Spoken Dialogue In Face-to-Face And Remote Collaborative Learning Environments

by

Arun Reddy Nelakurthi

A Thesis Presented in Partial Fulfillment
of the Requirement for the Degree
Master of Science

Approved July 2014 by the
Graduate Supervisory Committee:

Heather Pon-Barry, Chair
Kurt VanLehn
Erin Walker

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

Research in the learning sciences suggests that students learn better by collaborating with their peers than learning individually (Chi et al., 2008). Students working together as a group tend to generate new ideas more frequently and exhibit a higher level of reasoning (Johnson and Johnson, 2002). In this internet age with the advent of massive open online courses (MOOCs), students across the world are able to access and learn material remotely. This creates a need for tools that support distant or remote collaboration. In order to build such tools we need to understand the basic elements of remote collaboration and how it differs from traditional face-to-face collaboration.

The main goal of this thesis is to explore how spoken dialogue varies in face-to-face and remote collaborative learning settings. Speech data is collected from student participants solving mathematical problems collaboratively on a tablet. Spoken dialogue is analyzed based on conversational and acoustic features in both the settings. Looking for collaborative differences of transactivity and dialogue initiative, both settings are compared in detail using machine learning classification techniques based on acoustic and prosodic features of speech. Transactivity is defined as a joint construction of knowledge by peers. The main contributions of this thesis are: a speech corpus to analyze spoken dialogue in face-to-face and remote settings and an empirical analysis of conversation, collaboration, and speech prosody in both the settings. The results from the experiments show that amount of overlap is lower in remote dialogue than in the face-to-face setting. There is a significant difference in transactivity among strangers. My research benefits the computer-supported collaborative learning community by providing an analysis that can be used to build more efficient tools for supporting remote collaborative learning.

DEDICATION

To Savitha, my parents and all those friends
who made my life special.

ACKNOWLEDGEMENTS

I would like to express my most sincere thanks and gratitude to my advisor Dr. Heather Pon-Barry, who has guided me throughout my graduate study and gave me the freedom and support to conduct this research. Working with her has been one of the most enriching experience both professionally and personally.

I am also grateful to my thesis committee members, Dr. Kurt Vanlehn and Dr. Erin Walker for their valuable time serving on my thesis committee.

I would like to thank Nichola Lubold and Sree Aurovindo, my colleagues, for their invaluable inputs and help with data collection. I am also thankful to Sachin Grover at Dragoon lab for his immense help during the course of my research. I would like to express my gratitude to Michael Sullivan and all the staff members at the Hispanic Research Center for their help and support.

I am thankful for my friends Savitha, Archana, Prashant, Vishnu, Shruti and Vissu for their encouragement. I am also thankful to all my faculty, friends and seniors at Birla Institute of Technology for making me who I am today.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Research in the learning sciences suggests that students learn better by collaborating with their peers than learning individually (Chi et al., 2008). Students working together as a group tend to generate new ideas more frequently and exhibit a higher level of reasoning (Johnson and Johnson, 2002). In this internet age with the advent of massive open online courses (MOOCs), students across the world are able to access and learn material remotely. This creates a need for tools that support distant or remote collaboration. In order to build such tools we need to understand the basic elements of remote collaboration and how it differs from traditional face-to-face collaboration.

This thesis examines how spoken dialogue varies in face-to-face and remote collaborative learning environments. The speech corpus is collected from students collaboratively solving math problems on a tablet. The application is designed to support and provide formative assessment for K-12 students solving mathematical problems. The participants in the study are undergraduate students from Arizona State University with basic knowledge of algebra and geometry. A total of 40 native English speaking American students were recruited for the experiment. High quality speech data from both the participants in the experiment is collected for further analysis.

