken. Next to the number of times spoken, is the percentage of times spoken throughout the meeting per group. The final column of numbers is the normalization of each participant's number of times spoken. The final column is not divided per group but organized individually under the same range $[0, 1]$, with the minimum

Table 5.7: The table below shows an example of the individual participant's movement mean and variance. Each participant contains an ID and movement score separated by overall (O), first half(FH), second half(SH), first quarter(FQ), second quarter(SQ), third quarter(TQ), last quarter(LQ) movement scores.

ID	O Mean	O Variance	FH Mean	FH Variance	SH Mean	SH Variance	FQ Mean	FQ Variance	SQ Mean	SQ Variance	TQ Mean	TQ Variance	LQ Mean	LQ Variance
1_1	0.305	0.022	0.343	0.023	0.267	0.019	0.354	0.018	0.331	0.027	0.277	0.021	0.256	0.016
1_2	0.195	0.009	0.232	0.012	0.158	0.003	0.259	0.012	0.205	0.011	0.150	0.003	0.165	0.003
1_3	0.131	0.002	0.125	0.002	0.138	0.002	0.115	0.001	0.135	0.002	0.122	0.001	0.154	0.003
1_4	0.200	0.004	0.186	0.003	0.215	0.005	0.187	0.003	0.185	0.004	0.187	0.003	0.244	0.005
2_1	0.339	0.005	0.363	0.006	0.316	0.002	0.399	0.004	0.327	0.006	0.330	0.002	0.302	0.001
2_2	0.384	0.008	0.376	0.005	0.392	0.010	0.357	0.005	0.395	0.005	0.397	0.006	0.387	0.014
2_3	0.212	0.004	0.204	0.005	0.220	0.003	0.202	0.007	0.206	0.004	0.216	0.002	0.223	0.003
2_4	0.423	0.016	0.428	0.021	0.418	0.011	0.478	0.024	0.379	0.012	0.394	0.010	0.443	0.011
3_1	0.431	0.012	0.431	0.010	0.432	0.015	0.406	0.004	0.457	0.014	0.449	0.011	0.414	0.018
3_2	0.301	0.007	0.293	0.006	0.309	0.009	0.293	0.004	0.292	0.007	0.306	0.009	0.312	0.009
3_3	0.213	0.005	0.215	0.004	0.210	0.006	0.235	0.005	0.195	0.003	0.203	0.003	0.218	0.008
3_4	0.365	0.020	0.364	0.016	0.367	0.023	0.333	0.013	0.395	0.018	0.351	0.024	0.382	0.023
4_1	0.273	0.010	0.256	0.010	0.289	0.009	0.248	0.009	0.265	0.012	0.288	0.011	0.290	0.006
4_2	0.273	0.004	0.247	0.003	0.299	0.003	0.223	0.003	0.272	0.002	0.295	0.003	0.302	0.003
4_3	0.242	0.007	0.251	0.007	0.233	0.007	0.280	0.004	0.221	0.008	0.221	0.006	0.245	0.008
4_4	0.218	0.009	0.203	0.006	0.234	0.011	0.193	0.005	0.213	0.007	0.195	0.009	0.272	0.011
5_1	0.375	0.014	0.381	0.012	0.369	0.016	0.345	0.010	0.418	0.011	0.407	0.016	0.330	0.014
5_2	0.242	0.012	0.256	0.011	0.229	0.012	0.250	0.016	0.261	0.006	0.260	0.010	0.198	0.013
5_3	0.200	0.006	0.224	0.006	0.176	0.005	0.226	0.005	0.221	0.006	0.178	0.005	0.174	0.005
5_4	0.389	0.024	0.332	0.020	0.446	0.021	0.256	0.007	0.408	0.021	0.530	0.020	0.362	0.009
6_1	0.261	0.011	0.258	0.011	0.265	0.010	0.215	0.010	0.301	0.009	0.266	0.011	0.263	0.009
6_2	0.241	0.003	0.233	0.002	0.250	0.004	0.234	0.003	0.233	0.002	0.244	0.003	0.256	0.004
6_3	0.277	0.022	0.269	0.022	0.284	0.021	0.281	0.026	0.257	0.019	0.270	0.018	0.299	0.023
6_4	0.366	0.011	0.350	0.009	0.382	0.012	0.296	0.008	0.405	0.004	0.378	0.010	0.386	0.014
6_5	0.467	0.009	0.476	0.008	0.457	0.010	0.494	0.007	0.458	0.008	0.479	0.008	0.436	0.010
7_1	0.490	0.010	0.474	0.011	0.506	0.009	0.457	0.011	0.490	0.010	0.511	0.008	0.501	0.010
7_2	0.379	0.004	0.350	0.002	0.408	0.004	0.342	0.003	0.357	0.002	0.399	0.004	0.417	0.004
7_3	0.397	0.007	0.372	0.006	0.422	0.007	0.337	0.006	0.407	0.003	0.385	0.003	0.458	0.009

number of spoken turns being 15 and the maximum number of spoken turns being 170.

Table 5.8: The table below shows an example of the individual participant's Speaking Turn scores. Each participant contains an ID, number of speaking turns, percentage of speaking turns, and the normalization values of those speaking turns.

Participants	Number of Speaking Turns	Percentage of Speaking Turns	Normalization of Speaking Turns
1_1	70	0.248	0.355
1_2	97	0.344	0.529
1_3	60	0.213	0.290
1_4	55	0.195	0.258
2_1	106	0.281	0.587
2_2	97	0.257	0.529
2_3	80	0.212	0.419
2_4	94	0.249	0.510
3_1	85	0.215	0.452
3_2	147	0.372	0.852
3_3	68	0.172	0.342
3_4	95	0.241	0.516
4_1	88	0.259	0.471
4_2	55	0.162	0.258
4_3	152	0.447	0.884
4_4	45	0.132	0.194
5_1	66	0.185	0.329
5_2	67	0.188	0.335
5_3	164	0.461	0.961
5_4	59	0.166	0.284
6_1	119	0.332	0.671
6_2	81	0.226	0.426
6_3	91	0.254	0.490
6_4	30	0.084	0.097
6_5	37	0.103	0.142
7_1	155	0.455	0.903
7_2	89	0.261	0.477
7_3	97	0.284	0.529

5.3.2 Results of Word Frequency

Table 5.9 shows the results for the number of words spoken by a single participant during the small group meeting. In the table below, each row will contain the participant's ID number and the number of words spoken. Next to the number of words spoken, is the percentage of words spoken throughout the meeting per group. The final column of numbers is the normalization of each participant's number of words spoken. The final column is not divided per group but organized individually under the same range $[0, 1]$, with the minimum number of words spoken being 54 and the maximum number of words spoken being $1,498$.

5.3.3 Results of Keyword Frequency

Table 5.10 shows the results for the number of times a single participant spoke a keyword during the small group activity. Each participant is accompanied by their ID number and the number of keywords they spoke overall. Next to the number of keywords spoken, is the percentage of keywords spoken throughout the meeting per group. The final column of numbers is the normalization of each participant's number of keywords spoken. The final column is not divided per group but organized individually under the same range $[0, 1]$, with the minimum number of spoken turns being 0 and the maximum number of spoken turns being 86.

Table 5.9: The table below shows an example of the individual participant's Word Frequency scores. Each participant contains an ID, number of word frequency, percentage of word frequency, and the normalization values of those word frequency.

Participants	Number of Words	Percentage of Words	Normalization of Words
1_1	639	0.246	0.405
1_2	1119	0.430	0.738
1_3	435	0.167	0.264
1_4	408	0.157	0.245
2_1	922	0.294	0.601
2_2	822	0.262	0.532
2_3	681	0.217	0.434
2_4	713	0.227	0.456
3_1	640	0.261	0.406
3_2	850	0.347	0.551
3_3	362	0.148	0.213
3_4	597	0.244	0.376
4_1	668	0.274	0.425
4_2	419	0.172	0.253
4_3	1075	0.441	0.707
4_4	275	0.113	0.153
5_1	400	0.166	0.240
5_2	379	0.157	0.225
5_3	1352	0.561	0.899
5_4	278	0.115	0.155
6_1	944	0.391	0.616
6_2	512	0.212	0.317
6_3	673	0.279	0.429
6_4	54	0.022	0.000
6_5	232	0.096	0.123
7_1	821	0.413	0.531
7_2	671	0.337	0.427
7_3	498	0.250	0.307

Table 5.10: The table below shows an example of the individual participant's Keyword Frequency scores. Each participant contains an ID, number of keyword frequency, percentage of keyword frequency, and the normalization values of those keyword frequency.

Participants	Number of Keywords Used	Percentage of Keywords Used	Normalization of Keywords Used
1_1	29	0.199	0.337
1_2	65	0.445	0.756
1_3	35	0.240	0.407
1_4	17	0.116	0.198
2_1	60	0.326	0.698
2_2	56	0.304	0.651
2_3	32	0.174	0.372
2_4	36	0.196	0.419
3_1	23	0.176	0.267
3_2	58	0.443	0.674
3_3	18	0.137	0.209
3_4	32	0.244	0.372
4_1	34	0.306	0.395
4_2	15	0.135	0.174
4_3	50	0.450	0.581
4_4	12	0.108	0.140
5_1	20	0.139	0.233
5_2	22	0.153	0.256
5_3	86	0.597	1.000
5_4	16	0.111	0.186
6_1	44	0.383	0.512
6_2	22	0.191	0.256
6_3	34	0.296	0.395
6_4	3	0.026	0.035
6_5	12	0.104	0.140
7_1	62	0.470	0.721
7_2	36	0.273	0.419
7_3	34	0.258	0.395

CHAPTER 6: DISCUSSION

Summary of Results

In this thesis, Motion Dynamics metrics were created from raw video data, which was then analyzed using group performance scores and Communication Dynamics. First, previous studies [5, 40, 42, 44, 47, 54, 63, 76, 87] were utilized to better understand Motion Dynamics and Communication Dynamics in relation to Group Dynamics. Using the Database Coding Schema, I found four different team-focused datasets in the researched studies including, AMI [15], ELEA [65], UGI [8], and GAPS [81]. After looking at the information found from the four datasets and the previous studies I researched [5, 40, 42, 44, 47, 54, 63, 76, 87] I saw that Motion Dynamics were less studied in a number of papers published in respect to the number of papers published using Communication Dynamics. To figure out why Motion Dynamics were being studied less than Communication Dynamics I created Motion Dynamics metrics to track the motion found from the video data provided by the UGI corpus. To create these metrics, the optical flow method was used to track the motion by analyzing the pixel-to-pixel difference between the frames found in video data. I filtered the motion data to remove any noise, then created our motion-derived metrics by finding the mean and variance of the overall, each half, and each quarter of the motion data.

A total of 9 Communication Dynamics and 27 scores displaying team performance were found and used to test the Motion Dynamic metrics. I then used the new metrics and scores to understand the effects Motion Dynamics has on team performance. I saw from the 27 team performance scores, that five correlated with at least one Motion Dynamic metric. This could indicate that Motion Dynamics is related to team problem solving and therefore should be a rich area of future investigation. Using the metrics calculated, There was no direct correlation between

Motion Dynamics and Communication Dynamics.

Implications

The work in this thesis could potentially be beneficial in contributing to understanding team problem solving processes. Salas et al.(2014) state a number of different states, processes, and conditions that could influence teamwork [61]. Those states are **Cooperation**, **Conflict**, **Coordination**, **Communication**, **Coaching**, **Cognition**, **Composition**, **Context**, and **Culture**; can be seen described further in Salas's work however in relation to this paper three states will be further reviewed. The first state focused on was Coordination, this is a process is used to show the organization of different individuals which is necessary to produce an outcome[61]. Wiltshire et al.(2019) studied Motion Dynamics to investigate coordination in teamwork problem solving [83]. In Wiltshire's paper, the method used to obtain data to test their research was to calculate the pixel-to-pixel difference between each corresponding frame [83]. The method they used to find these differences came from Paxton and Dale's (2013) frame-differencing method which has been shown to highlight the motion found in the videos [57]. This work uses a similar method in optical flow to create motion data. I also created mean-variance Motion Dynamic metrics with motion data, which displayed potential correlation with the team performance scores I calculated, specifically the Team Scores vs Expert Scores (Disagreement Penalty) scores. In the variance portion of the Motion Dynamics metrics, I saw most of the correlations. The variance metrics show how each individual movement differs from the overall movement, the mean, of the team. Those correlations show differentiating movement could be relevant in studies dealing with understanding possible synchrony found in teams. Synchrony is a tendency where teams could develop similar patterns while interacting [18]. Understanding a team's synchrony could allow us to view improvements in a team's **Coordination** or if there are signs of **Conflict**. **Conflict**, found in Salas et al. (2014), is a negative effect on team performance where individuals are clashing due to their opinions and ideas [61]. The variance metrics found in this work could assist with studies dealing with synchrony by tracking whether those differentiating movements over time improve or deteriorate. Examining synchrony in the team is seen in other studies [1, 43, 50], for instance, Likens et al. (2021) present a synchronization method called windowed multiscale synchrony to assist with calculating synchrony in noisy, non-linear time series, this provides a window to focus on how the stability coordination in social interaction fluctuate over time [46]. They measure this stability by locating coordination between two signals of equal length over time across different frequencies. These frequencies can often be noisy which can cause difficulties when tracking synchrony, to address this issue Liken et al., focused on distribution, rather than the mean, of the frequency. This method allowed Liken et al. to look at coordination patterns even where there were sudden changes [46]. Similarly, the methodology and data present in this thesis could be useful in understanding synchrony because it provides a possible way to track synchrony through the previously explained variance metrics.

Group dynamics is the process emerging when people in a group interact [25], and to help understand this process I believe a combination of metrics stemming from different dynamics is necessary. Dale et al.(2020) also analyze synchrony in triads in body movement using the optical flow of the raw video data to extract body movement metrics [18]. Alongside body movement, Dale tested other non-motion metrics to view any potential relationships and saw interactions between these metrics that could improve or create multimodal models for understanding synchrony [18]. In terms of the states presented by Salas et al.(2014), the multimodal model could give us an understanding of a team's **Composition**, which can help determine factors that affect team performance[61]. Some metrics that can assist with the understanding of a team's Composition, are speaking turns, dialog acts, and body activity. This thesis contains some metrics from Motion Dynamics and Communication Dynamics that correlate with the group outcome team performance score, allowing the possibility to create multimodal models. An example of multimodal models being made from both Motion Dynamics and Communication Dynamics can be seen in Stewart et al. (2021). In this study, team problem solving is being analyzed using a unimodal model approach [19, 48], the singular data source used for this model was language data. Once the unimodal model was created then Stewart et al. added sound and video data to that unimodal, language, model to find any improvements from the newly created multimodal model[71, 72]. Stewart's paper showed that the unimodal, language, and model he created gave positive results but with the addition of the other metric, the new multimodal model was shown to be an enhancement to the original unimodal model in some of the categories that were tested [71]. Likewise, as seen in Stewart (2021) and Dale (2020), I believe the methodology, data, and additional statistical analysis can provide metrics that could be applied to existing unimodal models to enhance the model and understanding of group dynamics.

Ultimately the goal should be to combine the objective computational sensing and evaluation methods discussed in this thesis with the subjective survey instruments to understand how effective previous approaches are in assessing teamwork.

The work in the thesis explores teamwork performance and uses what I found to create metrics that could help further understand teams. The exploration of teamwork is a continuous study seen in research leveraging different types of data and approaches [2, 22, 60]. Previous studies I concentrated on used the idea of group dynamics in their work where the definition of group dynamics I are using is being both part communication dynamics and motion dynamics. I saw different topics such as speech recognition [28, 42, 76], dominance detection [54, 55, 64], and the improvement of group performance models [22, 31, 87] use one or both parts of group dynamics in their work. Our study focused on improving techniques that have been used to capture and process video data that could be used for motion dynamics. From the topics previously mention I found dominance detection, which researches individuals who start displaying traits of superiority or leadership over other members of the group [54, 55, 64], as well as the improvement of group performance models, which uses different feature derived from teamwork data to create measures to assist with already existing group performance models [22, 31, 87], to be the main contributor of the usage of video data in their study. Understanding these topics and what metrics they used to help further their study led us to study the datasets that the metrics were derived. Two dataset mentioned earlier that were used for their video data were the ELEA dataset [5, 54, 63] and the UGI dataset [47, 87]. From those datasets, ELEA [65] and UGI [8], as well as previous research using video data had led us to the thought process of how I could improve on these datasets to provide more data and variability to other studies. Beyond this study, the use of other video data sources could provide more avenues for different motion metrics. Currently, with the data I had available during COVID-19, I used the UGI dataset to create and test our own metrics using optical flow to process the video data. I saw metric creation use optical flow and other motion tracking programs have been used in previous studies to record data from video recordings [7, 54]. Dale et al. (2020) used optical flow to measure body movement and calculated the individual's magnitude of movement of the videos allotted time[18]. I found optical flow used in calculating individual movement trends throughout their meeting. Avci et al. (2016) use multimodal data to predict performance in groups that are performing tasks that deal with decisions making [5]. Their approach uses the ELEA dataset which provides both audio and video data, their audio data is processed to extract speaking cues from the group and each individual. The video data was filtered and measured to calculate the group's and individual's head and body movements alongside gaze by tracking head and eye movements. Gaze in this context is referring to what the individual is looking at. They calculated gaze by using the participant's head position and location to predict visual focus of attention [5, 32, 69]. Moving forward the addition of other motion tracking methods could lead to the creation of more metrics.

Limitations

Current teamwork datasets have their usages and provide positive results, however, overall they need to be remodeled. From what I saw with the UGI dataset and other datasets, video data is less studied and analyzed in comparison to audio data. From four datasets looked at in this work, those being AMI [15], ELEA [65], UGI [8], and GAPS [81], I saw all four datasets contained usable audio data that had been analyzed and used in research. Where only the ELEA and the UGI datasets contained usable video data to be analyzed for other works. The video data I used presented by the UGI dataset was recorded through the use of a ceiling-mounted motion sensor, this did provide unique motion capture data. However, there only being one sensor and the one sensor being ceiling-mounted, meant the study was limited to the motion data I could collect. This led to difficulties in creating a variety of Motion Dynamic metrics for our tests. Additionally, with the placement of the sensor, the data contained a number of issues making parts of the video unusable. An example of these issues can be seen when participants would move out of sight from the sensor. These issues caused us to edit or in some cases remove sections from the video data. A more in-depth statistical analysis did not show significant results due to the difficulties found in the data and a lack of variability in using current video data metrics.