First, we split speech data into turns and annotate for conceptual, surface and application level reasoning. We mark for transactivity as a measure of learning. Knowledge co-construction from (Hausmann et al., 2004) is called as transactivity in this thesis defining transactivity as the joint construction of knowledge by peers. Turns are also marked for dialogue initiative (Walker and Whittaker, 1990) as a measure of collaboration. Dialogue initiative tracks the leading participant in the conversation and thereby determines the current conversational focus. We also calculate conversational features like number of turns, turn duration, amount of overlaps. We extract acoustic and prosodic features including MFCC's, linear spectral frequencies, pitch, shimmer and jitter automatically using openSMILE (Eyben et al., 2010) software application. Finally, we build machine learning classifier models using the extracted features to predict transactivity, reasoning and dialogue initiative. The results from the experiments drive our analysis and comparison of spoken dialogue in face-to-face and remote settings. The thesis tries to answer the following questions:

1. How does spoken dialogue differ in a face-to-face and remote settings?
2. Does the amount of collaboration vary in these settings?
3. Does the remote setup pose challenges for collaboration?
4. Is it possible to detect transactive dialogue turns from acoustic features of the speech signal?

Answers to these questions will immensely help the computer supported collaborative learning community to build tools that better support effective remote collaboration. The ability to detect transactive dialogue turns with high accuracy can help the community to build tools which can evaluate students' collaborative learning based on their spoken dialogue.

The main contributions of this thesis are: a speech corpus to analyze spoken dialogue in face-to-face and remote settings and empirical analysis of conversation, collaboration, and speech prosody in both the settings. The speech corpus is transcribed and annotated for transactivity and initiative shifts. Annotation is done based on the coding manual developed for this dataset. Appendix A consists of the coding manual explained in detail with examples. The empirical results explained in Chapter 5, show that strangers feel more comfortable co-constructing in a face-to-face setting. However, based on the amount of contribution and collaborative measures calculated from dialogue initiative, there is no significant difference in the way students collaborate in both the settings. This thesis identifies the important acoustic and prosodic features for predicting transactivity and differentiating face-to-face and remote settings.

The organization of the next chapters are as follows. Chapter 2 gives an overview of the background and relevant research in the areas of collaborative learning, collaborative learning frameworks, assessing collaboration using machine learning techniques and face-to-face versus remote collaboration. The data collection and experimental procedure is explained in Chapter 3. Chapter 4 discusses speech data processing and annotation. Chapter 5 details the results and analysis of the experiments. I conclude my work in Chapter 6.

Chapter 2

BACKGROUND LITERATURE

This chapter surveys existing research in collaborative learning. The following sections provide a background on collaborative learning, frameworks of collaborative learning, assessing collaboration and comparisons of face-to-face and remote collaboration.

2.1 Collaborative Learning

Collaboration is a situation in which two or more people learn or attempt to learn something together (Dillenbourg, 1999). Chi et al. (2008) shows that collaborative learning is more effective in producing learning gains than individuals working alone. Commonly applied strategies by researchers for collaborative learning are story production, argumentation over an issue and problem-solving (Dillenbourg, 1999). Recent increasing interest in technology-supported collaborative learning in education has propelled the development of powerful and engaging learning environments (Oblinger et al., 2005). This has led to active research and development of new tools that support rich collaborative learning (Resta and Laferrière, 2007; Kersey et al., 2009).

According to (Hausmann et al., 2004), collaborative learning happens due to three possible mechanisms: other-directed explaining, co-construction and self-directed explaining. Other-directed explaining occurs when one peer instructs or explains another partner on how to solve a problem. Co-construction occurs when a peer shares knowledge, which can then be criticized or elaborated on by his/her partner. Self-directed explaining is to learn by listening to someone else self-explain. Learning from another person's explanation is similar to learning from a worked-out example, however in a collaborative setting the source of the worked-out example is not a textbook but a peer. In Hausmann et al. (2004), students with limited knowledge of physics, worked in pairs to solve problems in the domain of kinematics. All three mechanisms had an impact on learning but to different degrees. Self-directed explaining produced the strongest learning gains, occurring in 71% of self-directed explaining episodes, but this was only for the speaker. Other-directed explaining led to learning gains of 45% for the listener. Co-construction although relatively infrequent led to increased problem solving performance. It was beneficial to both the speaker and the listener when it occurred.

Hausmann et al. (2004) and Gweon (2012) proposed different co-construction frameworks to model task-oriented collaborative learning. Hausmann et al. (2004) proposed a knowledge co-construction framework (KCC) in which co-construction is defined as a joint construction of knowledge by peers. According to Hausmann, the process of constructing knowledge may proceed in a variety of ways, but the most natural way of collaboration for peers is to either elaborate or critically evaluate their partners contributions. In elaborative co-construction one partner adds a significant contribution to the discourse that develops on another persons idea. In the case of critical co-construction the dialogue between the dyads contains conflicts. The participants critically argue and reason to reach an agreement on the problem.

Gweon (2012) proposed the Idea Co-Construction (ICC) framework. ICC is the process of taking up, transforming, or otherwise building an idea expressed earlier in a conversation. According to ICC the instances of spoken dialogue are considered as co-constructive if they have explicit evidence of reasoning, either of an individual's own reasoning or where the individual builds upon prior reasoning statements in the discussion. The difference between KCC and ICC is that, the reasoning statements in KCC are based on the problem solving task, whereas the reasoning statements in ICC have to exhibit a strict compare and contrast or cause and affect relationship. In this thesis, I adopt the Hausmann et al. (2004) knowledge co-construction framework to analyze collaboration among students.

Another good measure of collaboration is initiative. According to (Walker and Whittaker, 1990), conversation is bidirectional; there is a two way flow of information between participants. Information is exchanged by mixed-initiative. As the initiative passes back and forth between the participants, the control over the conversation gets transferred from one participant to the other. Walker and Whitaker analyzed task oriented dialogues between experts and non-experts, the results showed that experts had control about 90% of utterances. While Walker and Whitaker claim that initiative encompasses both a dialogue and task initiative, Chu-Carroll and Brown (1997) claim that dialogue and task initiative are different. According to Chu-Carroll and Brown (1997), the dialogue initiative tracks who is leading the conversation and determines the current conversational focus. While the task initiative tracks the leader in the development of a plan to achieve a problem solving goal. Chu-Carroll and Brown (1997) proposed a model that tracks initiative shifts between participants. Initiative shifts tracks how the control switches from one participant to the other in a conversation. The model could correctly predict task initiative holders in 99.1% of turns and dialogue initiative holders in 87.8% of turns in the corpus on which the model was trained. In this thesis, we consider only dialogue initiative as defined by (Walker and Whittaker, 1990) as a measure of collaboration in spoken dialogue.

The concepts of control and initiative shifts in dialogue initiative can play a part in recognizing transactivity. Based on (Walker and Whittaker, 1990), one or more threads of control pass between participants in a dialogue. This implies that tracking transfer control can be useful in determining when transactivity is occurring. In theory, transfer of control from one participant to other indicates that they are working together collaboratively to solve the problem and also to co-construct knowledge.

Assessing pedagogical content using machine learning is an active area of research (Soller and Lesgold, 2003; McLaren et al., 2007; Gweon et al., 2009; Jain et al., 2012; Gweon et al., 2013; Martinez-Maldonado et al., 2013). Soller and Lesgold (2003) demonstrated how to analyze online knowledge sharing interactions to support collaborative distance learners. Utilizing Hidden Markov Model clustering and Multidimensional scaling Soller and Lesgold analyzed and distinguished between effective and ineffective student knowledge sharing interactions.

McLaren et al. (2007) made use of machine learning classifiers, to evaluate past e-discussions and use the results to provide awareness indicators for teachers and moderators in the context of new e-discussions. Awareness indicators are intended to alert the moderator of important events in the e-discussion, such as students not using critical reasoning in their contributions.

Gweon et al. (2009) showed using machine learning how to automatically assess project based learning groups. In the research they proposed a 5 dimensional assessment framework. They used prosodic features extracted from recorded speech from group meetings to predict these dimensions. The results correlated better with an objective observer rating of students than that of the instructor.