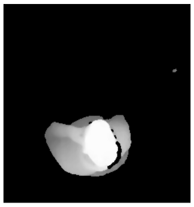
CHAPTER 7: CONCLUSION

As issues become more challenging, not only the demand, but the quality of teamwork to solve these matters become more pertinent than ever. This thesis took the first steps to study potential objective computational sensing methods to automatically understand teamwork measures. I looked at past literature and studies to better understand different datasets and measures used to analyze teamwork. I need to first break down existing teamwork data into a new taxonomy to be able to understand the fundamentals of small group teaming and their processes. Once a new taxonomy was created I then categorized and examined major research using four different selections of teamwork datasets that used unobtrusive sensing techniques. This research looked at ways these approaches could be improved to construct a novel constant-less sensing unobtrusive senor room to capture small team dynamics and processes. It is important to develop a methodology to automatically assess teams to understand how to help advance collaborative problem solving.

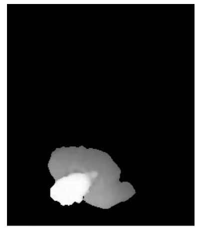
From the examined research and selection of dataset, I selected the UGI corpus to study in more detail and test our initial computational approaches. The UGI corpus had preexisting recorded video and audio data of teams conducting the *Lunar Survival Task*, as well as teamwork-based scoring for each participant and group. I calculated several team performance scores based on the objective rankings provided by UGI. This allowed a set of data to test correlations from the video data metrics and audio data metrics I calculated. I found a few potential areas that will narrow future studies looking at this area. This shows that current teamwork datasets have their usages and provide positive results but overall need to be remodeled with more systematic data capture approaches. From what I saw from the UGI dataset, as well as the other datasets, is that their video data is often poorly captured and processed. With video capturing technology continuing to improve in many different filters and channels, a simple video recording is no longer enough. Meaning a simple recording alongside other multi-modal channels could provide more rich data by giving different measurements to use for testing.

For future research into improving the quality of teamwork, I would create our own teamwork corpus using the lessons learned studying these other datasets in this thesis. This corpus, much like the other, would contain the recording, both in audio and video, of teams completing a task together. The room will be equipped with an improved audio recording device to pick up the conversation between the team and then transcribed using the AMI transcription method. Other improvements would include a variety of advanced video recordings these devices would allow a simple recording as well as the possibility of recording different channels as the UGI corpus did with the depth channel. In addition to the improved technology, I would include an assortment of sensors to track a multitude of additional variables. Those possible variables are the participants' vitals, gases found in the air, or tracking movement. This modernization accompanied by proven methods would create a dataset that may further develop future teamwork research.