Jain et al. (2012) used an unsupervised dynamic Bayesian modeling approach to model speech style accommodation in face-to-face interactions. Speech style accommodation refers to shifts in style that are used to achieve strategic goals with in interactions. Gweon et al. (2013) used the dynamic bayesian modeling approach (Jain et al., 2012) for estimating prevalence other oriented transacts in dyadic situations. Transactive contribution is one where reasoning is made explicit, and where the reasoning builds on a prior reasoning statement within the discussion. Other-oriented transacts are contributions that build on prior contribution of a conversational partner.

Martinez-Maldonado et al. (2013) analyzed students collaboration when working around a multi-touch table top enriched with sensors for identifying users, their actions and their verbal interactions. They have used classification models and hierarchical clustering to identify patterns in student interactions.

2.2 Face-to-face Versus Remote Collaboration

In most learning environments students collaborate in groups to undertake collective tasks like group assignments and projects. Collaborative activities may differ depending on the environment, from face-to-face in the same room to remote, separated by distance. The influence of different collaborative settings on the students' learning outcome still remains unclear. This thesis contributes to the understanding of how conversation in collaborative learning differs in face-to-face versus remote environments.

O'Conaill et al. (1993) showed that video-mediated meetings are characterized by highly formal conversational behaviors compared to face-to-face meetings. Listeners produced fewer backchannels and interrupted less often than in video-mediated meetings. Sellen (1995) compared different video conferencing systems with same-room and audio-only conversations. The results showed that people in the same-room produced more interruptions and fewer-formal handovers of the floor compared to any of the technology mediated conditions. The audio-only and video conferencing conditions were equivalent.

Basque and Pudelko (2004) examined the effect of co-elaborating a knowledge model in dyads at a distance on performance and on learning. A knowledge model is similar to a concept map, in which different types of knowledge objects are represented with different shapes and there is a typology of predefined links to use. The results showed that the collaborative knowledge sharing task is superior to use for dyads working face-to-face when compared to those dyads who communicated asynchronously at a distance using chat software. Remote partners who collaborated synchronously actually learned more than participants in the face-to-face and groups communicating asynchronously at a distance. Tutty and Klein (2008) investigated the impact of online and face-to-face collaboration on the students learning outcomes using post-tests and a total projects assessment. Online groups appeared to be more efficient than face-to-face groups on the total project, whereas the face-to-face groups were more successful in the post-test procedure.

Chapter 3

DATA COLLECTION METHODOLOGY

This chapter presents the methodology for collection and analysis of a speech corpus of collaborative learning dialogues. Speech data is collected from students collaboratively solving math problems in face-to-face and remote settings.

3.1 Speech Corpus

The speech corpus is collected from students collaboratively solving math problems, each student works on a tablet containing an Android-based Formative Assessment with Computation Technologies (FACT)[1] application. The application is designed to support and provide formative assessment for K-12 students solving mathematical problems. It runs on a touch-based tablet with stylus support that allows free-hand drawing and writing. In all the studies students work in pairs and each student works on their own tablet. The tablet workspace is shared, allowing both the students to simultaneously write and see each others changes.

The mathematical problems that come with the FACT application are part of Mathematics Assessment Project[2]. They are designed with a goal to make knowledge and reasoning visible. The iterative refinement required to solve the problem is intended to generate conversation and drive collaboration.

Given below is a sample math problem. Figure 3.1 is a screenshot of the FACT application. It illustrates participants solving the problem by writing on the movable cards. The writing of different participants is represented by different font colors.

[1] http://fact.engineering.asu.edu/

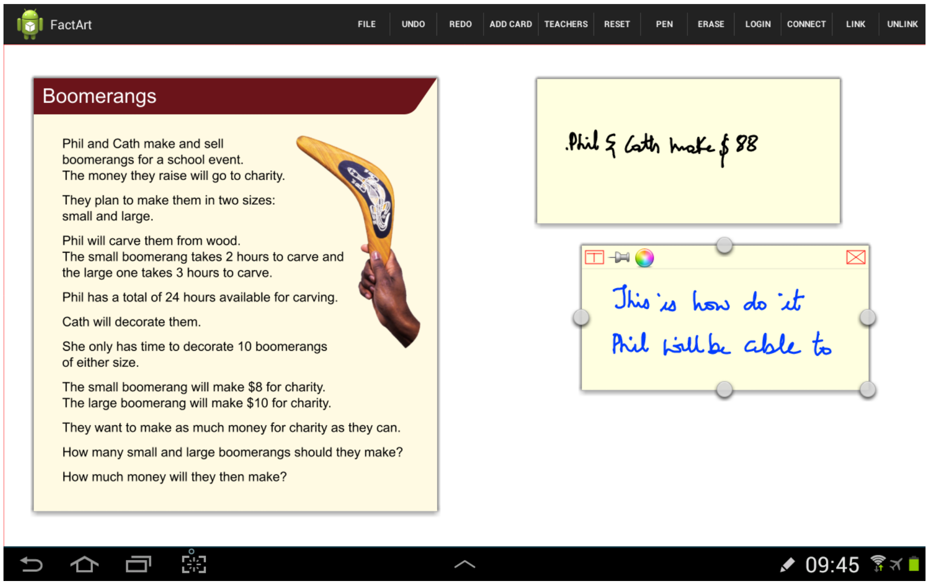
[2] http://map.mathshell.org

Boomerangs Problem

Phil and Cathy make and sell boomerangs for a school event. The money they raise will go to charity. They make them in two sizes, small and large. Phil will carve them from wood. The small one takes 2 hours to carve and the large takes 3 hours. Cathy will decorate them. Phil has 24 hours available for carving, Cathy can decorate only 10 boomerangs. The small boomerang makes $8 for charity. The large boomerang will make $10. They want to make as much for charity as they can.

1. How many small and large boomerangs should they make?

2. How much money will they make?



Boomerangs

Phil and Cath make and sell boomerangs for a school event.
The money they raise will go to charity.

They plan to make them in two sizes: small and large.

Phil will carve them from wood.
The small boomerang takes 2 hours to carve and the large one takes 3 hours to carve.

Phil has a total of 24 hours available for carving.

Cath will decorate them.

She only has time to decorate 10 boomerangs of either size.

The small boomerang will make $8 for charity.
The large boomerang will make $10 for charity.

They want to make as much money for charity as they can.

How many small and large boomerangs should they make?

How much money will they then make?

Phil & Cath make $ 88

This is how do it
Phil will be able to

Figure 3.1: Boomerang Problem.

3.2 Participants

The participants in the study are undergraduate students from Arizona State University with basic knowledge of algebra and geometry. A total of 40 native English speaking American students are recruited for the experiment. Each experiment consists of two participants. A set of 20 students (10 pairs) participated in the face-to-face experiments and another 20 students (10 pairs) in the remote experiments. It is observed that several of the participants in the experiments are friends. In the face-to-face setup, 2 groups are friends and the remaining 8 groups are strangers. In the remote setting 4 groups are friends and the remaining 6 are strangers. Table 3.1 shows the details of the participants.

Scenario	**Strangers**	**Friends**	**Total**
Face-to-face	16 (8 groups)	4 (2 groups)	20
Remote	12 (6 groups)	8 (4 groups)	20

Table 3.1: Distribution of Participants in Face-to-face and Remote Settings.

The students are compensated for their participation in the study. The participants are not trained before the experiment on mathematical topics required to solve the problem.

3.3 Experimental Procedure

The experimental procedure is illustrated in Table 3.2. The total duration for each experiment is 90 minutes with no break. In order to get familiar with the software, participants perform a warm-up activity. After warm-up participants take a pre-test. Participants are then given two math problems to solve. The participants are given more problems if they complete the initial set of problems given to them. The aim is to ensure participants collaborate and solve problems for 40 minutes. The time to solve each problem and the problem solution is recorded for further analysis. At the end participants take a post-test with the same set of questions as the pre-test.