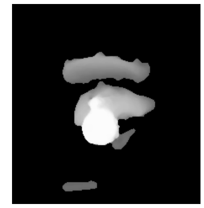
APPENDIX A: PROCESSED IMAGES



Team Member 1_1



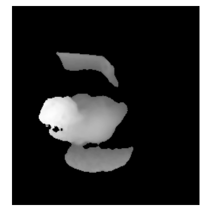
Team Member 1_2



Team Member 1_3

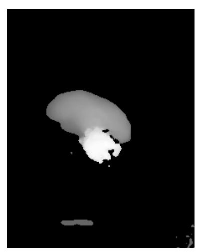
Team Member 1_4

Team Member 2_1

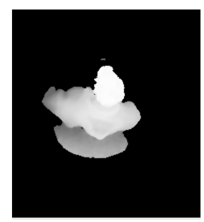


Team Member 2_2

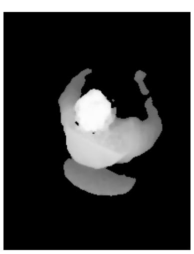
Team Member 2_3



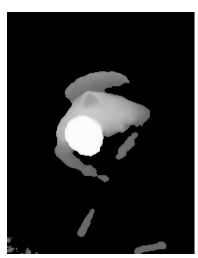
Team Member 2_4



Team Member 3_1

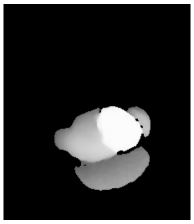


Team Member 3_2



Team Member 3_3

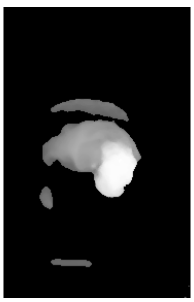
Team Member 3_4



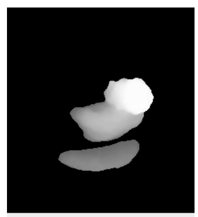
Team Member 4_1

Team Member 4_2

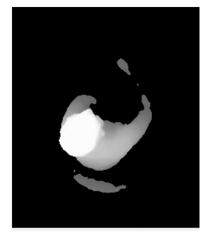
Team Member 4_3



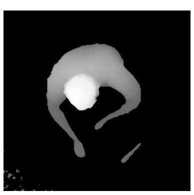
Team Member 4_4



Team Member 5_1

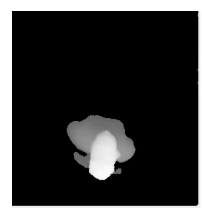


Team Member 5_2

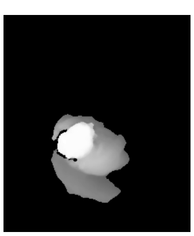


Team Member 5_3

Team Member 5_4

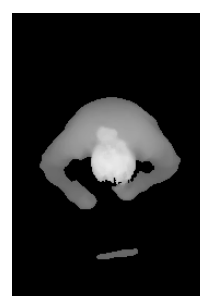


Team Member 6_1



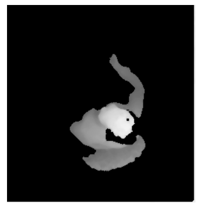
Team Member 6_2

Team Member 6_3



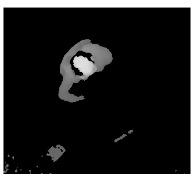
Team Member 6_4

Team Member 6_5

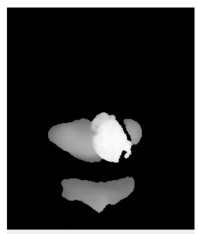


Team Member 7_1

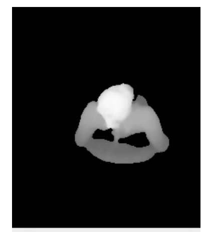
Team Member 7_2



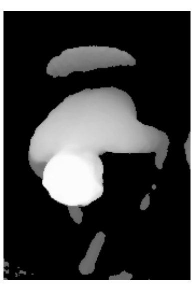
Team Member 7_3



Team Member 8_1



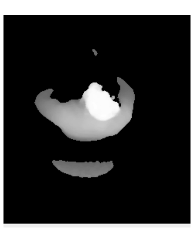
Team Member 8_2



Team Member 8_3

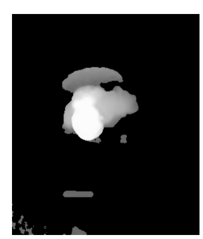
Team Member 8_4

Team Member 8_5

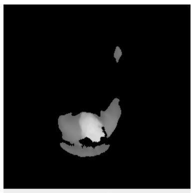


Team Member 9_1

Team Member 9_2



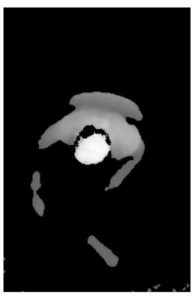
Team Member 9_3



Team Member 10_1

Team Member 10_2

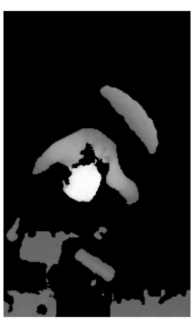
Team Member 10_3



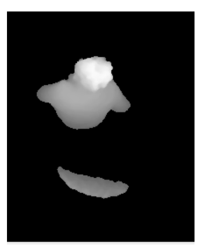
Team Member 10_4

Team Member 11_1

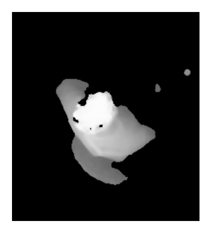
Team Member 11_2



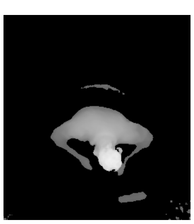
Team Member 11_3



Team Member 12_1

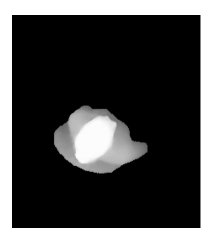


Team Member 12_2



Team Member 12_3

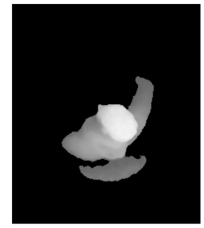
Team Member 13_1



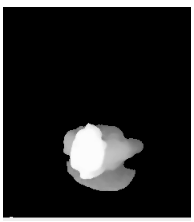
Team Member 13_2

Team Member 13_3

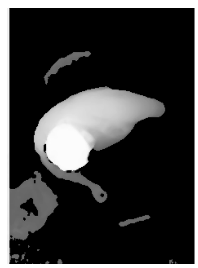
Team Member 13_4



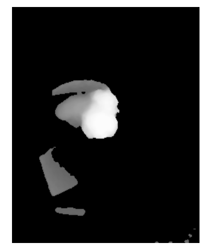
Team Member 14_1



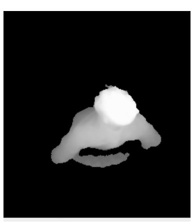
Team Member 14_2



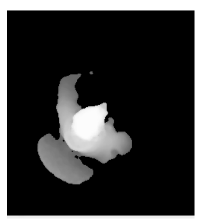
Team Member 14_3



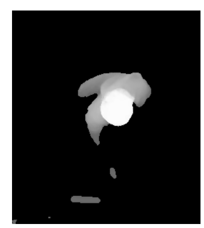
Team Member 14_4



Team Member 15_1

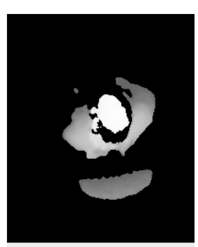


Team Member 15_2

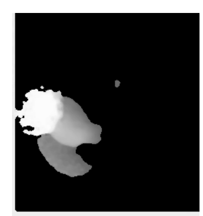


Team Member 15_3

Team Member 15_4

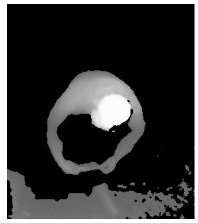


Team Member 16_1

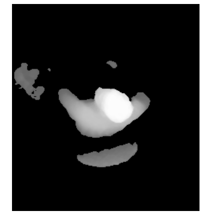


Team Member 16_2

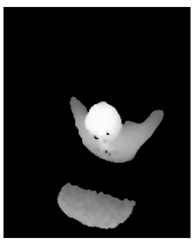
Team Member 16_3



Team Member 16_4



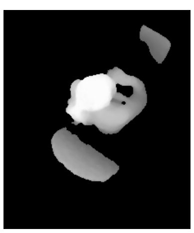
Team Member 17_1



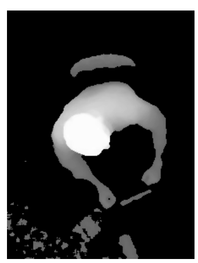
Team Member 17_2

Team Member 17_3

Team Member 18_1

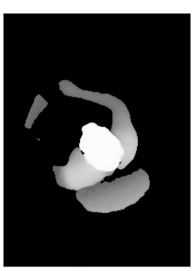


Team Member 18_2

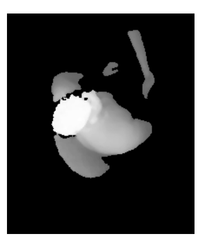


Team Member 18_3

Team Member 18_4

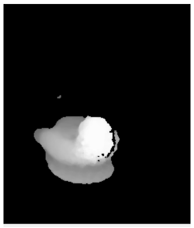


Team Member 19_1

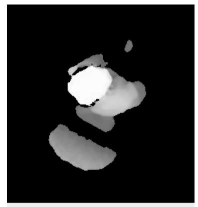


Team Member 19_2

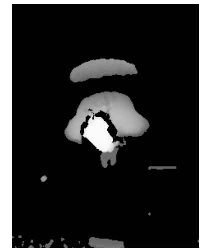
Team Member 19_3



Team Member 20_1

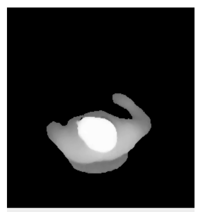


Team Member 20_2



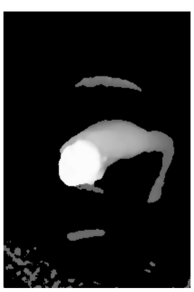
Team Member 20_3

Team Member 20_4



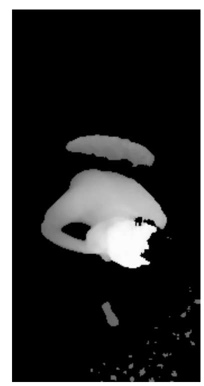
Team Member 21_1

Team Member 21_2

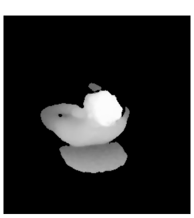


Team Member 21_3

Team Member 21_4

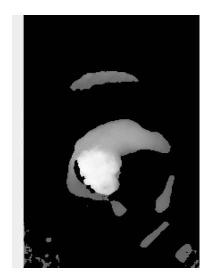


Team Member 21_5



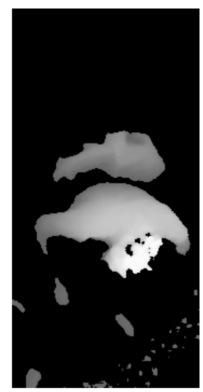
Team Member 22_1

Team Member 22_2



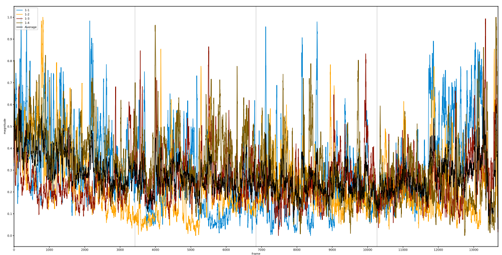
Team Member 22_3

Team Member 22_4

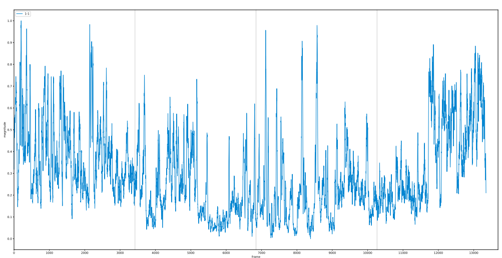


Team Member 22_5

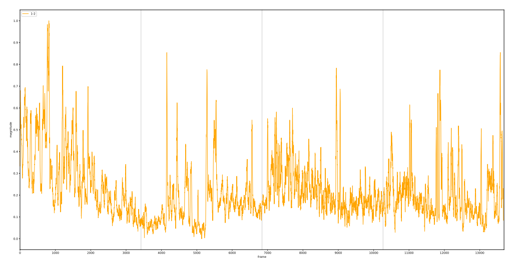
APPENDIX B: CONVOLUTION GRAPHS



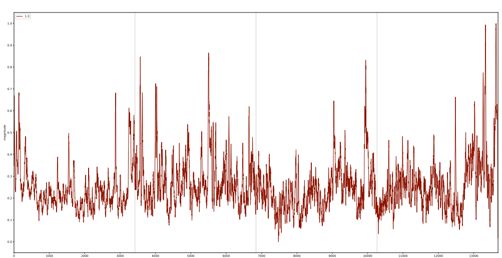
(a) Team 1



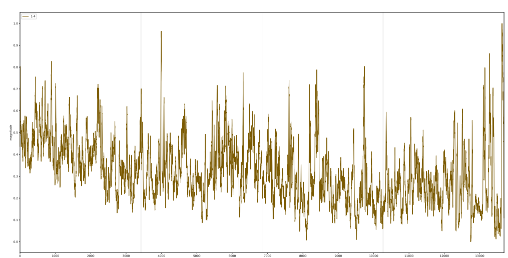
(b) Team Member 1_1



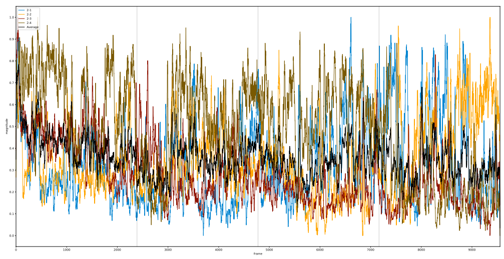
(c) Team Member 1_2



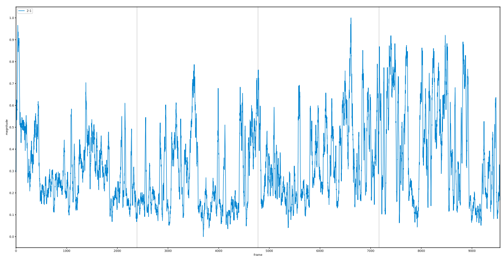
(d) Team Member 1_3



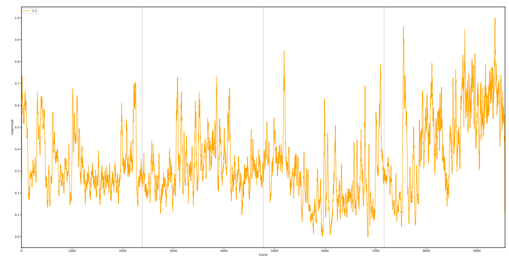
(e) Team Member 1_4



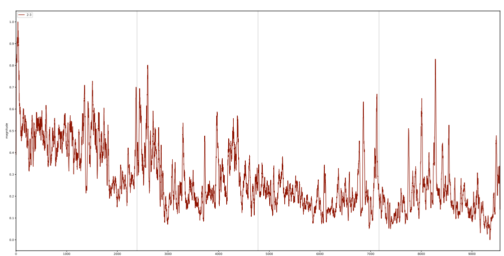
(a) Team 2



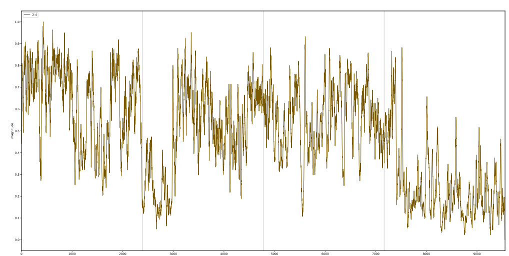
(b) Team Member 2_1



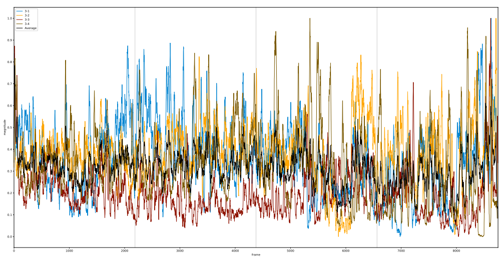
(c) Team Member 2_2



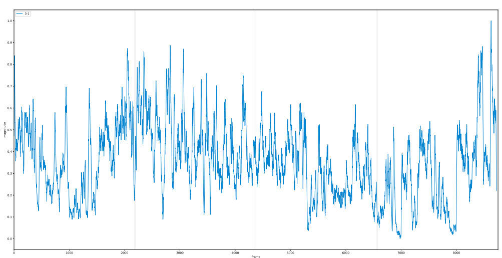
(d) Team Member 2_3



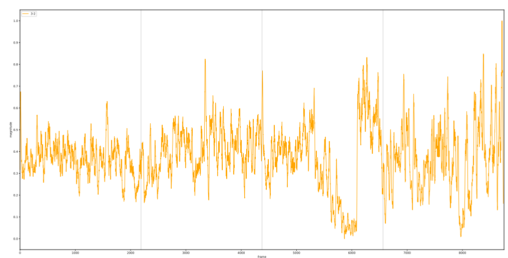
(e) Team Member 2_4



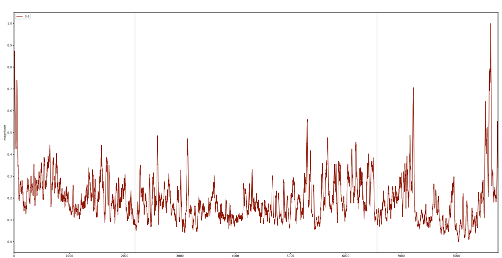
(a) Team 3



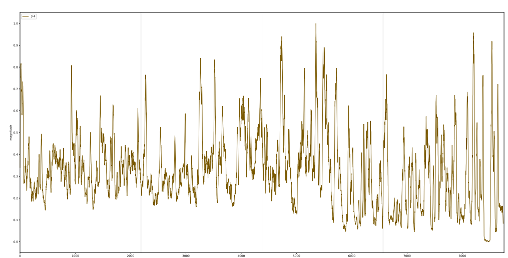
(b) Team Member 3_1



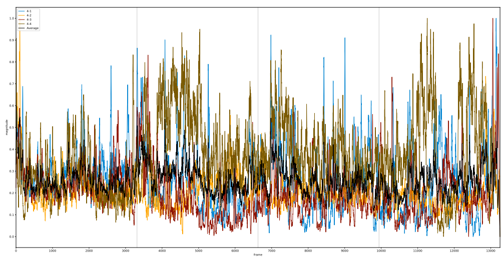
(c) Team Member 3_2



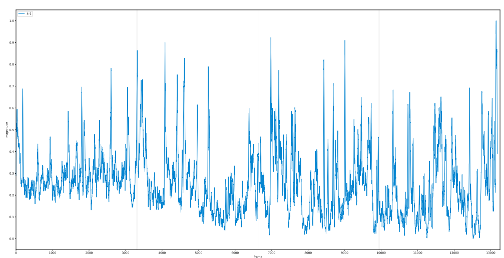
(d) Team Member 3_3



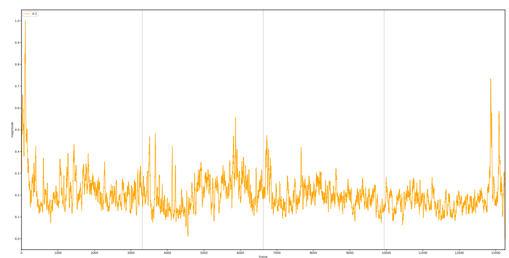
(e) Team Member 3_4



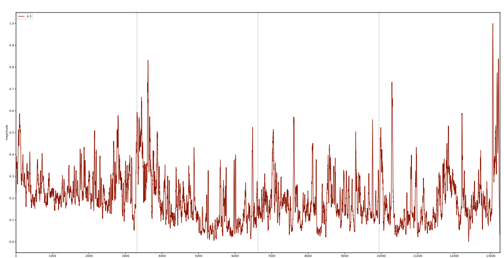
(a) Team 4



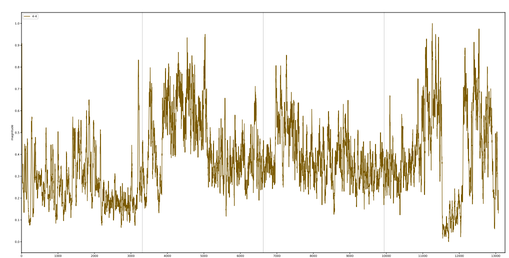