Steps	**Description**	**Length**
Step 1	Consent form and personal information	5 min
Step 2	Pre-test	20 min
Step 1	Fact Application warm-up	5 min
Step 3	Math problem 1	15-20 min
Step 4	Math problem 2	15-20 min
Step 5	Post-test	15 min
Step 6	Questionnaire	5 min

Table 3.2: The Experiment Procedure

The experiments are conducted in a lab on campus in a quiet setting. The assignment of students to different scenarios is done at random. In the face-to-face setup the participants sit together in a room and solve mathematical problems on a tablet. Each participant is given a tablet and the workspace is shared between the participants. Both the participants can write and view each others' changes simultaneously. Speech from both the participants is recorded using unidirectional microphones. The Audacity[3] software application is used to record the audio. Figure 3.2 shows two participants solving the problem in the face-to-face scenario.

In the remote setup the participants sit in different rooms. The communication between two participants is facilitated using Skype, a voice-over-IP service. Both the participants can see and are able to speak to each other. Similar to the face-to-

[3] http://audacity.sourceforge.net

face setup participants solve mathematical problems on the tablet and the workspace is shared between the participants. Figure 3.3 illustrates the procedure involved in recording speech data in the remote setup. Left represents participant sitting in Room 1, Right represents participant in Room 2. A is webcam capturing the work on tablet for analysis. B is webcam used by Skype for communication between participants. The audio from A and B are recorded as stereo - left and right in laptop.

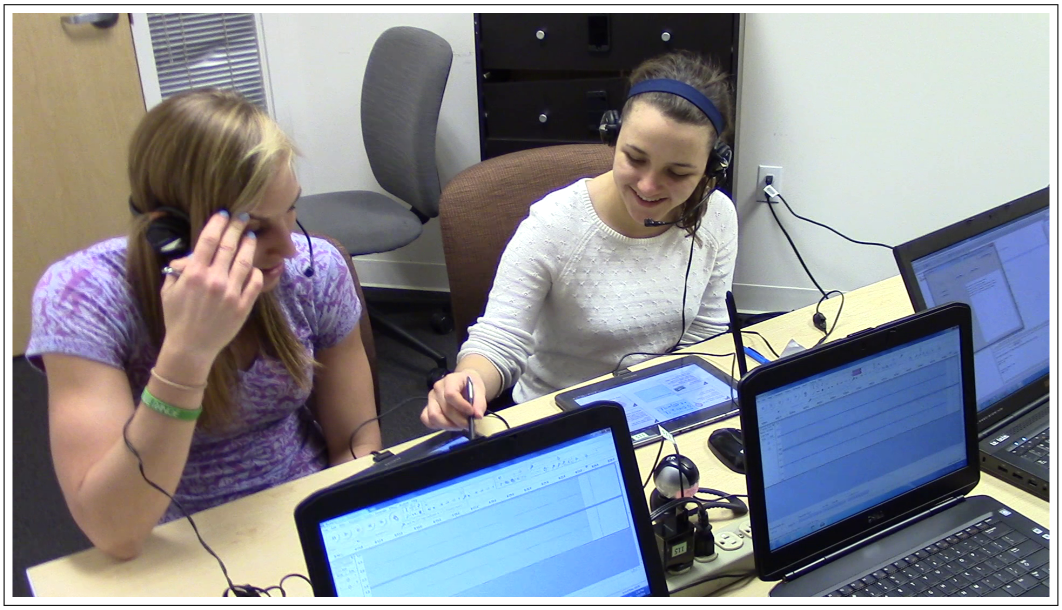


Figure 3.2: Face-to-face Collaboration.