(b) Team Member 4_1



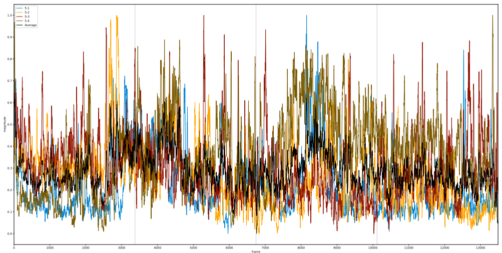
(c) Team Member 4_2



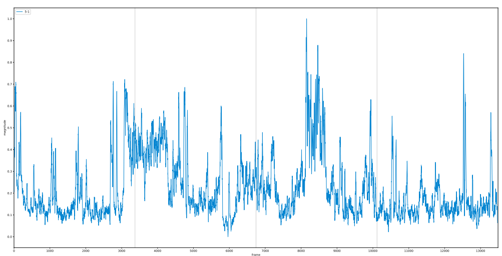
(d) Team Member 4_3



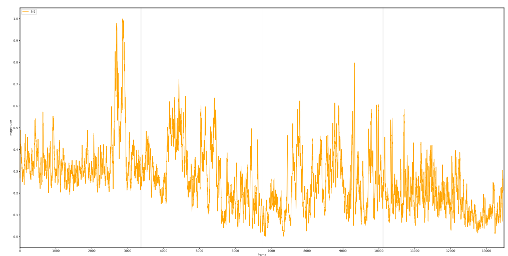
(e) Team Member 4_4



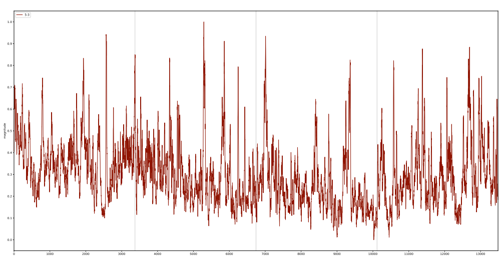
(a) Team 5



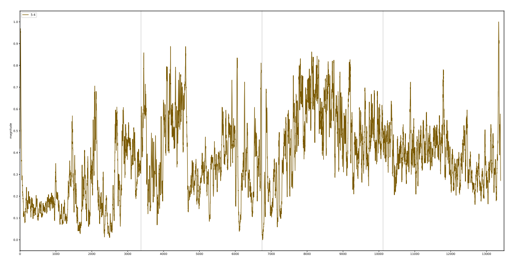
(b) Team Member 5_1



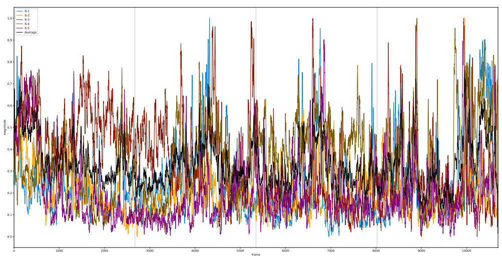
(c) Team Member 5_2



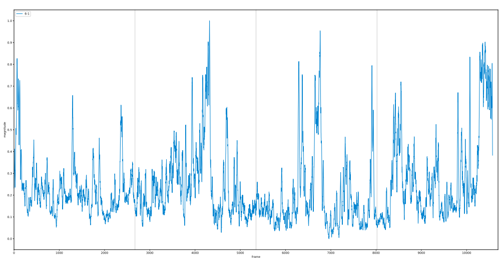
(d) Team Member 5_3



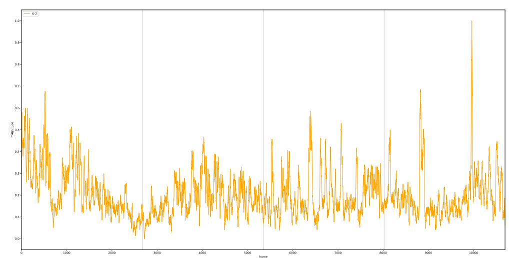
(e) Team Member 5_4



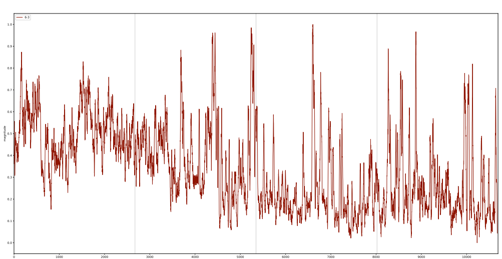
(a) Team 6



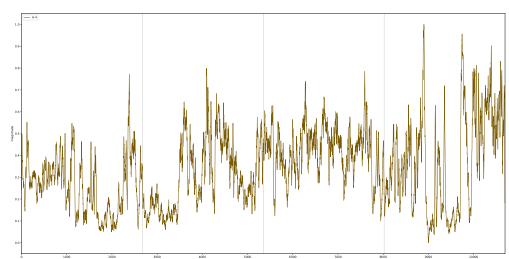
(b) Team Member 6_1



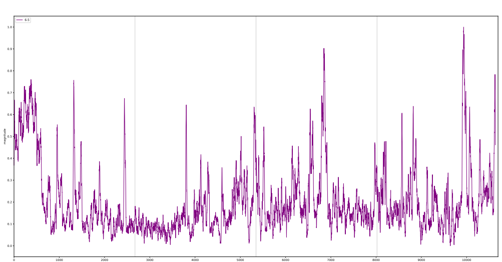
(c) Team Member 6_2



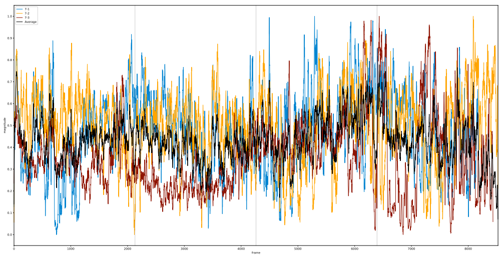
(d) Team Member 6_3



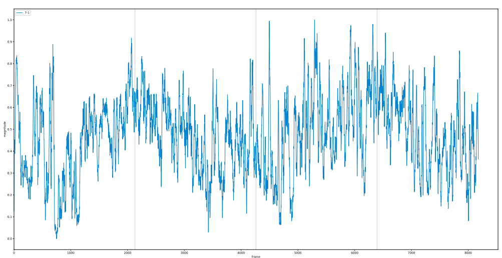
(e) Team Member 6_4



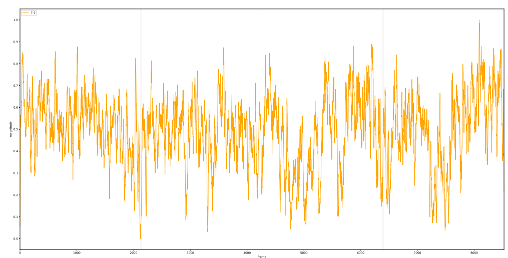
(f) Team Member 6_5



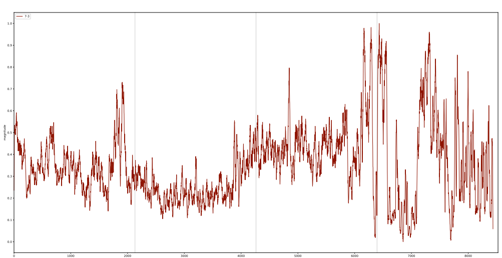
(a) Team 7



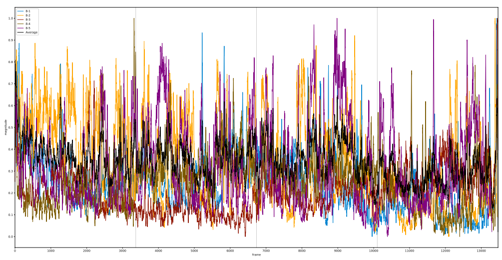
(b) Team Member 7_1



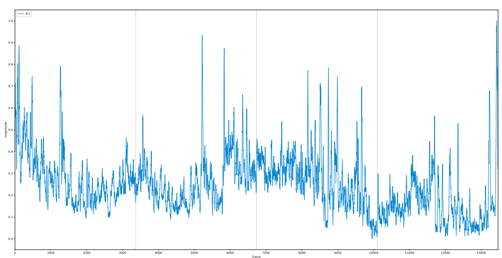
(c) Team Member 7_2



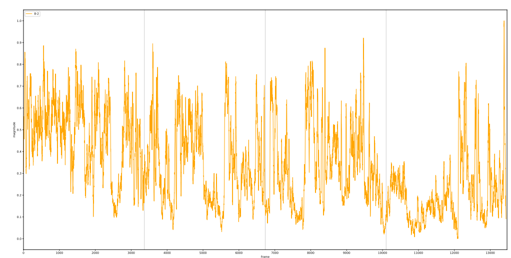
(d) Team Member 7_3



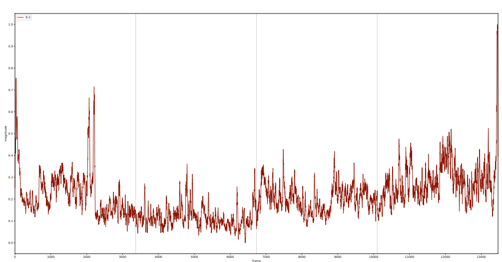
(a) Team 8



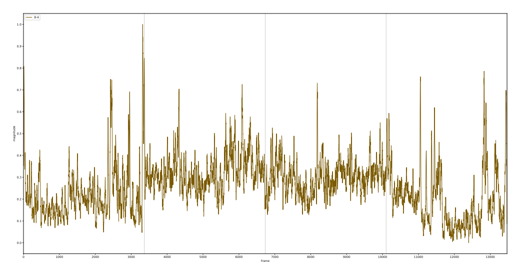
(b) Team Member 8_1



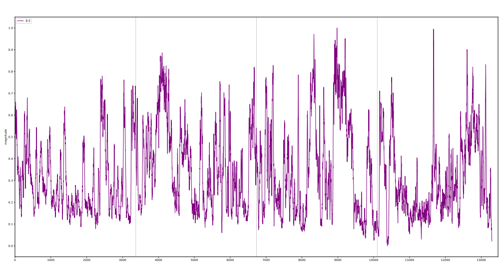
(c) Team Member 8_2



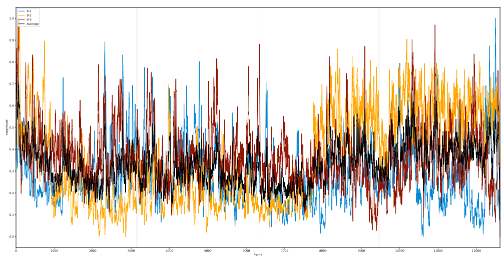
(d) Team Member 8_3



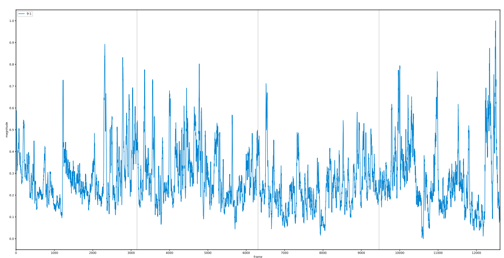
(e) Team Member 8_4



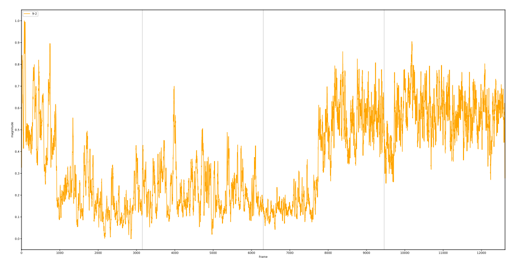
(f) Team Member 8_5



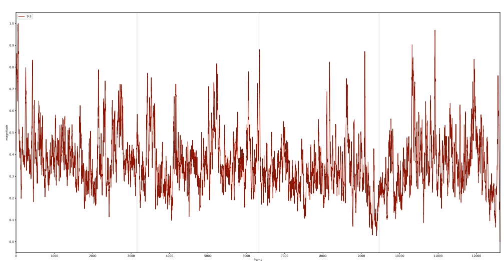
(a) Team 9



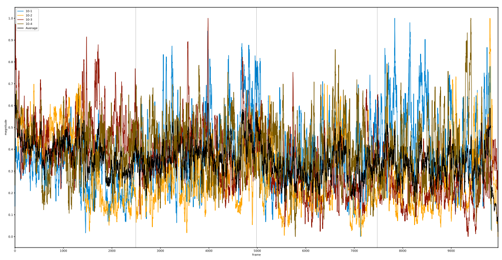
(b) Team Member 9_1



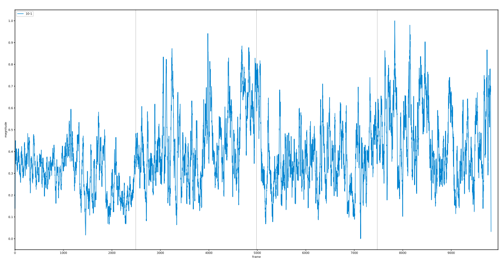
(c) Team Member 9_2



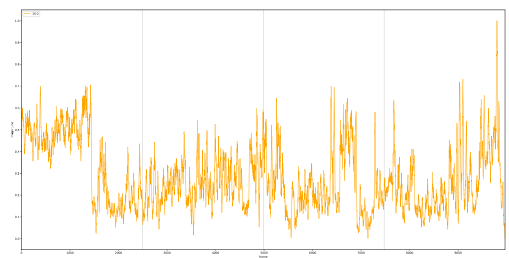
(d) Team Member 9_3



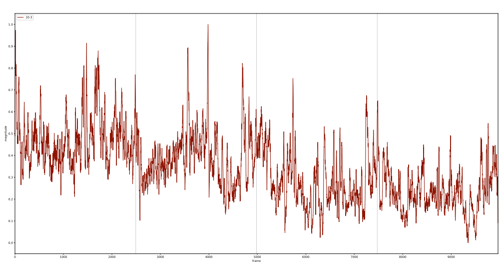
(a) Team 10



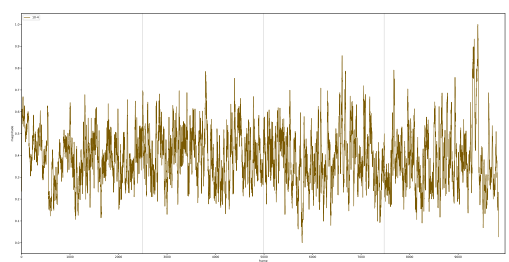
(b) Team Member 10_1



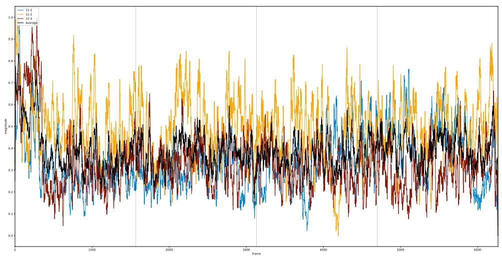
(c) Team Member 10_2



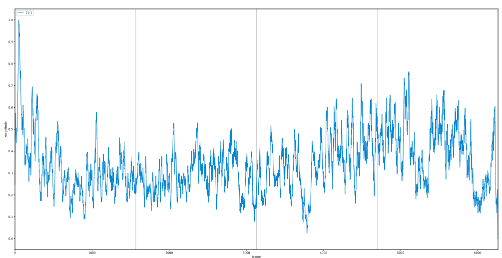
(d) Team Member 10_3



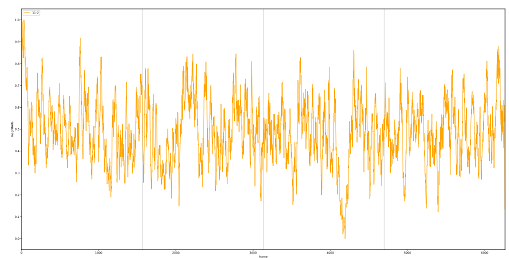
(e) Team Member 10_4



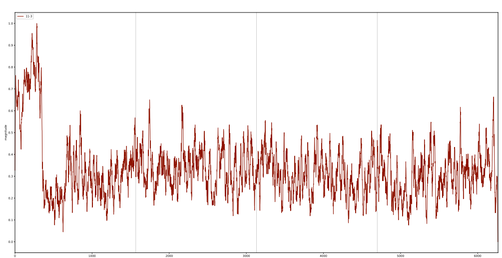
(a) Team 11



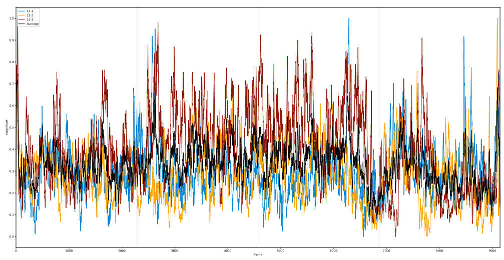
(b) Team Member 11_1



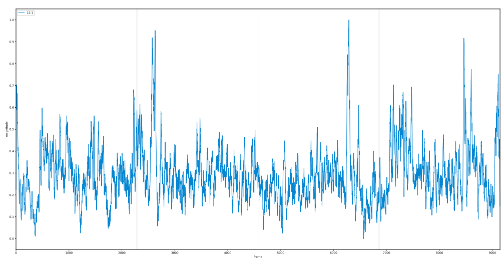
(c) Team Member 11_2



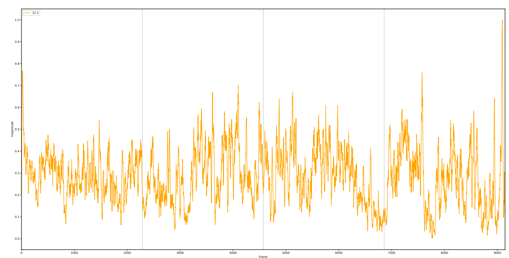
(d) Team Member 11_3



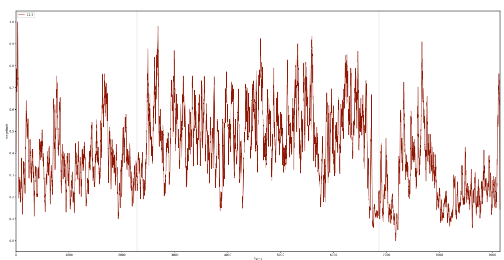
(a) Team 12



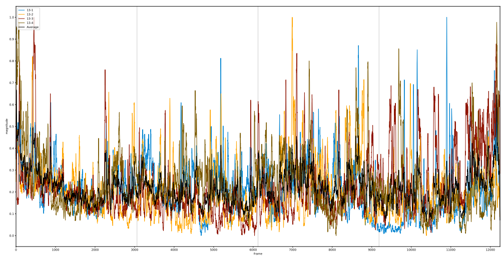
(b) Team Member 12_1



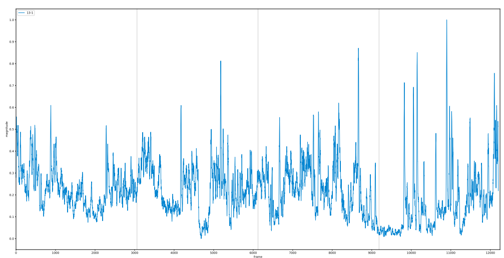
(c) Team Member 12_2



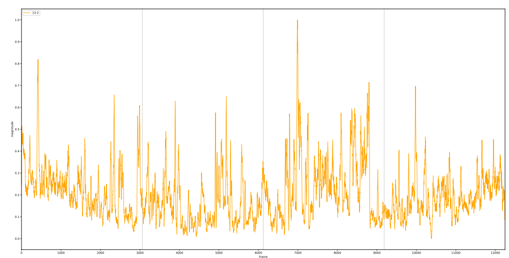
(d) Team Member 12_3



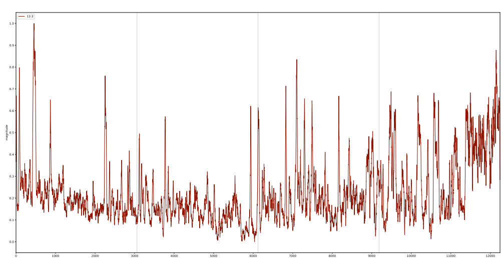
(a) Team 13



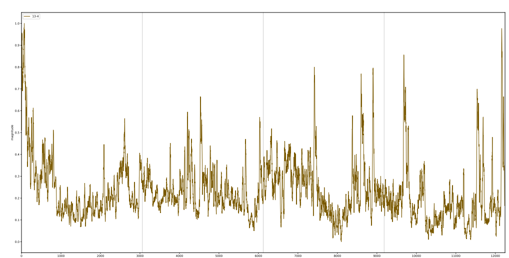
(b) Team Member 13_1



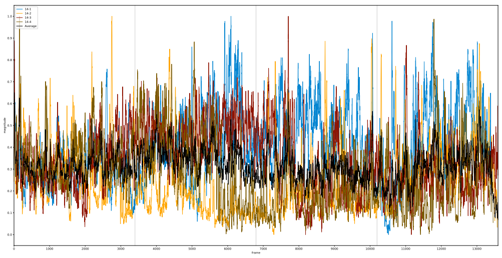
(c) Team Member 13_2



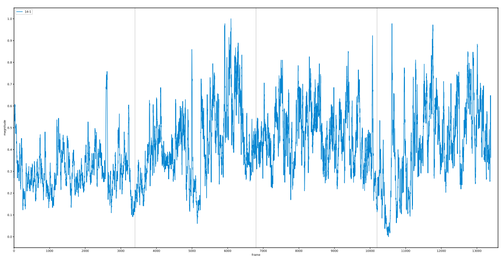
(d) Team Member 13_3



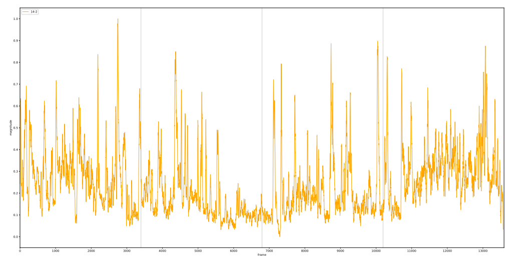
(e) Team Member 13_4



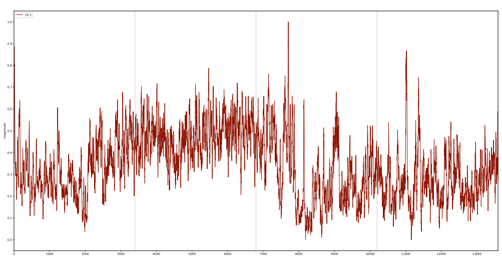
(a) Team 14



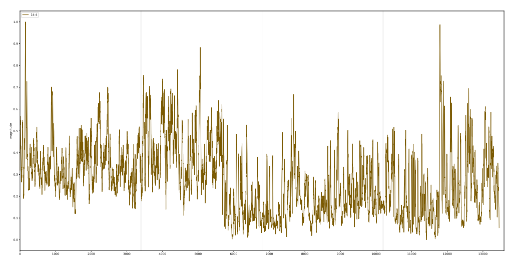
(b) Team Member 14_1



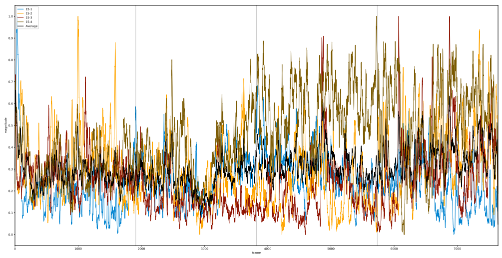
(c) Team Member 14_2



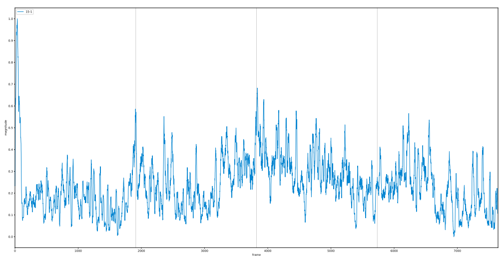
(d) Team Member 14_3



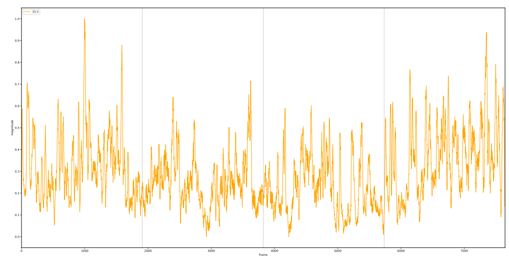
(e) Team Member 14_4



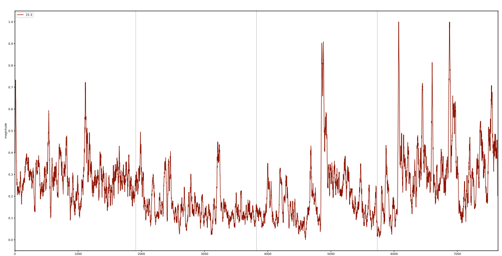
(a) Team 15



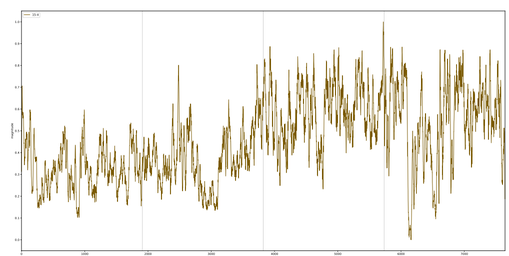
(b) Team Member 15_1



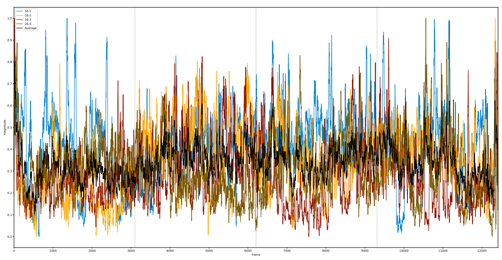
(c) Team Member 15_2



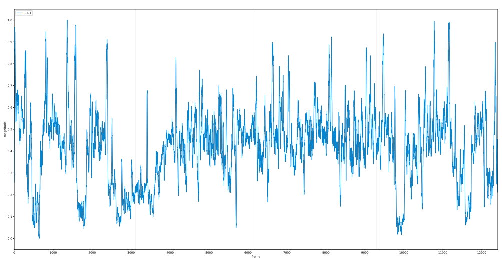
(d) Team Member 15_3



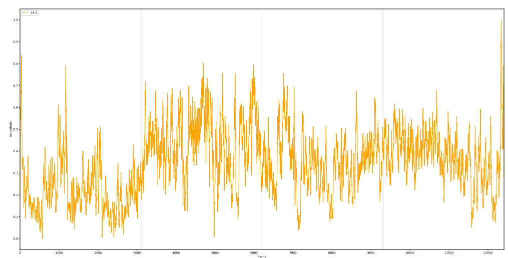
(e) Team Member 15_4



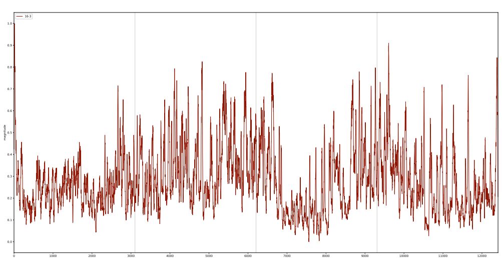
(a) Team 16



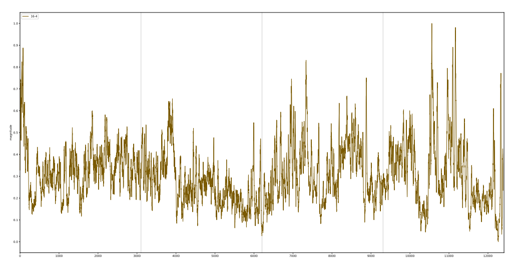
(b) Team Member 16_1



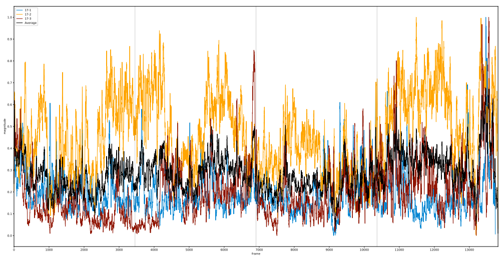
(c) Team Member 16_2



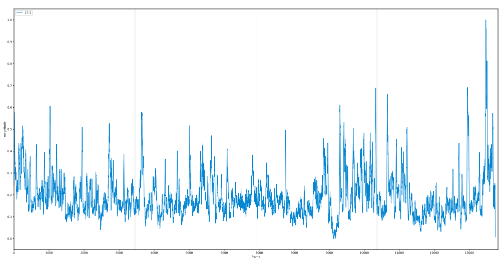
(d) Team Member 16_3



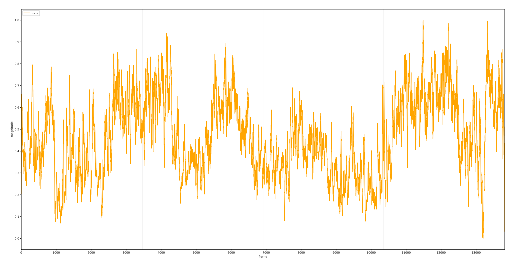
(e) Team Member 16_4



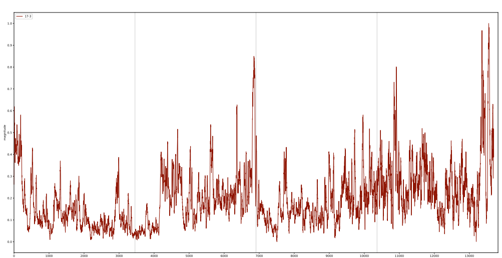
(a) Team 17



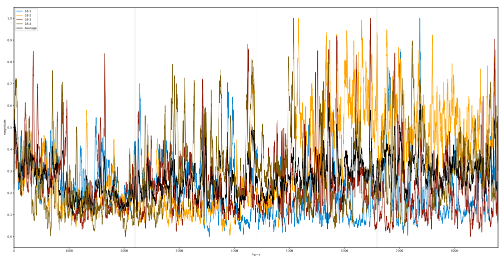
(b) Team Member 17_1



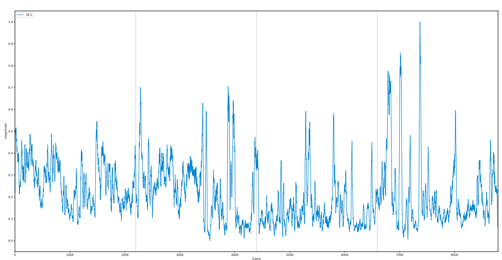
(c) Team Member 17_2



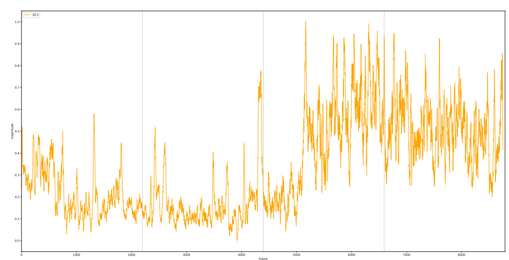
(d) Team Member 17_3



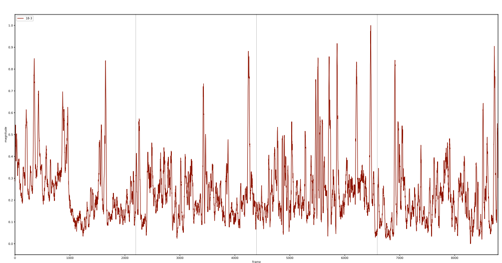
(a) Team 18



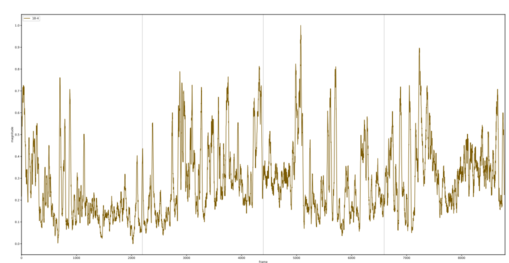
(b) Team Member 18_1



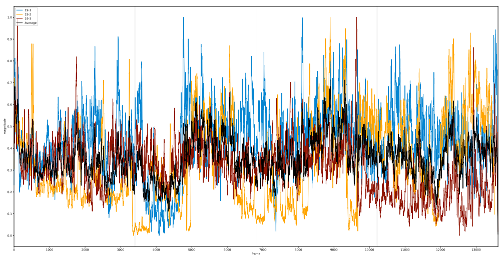
(c) Team Member 18_2



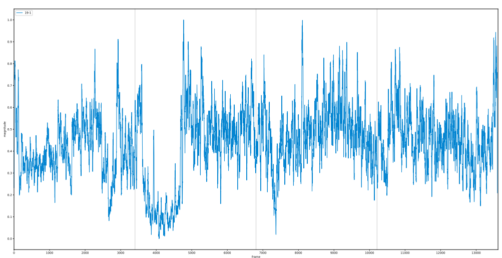
(d) Team Member 18_3



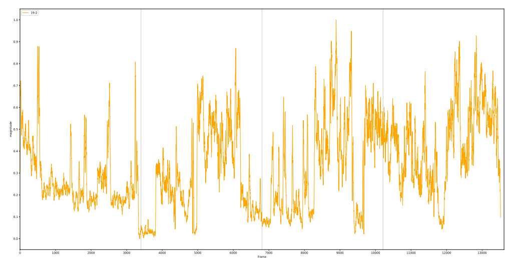
(e) Team Member 18_4



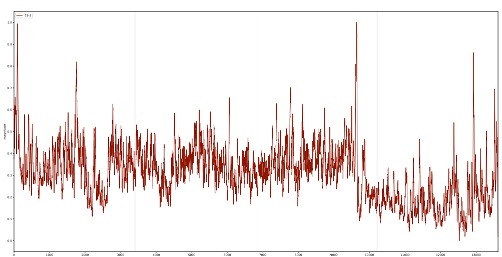
(a) Team 19



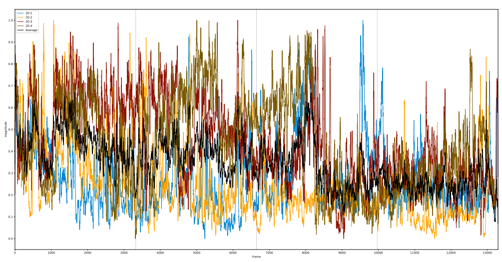
(b) Team Member 19_1



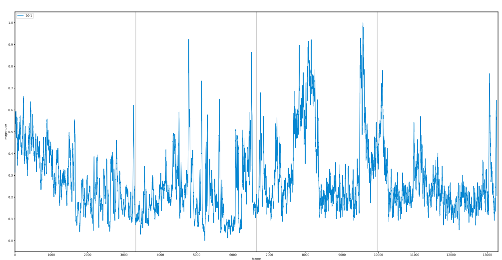
(c) Team Member 19_2



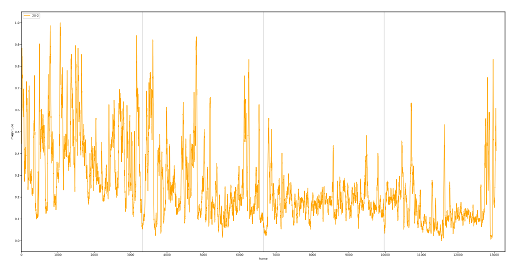
(d) Team Member 19_3



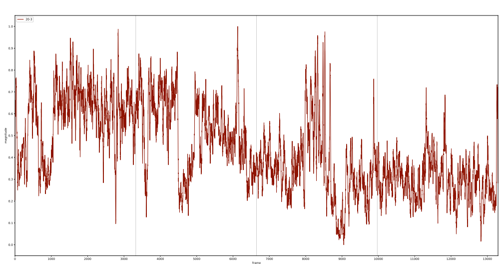
(a) Team 20



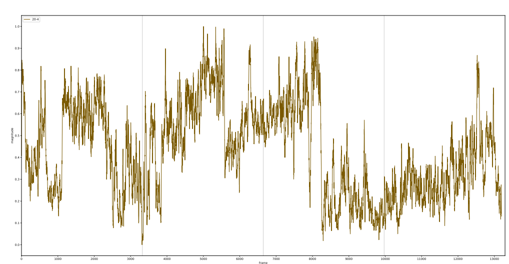
(b) Team Member 20_1



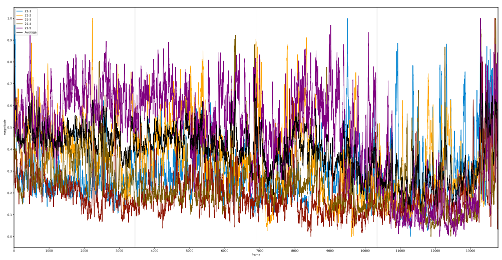
(c) Team Member 20_2



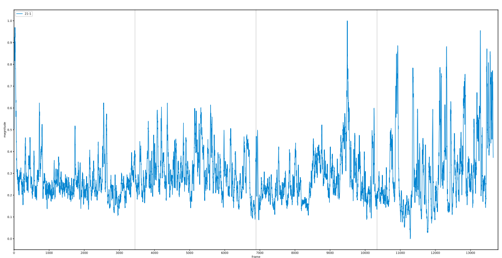
(d) Team Member 20_3



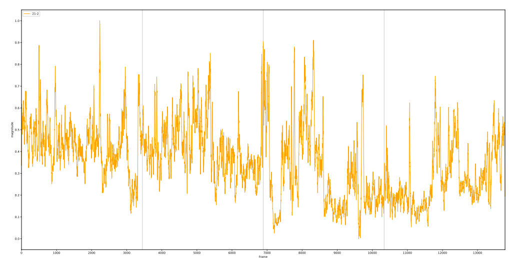
(e) Team Member 20_4



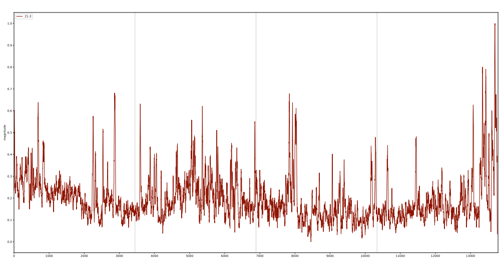
(a) Team 21



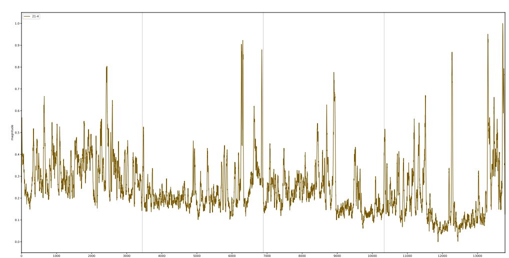
(b) Team Member 21_1



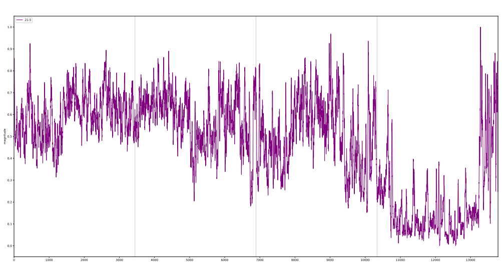
(c) Team Member 21_2



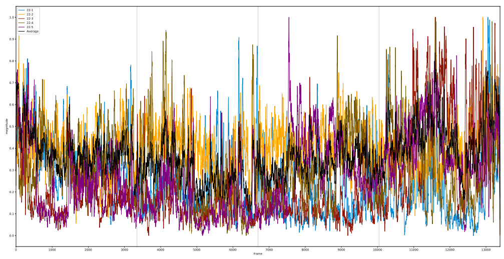
(d) Team Member 21_3



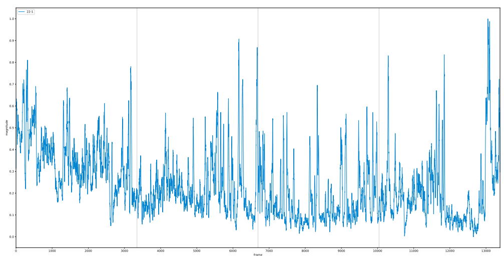
(e) Team Member 21_4



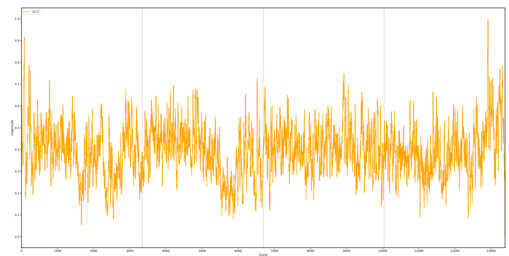
(f) Team Member 21_5



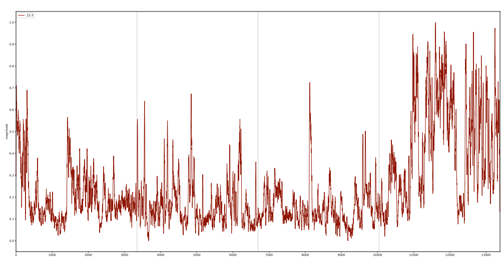
(a) Team 22



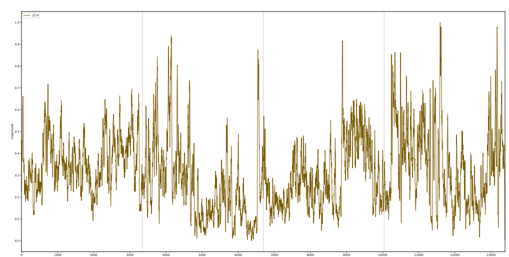
(b) Team Member 22_1



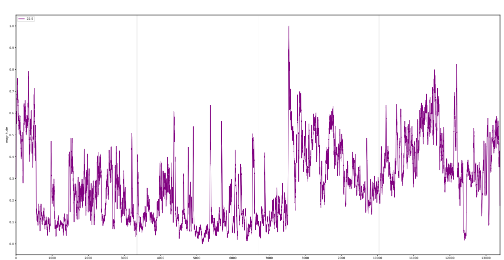
(c) Team Member 22_2



(d) Team Member 22_3

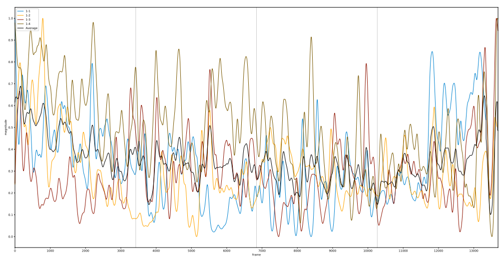


(e) Team Member 22_4

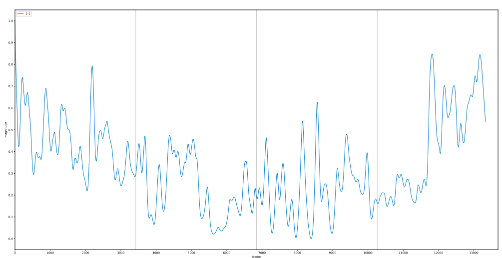


(f) Team Member 22_5

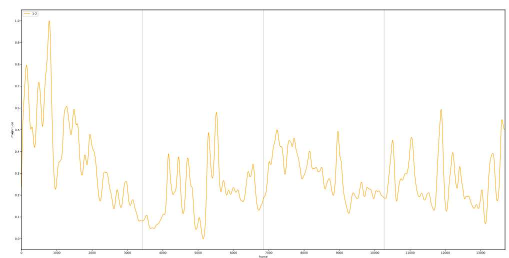
APPENDIX C: LOW BAND PASS BUTTERWORTH



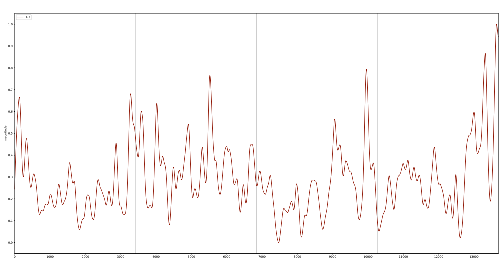
(a) Team 1



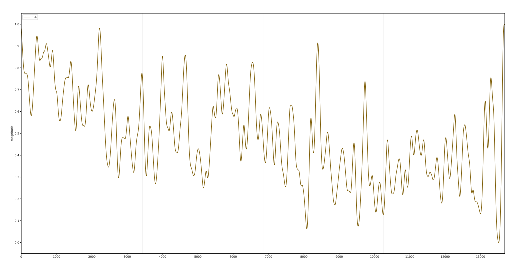
(b) Team Member 1_1



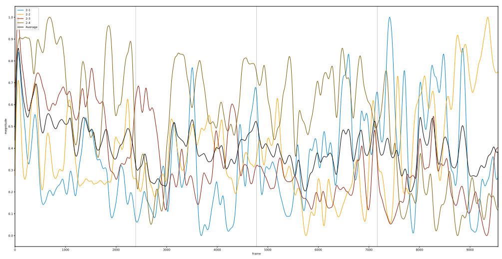
(c) Team Member 1_2



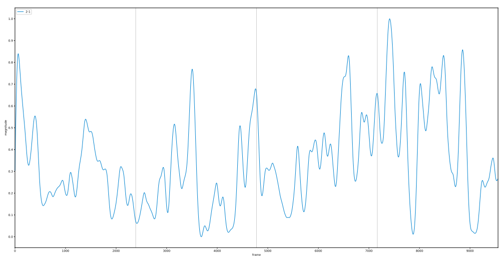
(d) Team Member 1_3



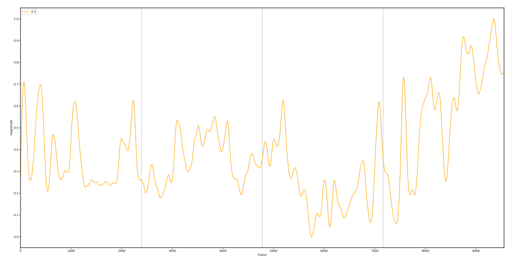
(e) Team Member 1_4



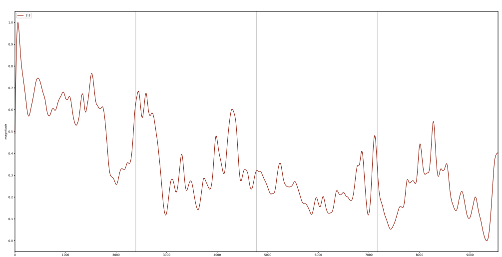
(a) Team 2



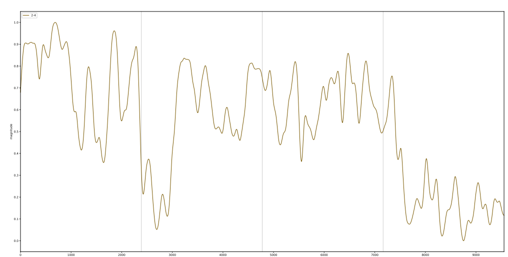
(b) Team Member 2_1



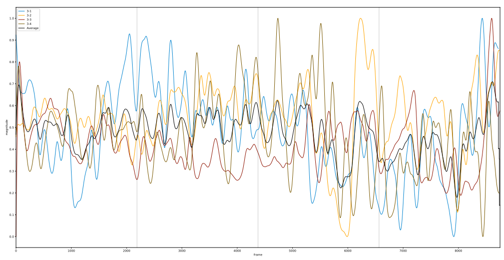
(c) Team Member 2_2



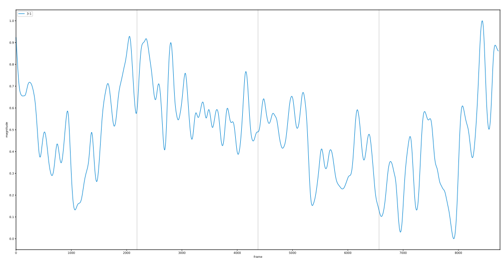
(d) Team Member 2_3



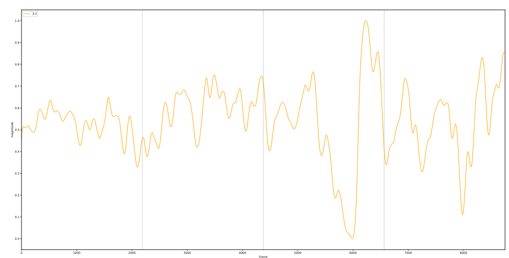
(e) Team Member 2_4



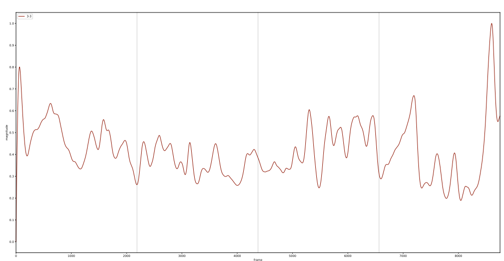
(a) Team 3



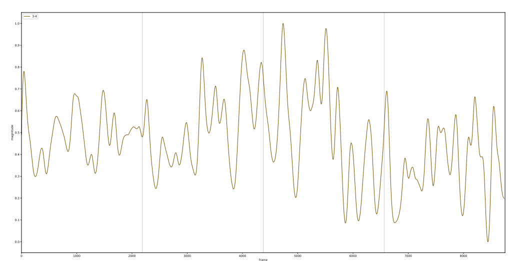
(b) Team Member 3_1



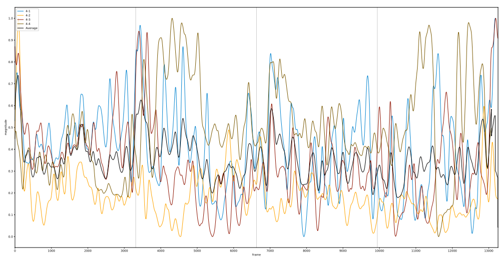
(c) Team Member 3_2



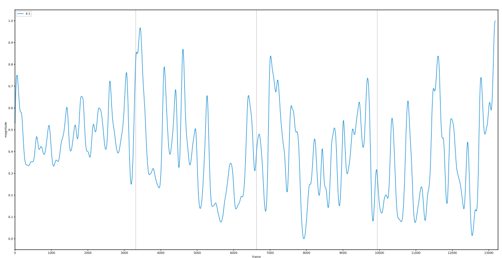
(d) Team Member 3_3



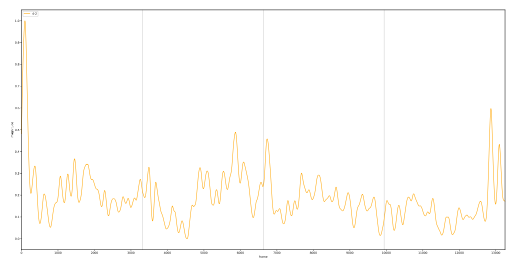
(e) Team Member 3_4



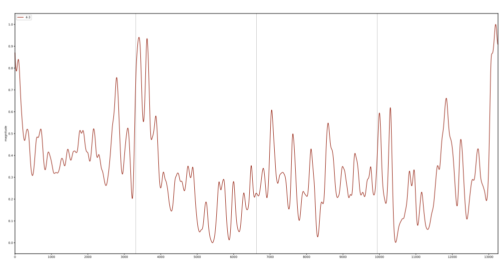
(a) Team 4



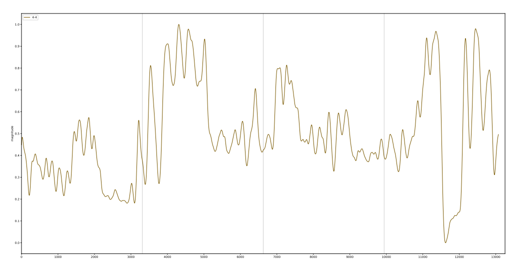
(b) Team Member 4_1



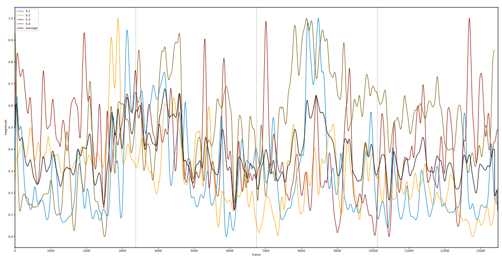
(c) Team Member 4_2



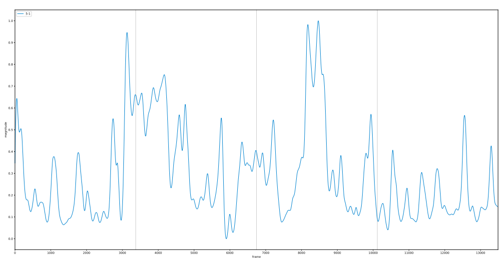
(d) Team Member 4_3



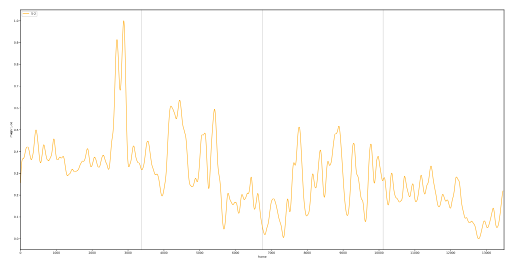
(e) Team Member 4_4



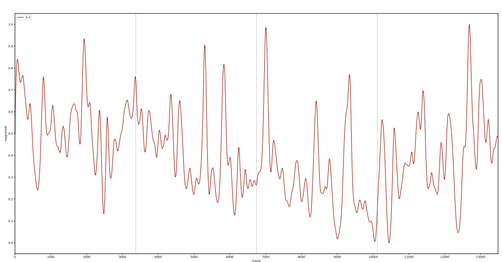
(a) Team 5



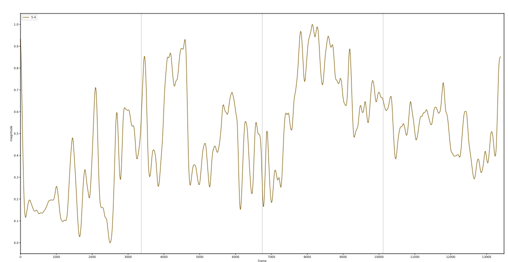
(b) Team Member 5_1



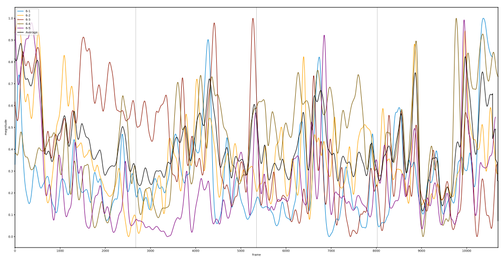
(c) Team Member 5_2



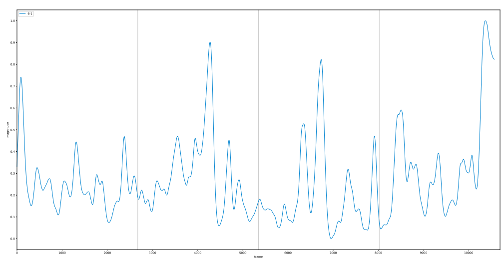
(d) Team Member 5_3



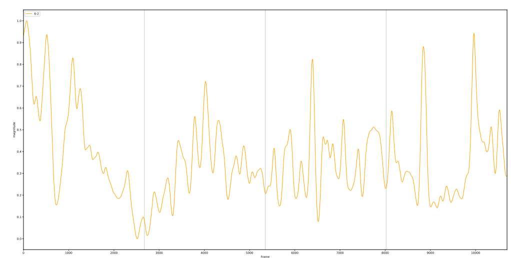
(e) Team Member 5_4



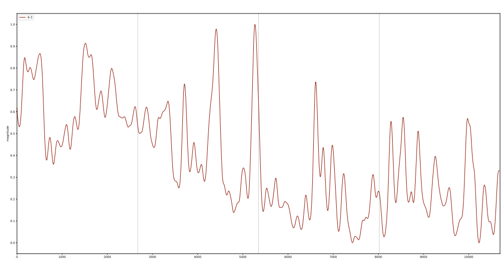
(a) Team 6



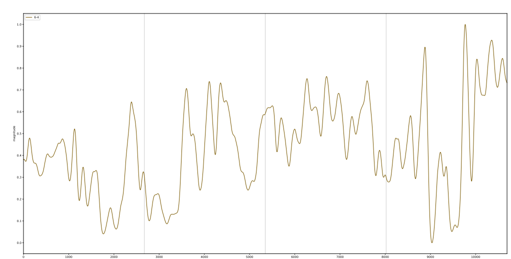
(b) Team Member 6_1



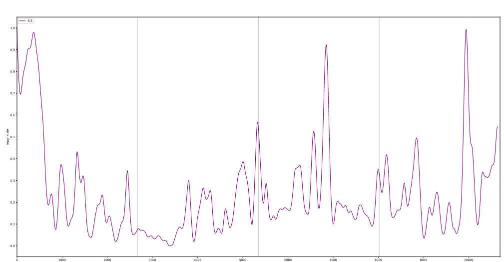
(c) Team Member 6_2



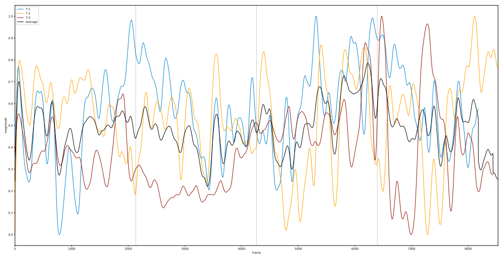
(d) Team Member 6_3



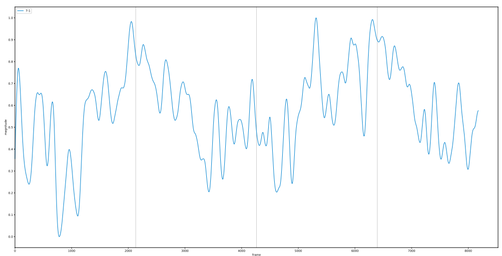
(e) Team Member 6_4



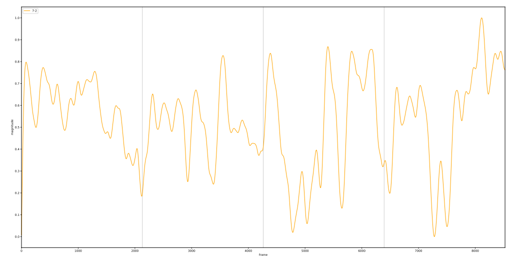
(f) Team Member 6_5



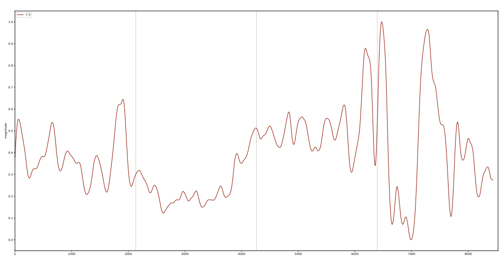
(a) Team 7



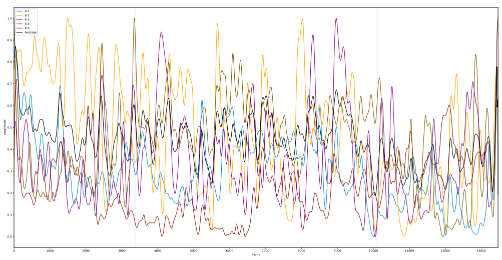
(b) Team Member 7_1



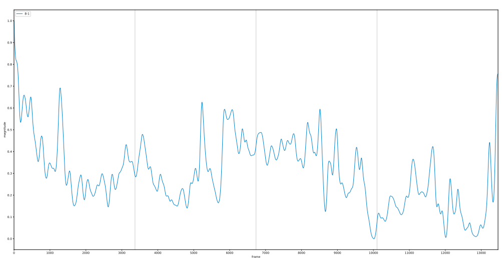
(c) Team Member 7_2



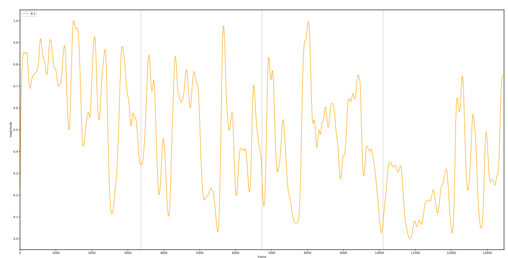
(d) Team Member 7_3



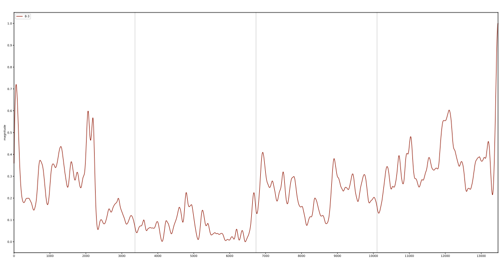
(a) Team 8



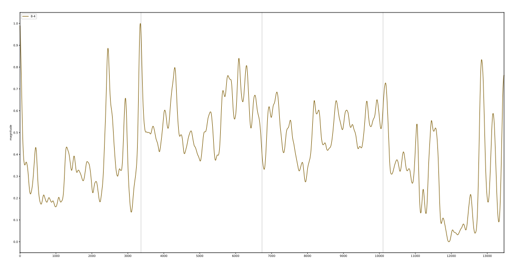
(b) Team Member 8_1



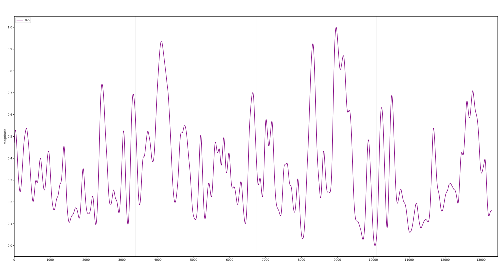
(c) Team Member 8_2



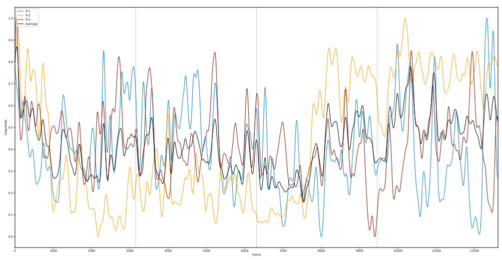
(d) Team Member 8_3



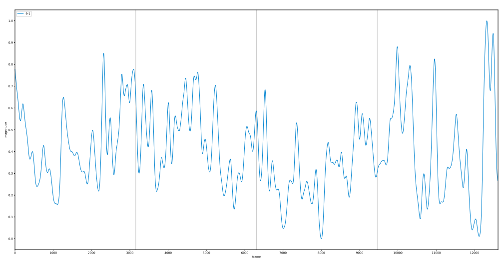
(e) Team Member 8_4



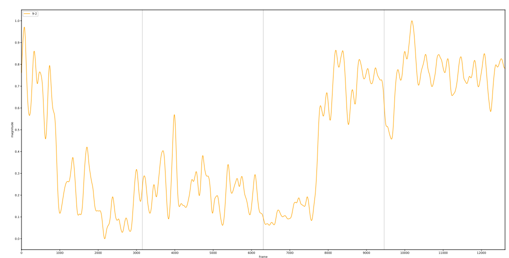
(f) Team Member 8_5



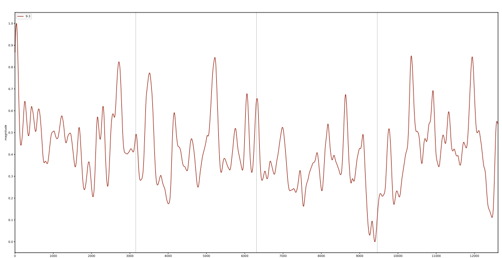
(a) Team 9



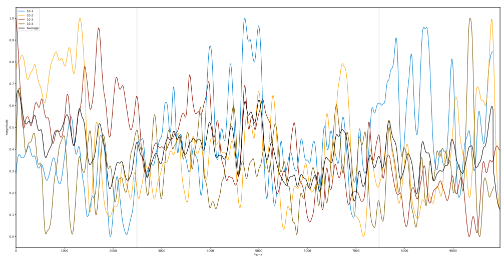
(b) Team Member 9_1



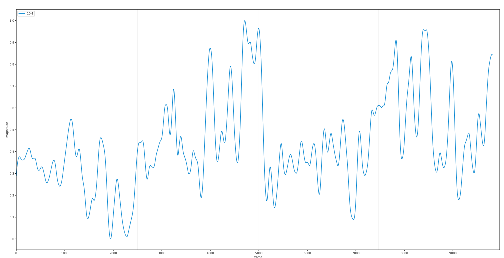
(c) Team Member 9_2



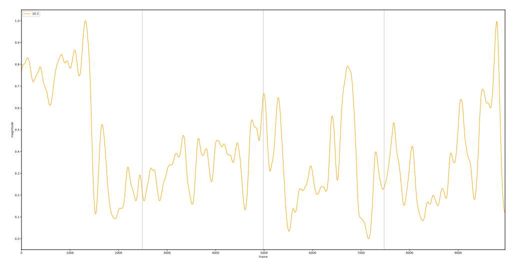
(d) Team Member 9_3



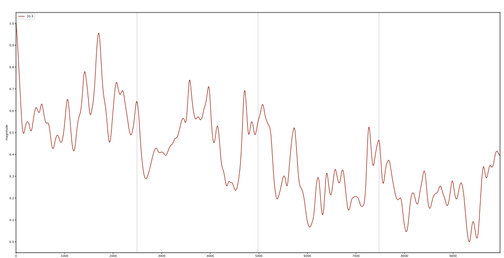
(a) Team 10



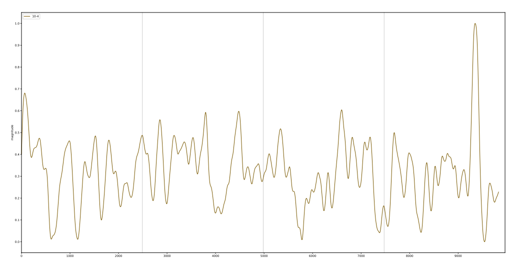
(b) Team Member 10_1



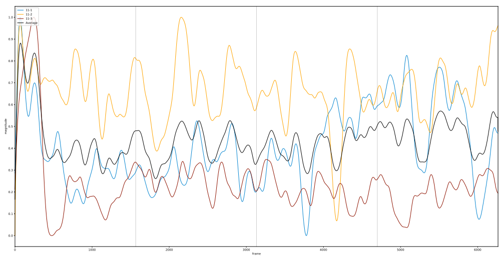
(c) Team Member 10_2



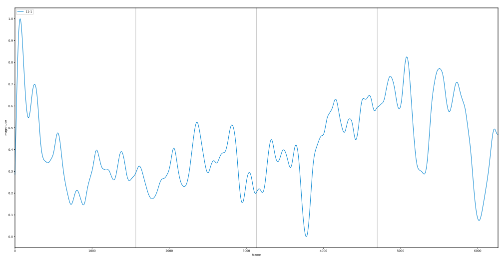
(d) Team Member 10_3



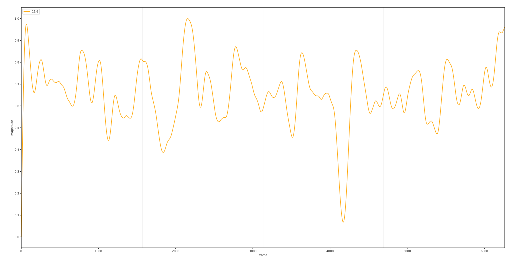
(e) Team Member 10_4



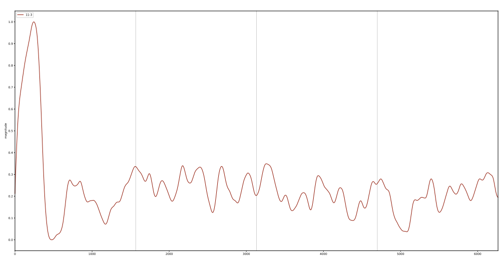
(a) Team 11



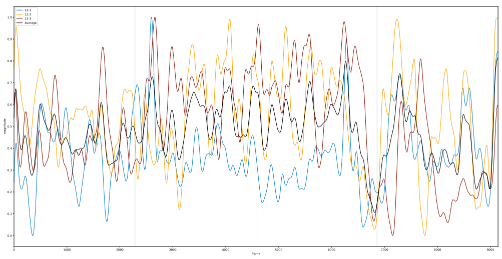
(b) Team Member 11_1



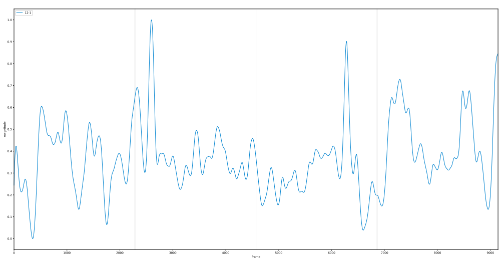
(c) Team Member 11_2



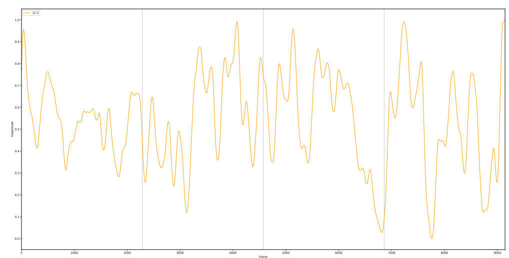
(d) Team Member 11_3



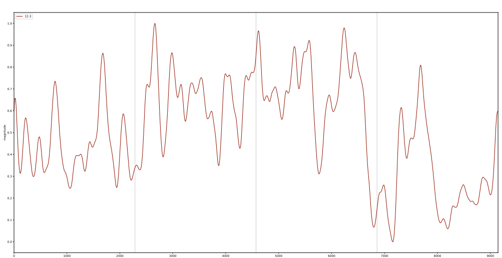
(a) Team 12



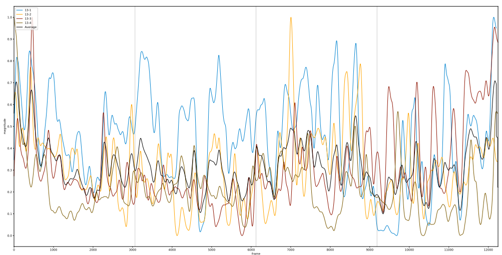
(b) Team Member 12_1



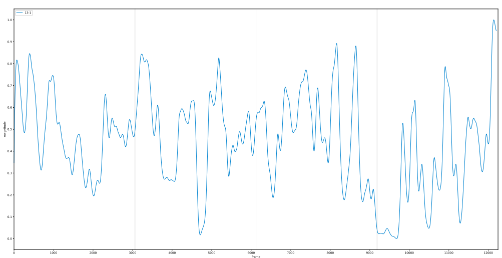
(c) Team Member 12_2



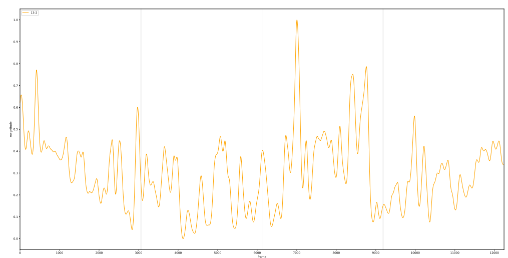
(d) Team Member 12_3



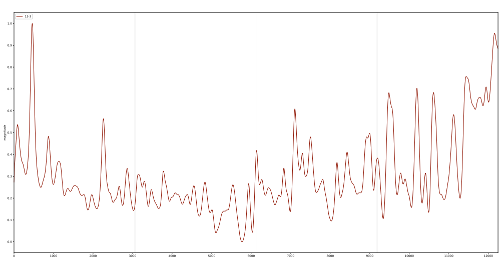
(a) Team 13



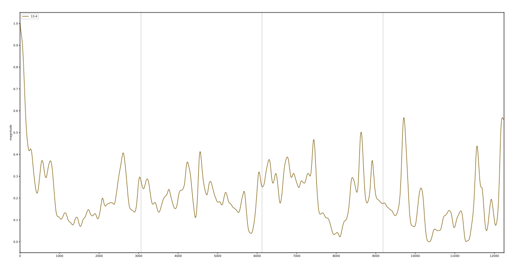
(b) Team Member 13_1



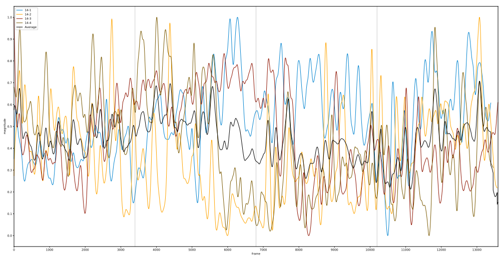
(c) Team Member 13_2



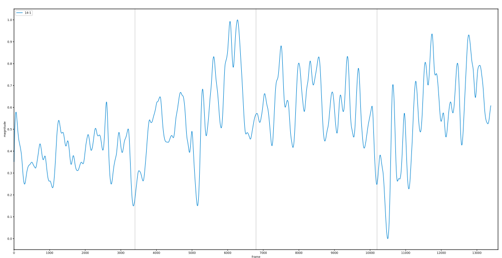
(d) Team Member 13_3



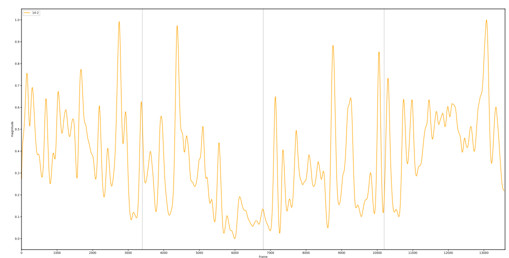
(e) Team Member 13_4



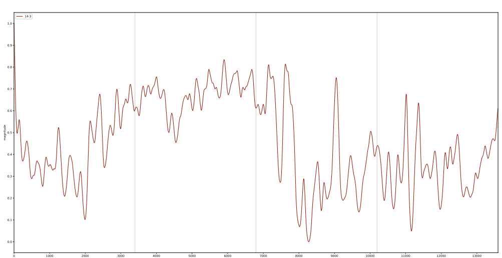
(a) Team 14



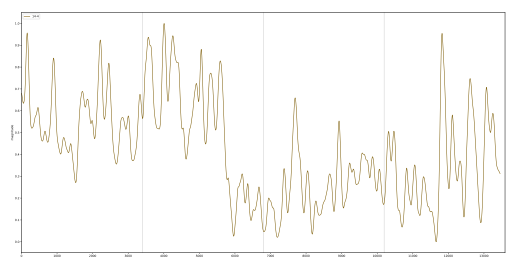
(b) Team Member 14_1



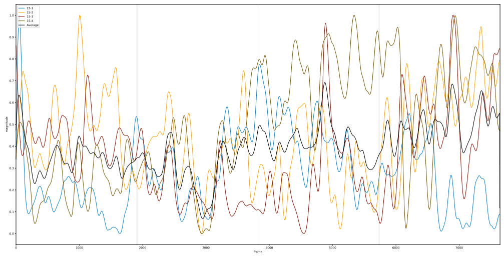
(c) Team Member 14_2



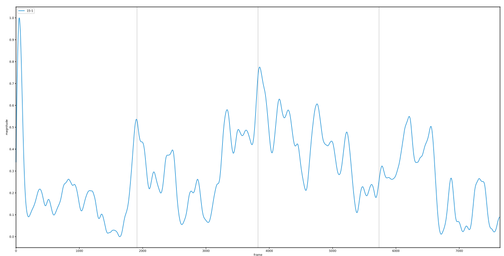
(d) Team Member 14_3



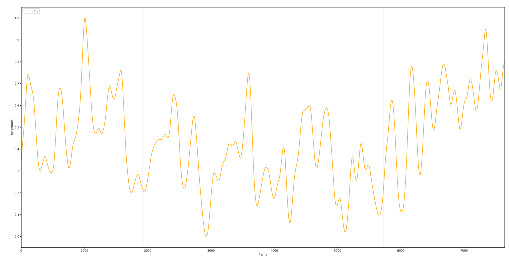
(e) Team Member 14_4



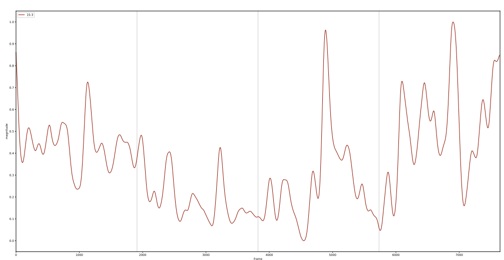
(a) Team 15



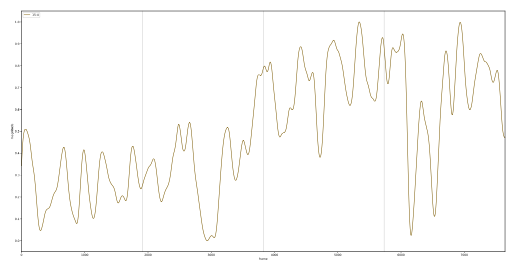
(b) Team Member 15_1



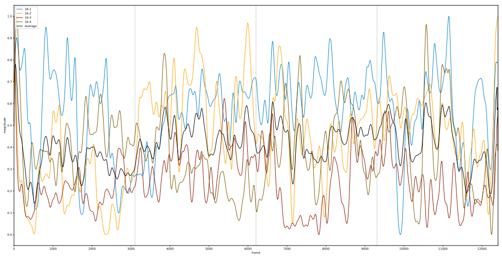
(c) Team Member 15_2



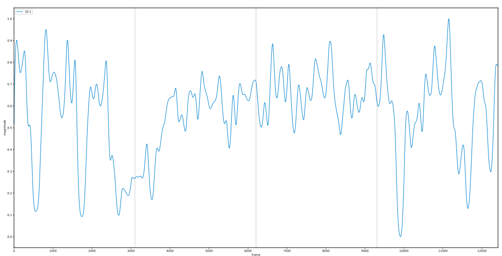
(d) Team Member 15_3



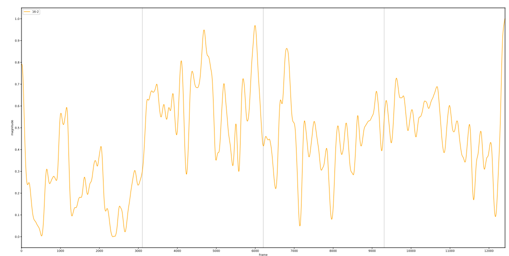
(e) Team Member 15_4



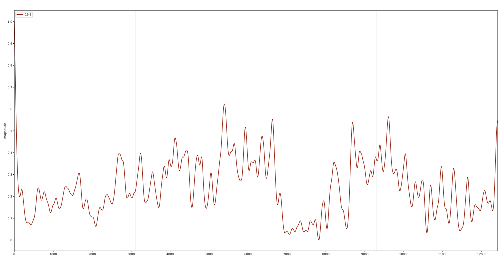
(a) Team 16



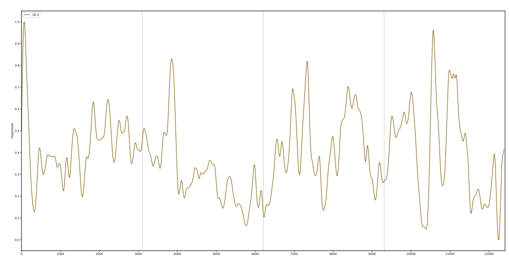
(b) Team Member 16_1



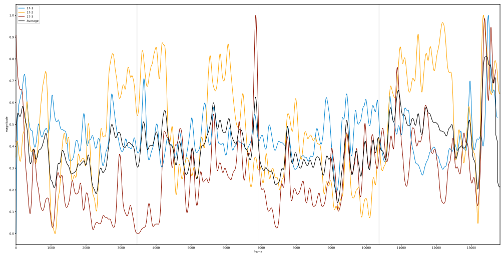
(c) Team Member 16_2



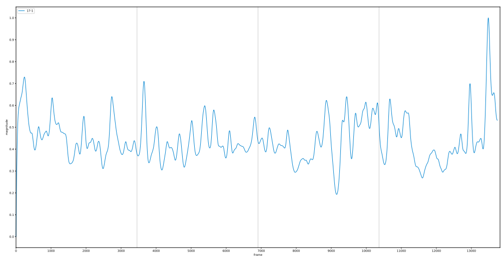
(d) Team Member 16_3



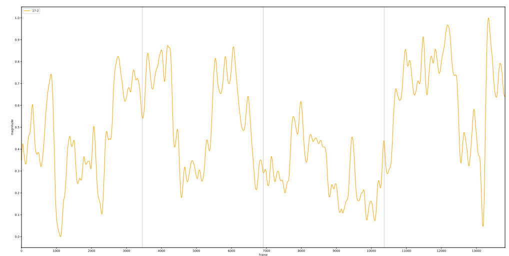
(e) Team Member 16_4



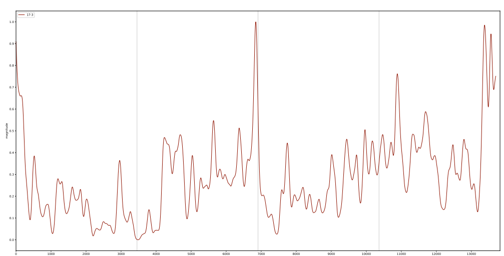
(a) Team 17



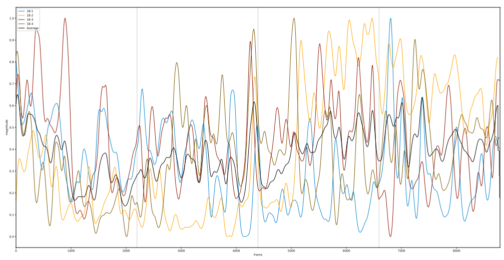
(b) Team Member 17_1



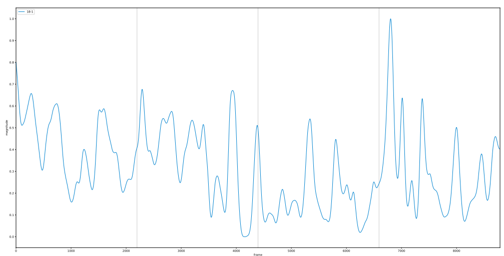
(c) Team Member 17_2



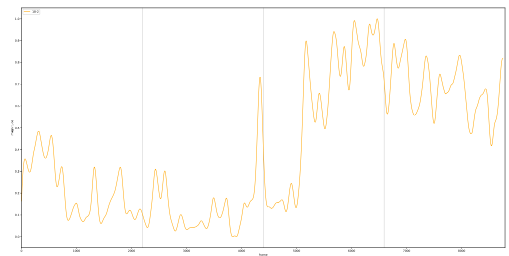
(d) Team Member 17_3



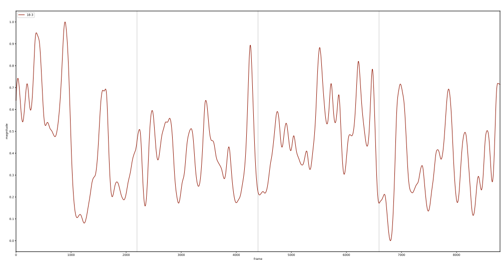
(a) Team 18



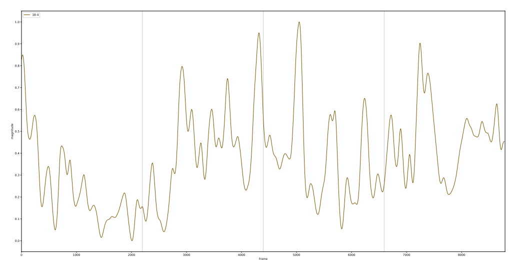
(b) Team Member 18_1



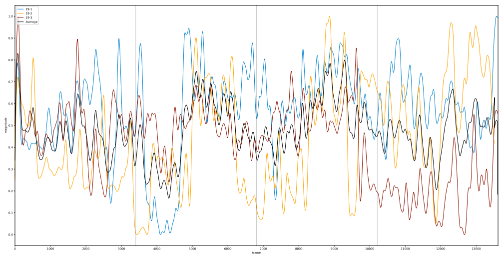
(c) Team Member 18_2



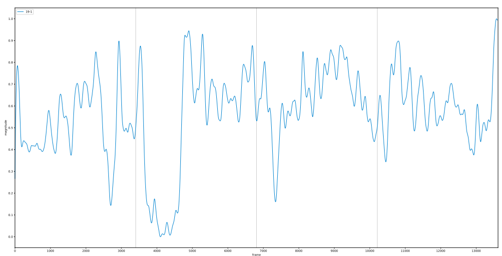
(d) Team Member 18_3



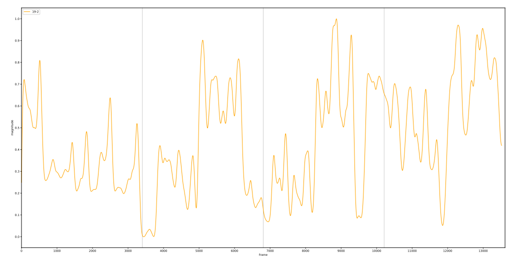
(e) Team Member 18_4



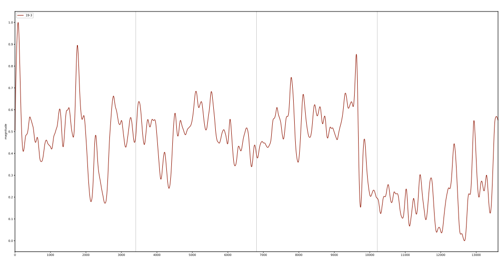
(a) Team 19



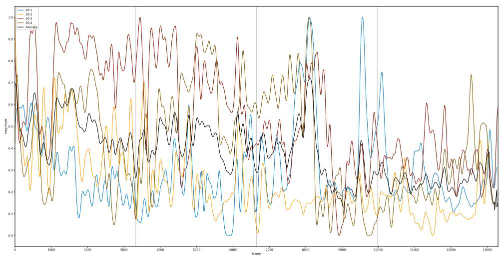
(b) Team Member 19_1



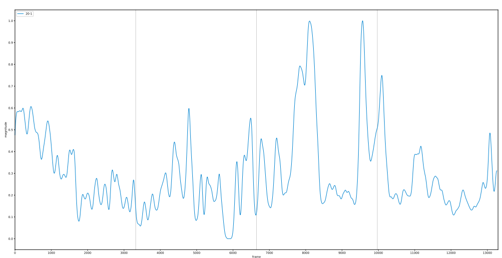
(c) Team Member 19_2



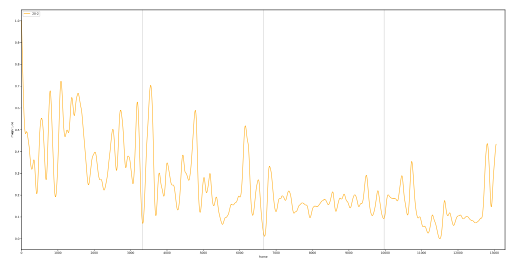
(d) Team Member 19_3



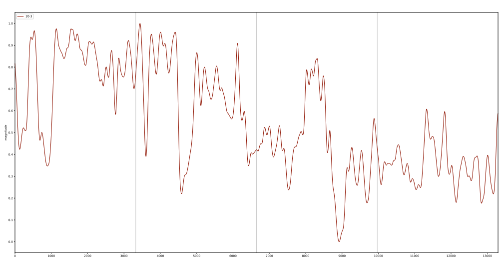
(a) Team 20



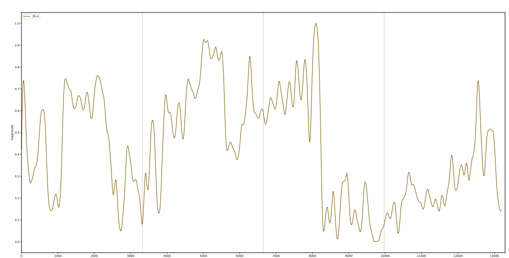
(b) Team Member 20_1



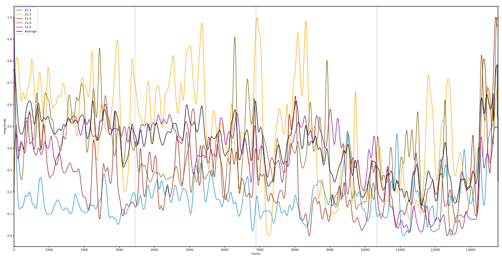
(c) Team Member 20_2



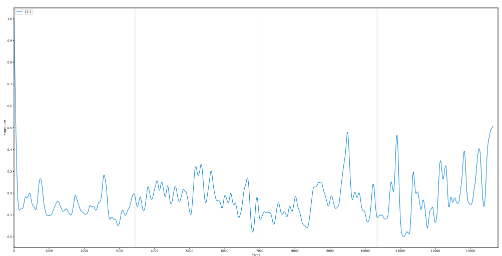
(d) Team Member 20_3



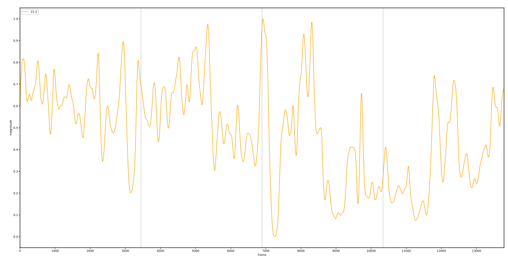
(e) Team Member 20_4



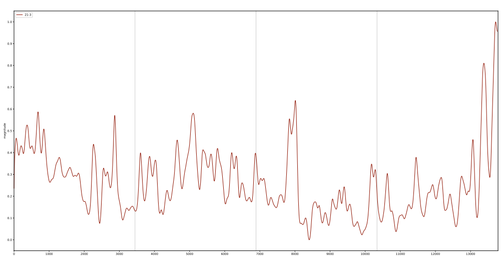
(a) Team 21



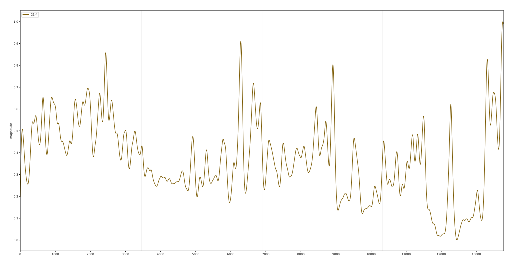
(b) Team Member 21_1



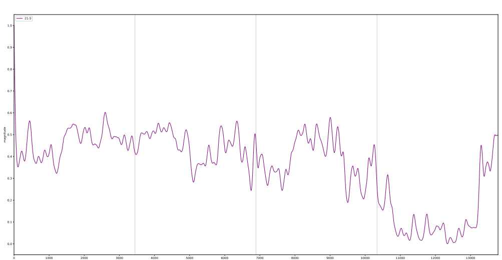
(c) Team Member 21_2



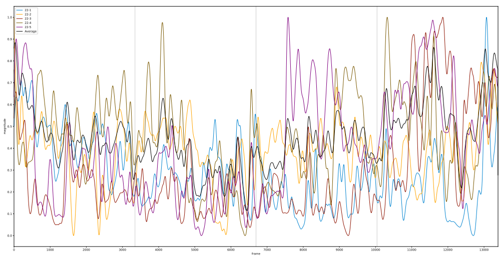
(d) Team Member 21_3



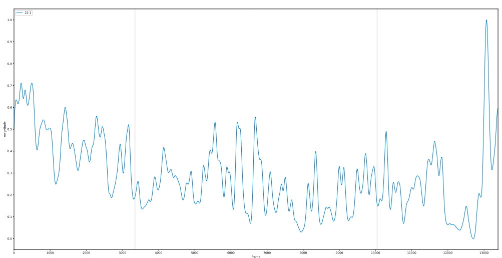
(e) Team Member 21_4



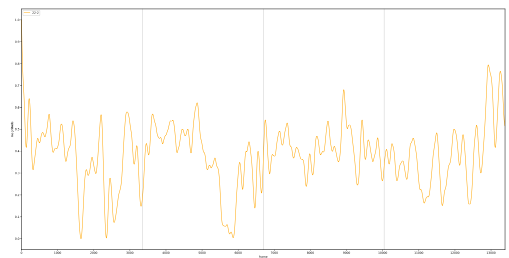
(f) Team Member 21_5



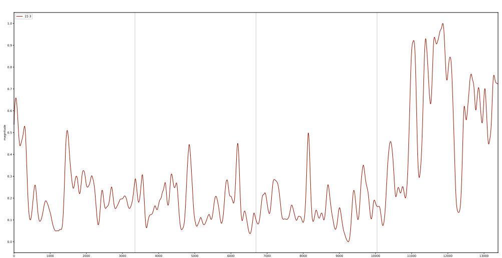
(a) Team 22



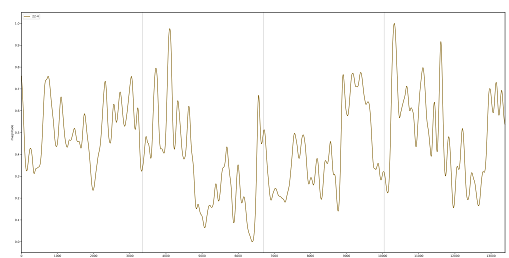
(b) Team Member 22_1



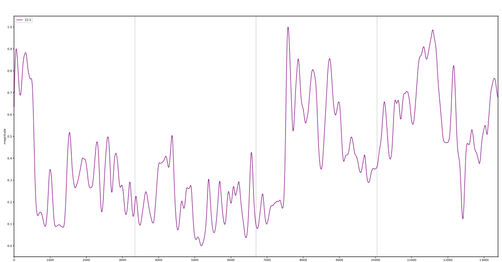
(c) Team Member 22_2



(d) Team Member 22_3



(e) Team Member 22_4



(f) Team Member 22_5

LIST OF REFERENCES

[1] Allsop, Jamie S.. 2016. Coordination and Collective Performance: Cooperative Goals Boost Interpersonal Synchrony and Task Outcomes. *Front Psychol*.

[2] Andrei, Oana. 2018. Interpreting Models of Social Group Interactions in Meetings with Probabilistic Model Checking. *The Group Interaction Frontiers in Technology*

[3] Aran, Oya. 2013. One of a kind: inferring personality impressions in meetings. *In Proceedings of the 15th ACM on International conference on multimodal interaction (ICMI '13)*

[4] Avci, Umut. 2014. Effect of nonverbal behavioral patterns on the performance of small groups. *In Proceedings of the 2014 workshop on Understanding and Modeling Multiparty, Multimodal Interactions (UM3I '14).*

[5] Avci, Umut. 2016. Predicting the Performance in Decision-Making Tasks: From Individual Cues to Group Interaction. *In Proceedings of the 2014 workshop on Understanding and Modeling Multiparty, Multimodal Interactions (UM3I '14).*

[6] Bahr, Michael W.. 2006. The Need for Problem-Solving Teams: Introduction to the Special Issue. *Remedial and Special Education*.

[7] Beyan, Cigdem. 2018. Investigation of Small Group Social Interactions Using Deep Visual Activity-Based Nonverbal Features. *In Proceedings of the 26th ACM international conference on Multimedia (MM '18).*

[8] Bhattacharya, Indrani. 2019. The unobtrusive group interaction (UGI) corpus. *In Proceedings of the 10th ACM Multimedia Systems Conference (MMSys '19).*

[9] Bonthu, Harika. 2021. Rule-Based Sentiment Analysis in Python for Data Scientists. *Analytics Vidhya*.

[10] Börner, Katy. 2010. A Multi-Level Systems Perspective for the Science of Team Science. *Science translational medicine*.

[11] Braley, McKenzie. 2018. The Group Affect and Performance (GAP) Corpus. *In Proceedings of the Group Interaction Frontiers in Technology (GIFT'18)*.

[12] Campaign_Creators. Free Image on Pixabay - Work, Team, Office, Meeting. https://pixabay.com/photos/work-team-office-meeting-teamwork-3828297/.

[13] Carletta, Jean. 2006. AMI Corpus Annotations. *The ELRA Newsletter 11(1)*.

[14] Carletta, Jean. 2006. AMI Corpus Meeting Rooms. *The ELRA Newsletter 11(1)*.

[15] Carletta, Jean. 2006. AMI Corpus Overview. *The ELRA Newsletter 11(1)*.

[16] Carletta, Jean. 2006. The AMI Meeting Corpus: A Pre-announcement. *The ELRA Newsletter 11(1)*.

[17] Collins, Allan. 1989. Cognitive Apprenticeship. *American Educator*.

[18] Dale, Rick. 2020. Body synchrony in triadic interaction.. *Royal Society*.

[19] Dehghani, Mohammad. 2021. Teamwork Optimization Algorithm: A New Optimization Approach for Function Minimization/Maximization. *Sensors*.

[20] Delaherche, Emilie. 2012. Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Transactions on Affective Computing*.

[21] Eskevich, Maria. 2014. Exploring speech retrieval from meetings using the AMI corpus. *Computer Speech & Language*

[22] Fang, Sheng. 2016. Personality classification and behaviour interpretation: an approach based on feature categories. *In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*.

[23] Farnebäck, Gunnar. 2003. Two-Frame Motion Estimation Based on Polynomial Expansion. *Image Analysis*.

[24] Fiore, Stephen M. 2010. Collaborative problem-solving education for the twenty-first-century workforce. *Nature Human Behaviour*.

[25] Fiore, Stephen M. 2018. Toward an Understanding of Macrocognition in Teams: Predicting Processes in Complex Collaborative Contexts. *Human Factors*

[26] GeeksforGeeks. 2020. OpenCV - The Gunnar-Farneback optical flow. https://www.geeksforgeeks.org/opencv-the-gunnar-farneback-optical-flow/.

[27] Graesser, Arthur C.. 2018. Advancing the Science of Collaborative Problem Solving. *Psychological Science in the Public Interest*.

[28] Hsueh, Pei-Yun. 2008. Automatic Decision Detection in Meeting Speech. *Machine Learning for Multimodal Interaction*.

[29] Hsueh, Pei-yun. 2006. Automatic topic segmentation and labeling in multiparty dialogue. *2006 IEEE Spoken Language Technology Workshop*.

[30] IBM Cloud Education. 2021. What is Natural Language Processing?. https://www.ibm.com/cloud/learn/natural-language-processing.

[31] Jayagopi, Dineshbabu. 2012. Linking speaking and looking behavior patterns with group composition, perception, and performance. *In Proceedings of the 14th ACM international conference on Multimodal interaction (ICMI '12)*.

[32] Jayagopi, Dinesh Babu. 2008. Predicting the dominant clique in meetings through fusion of nonverbal cues. *In Proceedings of the 16th ACM international conference on Multimedia (MM '08)*.

[33] Jones, Gareth J. F.. 2010. Towards methods for efficient access to spoken content in the ami corpus. *In Proceedings of the 2010 international workshop on Searching spontaneous conversational speech (SSCS '10).*

[34] Joshi, Preetam. 2021. RBO v/s Kendall Tau to compare ranked lists of items. *Towards Data Science*,

[35] Jovanovic, Natasa. 2006. A corpus for studying addressing behaviour in multi-party dialogues. *Language Resources and Evaluation*.

[36] Kilgour, Jonathan. 2006. NITE XML Toolkit - Edinburgh Home Page. https://groups.inf.ed.ac.uk/nxt/.

[37] Kindiroglu, Ahmet Alp. 2017. Multi-domain and multi-task prediction of extraversion and leadership from meeting videos. *EURASIP Journal on Image and Video Processing*

[38] Knowledge Transfer 2020. How to Scale data into the 0-1 range using Min-Max Normalization.. https://androidkt.com/how-to-scale-data-to-range-using-minmax-normalization/.

[39] Koutsombogera, Maria. 2019. Observing Collaboration in Small-Group Interaction. *Multimodal Technologies and Interaction*.

[40] Kubasova, Uliyana.2019. 2019. Analyzing Verbal and Nonverbal Features for Predicting Group Performance. *arXiv*

[41] Kubasova, Uliyana. 2020. Group Performance Prediction with Limited Context. *In Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion).*

[42] Lai, Catherine. 2018. Predicting group satisfaction in meeting discussions. *In Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data (MCPMD '18).*

[43] Lakens, Daniel. 2010. Movement synchrony and perceived entitativity. *Journal of Experimental Social Psychology*.

[44] Leamy, Paul. 2019. Re-annotation of cough events in the AMI corpus. *2019 30th Irish Signals and Systems Conference (ISSC)*.

[45] Letsky, Michael P.. 2017. Macrocognition in Teams: Theories and Methodologies. *CRC Press*

[46] Likens, Aaron D.. 2021. Windowed multiscale synchrony: modeling time-varying and scale-localized interpersonal coordination dynamics.. *Social Cognitive and Affective Neuroscience*.

[47] Lin, Yun-Shao. 2020. Predicting Performance Outcome with a Conversational Graph Convolutional Network for Small Group Interactions. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[48] Luotsinen, Linus J.. 2008. Role-based teamwork activity recognition in observations of embodied agent actions. *In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 2 (AAMAS '08)*.

[49] Mathieu, J. E., Luciano. 2020. The development and construct validity of a team processes survey measure. *Organizational Research Methods*.

[50] Miles, Lynden K.. 2017. Coordination Matters: Interpersonal Synchrony Influences Collaborative Problem-Solving. *Psychology*.

[51] Miura, Go. 2019. Task-independent Multimodal Prediction of Group Performance Based on Product Dimensions. *In 2019 International Conference on Multimodal Interaction (ICMI '19)*.

[52] Nisbett, Richard E.. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*.

[53] Okada, Shogo. 2016. Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. *In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*.

[54] Okada, Shogo. 2019. Modeling Dyadic and Group Impressions with Intermodal and Interperson Features. *ACM Trans. Multimedia Comput. Commun. Appl. 15*.

[55] Okada, Shogo. 2015. Personality Trait Classification via Co-Occurrent Multiparty Multimodal Event Discovery. *In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*.

[56] OpenCV: Optical Flow. https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html.

[57] Paxton, Alexandra. 2013. Argument disrupts interpersonal synchrony. *Q J Exp Psychol (Hove)*.

[58] Pipsels. Team whiteboard brainstorm photo, Women, Men, Business, Friends, Meetings, Marketing, Presentation — Piqsels. https://www.piqsels.com/en/public-domain-photo-zkhnc.

[59] PositivePsychology.com. 2016. The Psychology of Teamwork: The 7 Habits of Highly Effective Teams. *Positive Workplace*.

[60] Raaijmakers, Stephan. 2008. Multimodal subjectivity analysis of multiparty conversation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP '08*.

[61] Salas, Eduardo. 2015. Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*.

[62] , An audio visual corpus for emergent leader analysis. 2011. http://infoscience.epfl.ch/record/192622.

[63] Sanchez-Cortes, Dairazalia. 2012. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups, *IEEE Transactions on Multimedia*,

[64] Sanchez-Cortes, Dairazalia. 2013. Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *Journal on Multimedia User Interfaces*,

[65] Sanchez-Cortes, Dairazalia. 2012. ELEA. https://www.idiap.ch:/en/dataset/elea/index_html,

[66] Sanchez-Cortes, Dairazalia. 2010. Identifying emergent leadership in small groups using non-verbal communicative cues. In International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10).

[67] Sapru, Ashtosh. 2012. Automatic speaker role labeling in AMI meetings: Recognition of formal and social roles. *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[68] Schiff, Lauren. 2018. Teaching and Sustaining a Shared Mental Model for Intraoperative Communication and Teamwork. *Obstetrics & Gynecology*.

[69] Shen, Zhihao. 2020. Understanding nonverbal communication cues of human personality traits in human-robot interaction. *IEEE/CAA Journal of Automatica Sinica*.

[70] Statistics Solutions. Table of Critical Values: Pearson Correlation. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/pearsons-correlation-coefficient/table-of-critical-values-pearson-correlation/.

[71] Stewart, Angela EB. 2021. Multimodal modeling of collaborative problem-solving facets in triads.. *User Modeling and User-Adapted Interaction*.

[72] Subburaj, Shree Krishna. 2020. Multimodal, Multiparty Modeling of Collaborative Problem Solving Performance. *ICMI '20: INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*.

[73] Sun, Xiaofan. 2011. A Multimodal Database for Mimicry Analysis. *Affective Computing and Intelligent Interaction*

[74] TeachingEnglish | British Council | BBC. Prosodic features. https://www.teachingenglish.org.uk/article/prosodic-features.

[75] TMT Group Inc. Careers – Technology Management Training Group. https://www.tmtgroupinc.com/jobs/.

[76] Tur, Gokhan. 2010. The CALO Meeting Assistant System. *IEEE Transactions on Audio, Speech, and Language Processing*.

[77] University of Colorado. Team Effectiveness Questionnaire. https://www.cu.edu/sites/default/files/Team_effectiveness_questionnaire.pdf.

[78] Valente, Fabio. 2012. Annotation and recognition of personality traits in spoken conversations from the ami meetings corpus. *INTERSPEECH 2012*.

[79] Valente, Fabio. 2011. Language-independent socioemotional role recognition in the ami corpus. *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*.

[80] van Leeuwen, David A.. 2006. The AMI Speaker Diarization System for NIST RT06s Meeting Data. *Machine Learning for Multimodal Interaction*.

[81] Vinciarelli, Alessandro. 2017. Introduction: Social Signal Processing. *Cambridge University Press*.

[82] Vinciarelli, Alessandro. 2011. Understanding social signals in multi-party conversations: Automatic recognition of socio-emotional roles in the AMI meeting corpus. *2011 IEEE International Conference on Systems, Man, and Cybernetics*.

[83] Wiltshire, Travis J.. 2019. Multiscale movement coordination dynamics in collaborative team problem solving.. *Applied Ergonomics*.

[84] Wong, Bonnie. 2018. Tips on How to Avoid People Pleasing in Group Projects. *Study Break*

[85] Yanchus, Nancy J.. 2017. 'You just can't do it all': a secondary analysis of nurses' perceptions of teamwork, staffing and workload. *Journal of Research in Nursing*.

[86] Zamalloa, Maider. 2010. Low-latency online speaker tracking on the AMI Corpus of meeting conversations. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*.

[87] Zhang, Lingyu. 2019. Improved Visual Focus of Attention Estimation and Prosodic Features for Analyzing Group Interactions. *In 2019 International Conference on Multimodal Interaction (ICMI '19)*.