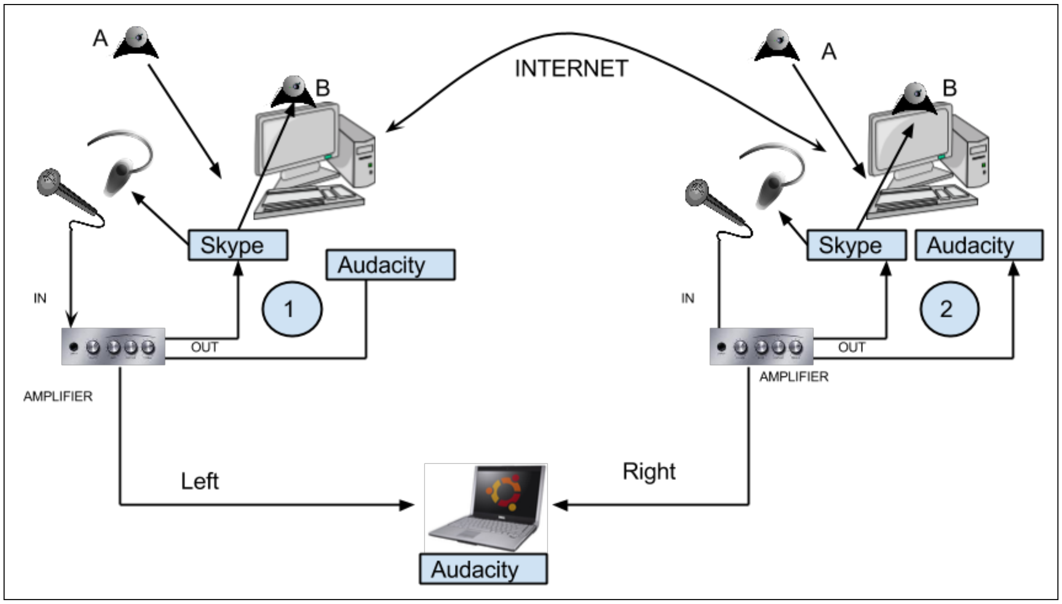


Figure 3.3: Remote Setup Illustration.

During data collection process, human errors while operating tablets led to interruptions in software. The FACT software used for data collection is currently in development stage. Bugs in software led to software crashes interrupting the students solving math problems. Table 3.3 shows the number of interruptions for each experiment. Sessions with more than 5 crashes were removed from the analysis. The average number of interruptions per session for face-to-face is 2 and for remote it is 1.

Experiment	**Face-to-face**	**Remote**
1	3	4
2	4	3
3	5	2
4	2	3
5	2	0
6	1	3
7	0	0
8	0	1
9	2	2
10	4	0
Average	2.3	1.8

Table 3.3: Interruptions per Session in Face-to-face and Remote Settings.

Chapter 4

SPEECH ANNOTATION AND ANALYSIS

This chapter presents the data processing and annotation of the collected speech corpus. The speech data is annotated for transactivity and dialogue initiative. The data is then analyzed using machine learning techniques. The following sections provide insight into the speech data processing, acoustic feature extraction from spoken dialogue and machine learning analysis.

4.1 Speech Data Processing

We consider only the dialogue spoken during problem solving episodes. The speech data of each problem solving episode is marked for speaker turns. A turn, by definition, is a continuous speech utterance with filled pauses by a single speaker (Traum and Heeman, 1997). The turn boundaries are marked using Elan[1]. Figure 4.1 illustrates an annotated experiment. All the turns are transcribed by a professional transcriber using Elan. The turn beginning is marked with the start of an utterance. The turn end is marked when the participant concludes the utterance. Laughter and filled pauses are included and marked in the turns. The turns in which the speech is either not clear or inaudible are marked as inaudible. After marking the turn boundaries, the audio file from each problem solving episode is segmented at the turn level. The distribution of the turns for various settings is shown in Table 4.1.

[1] http://tla.mpi.nl/tools/tla-tools/elan/

Scenarios	Experiments	# Turns	Duration (min)
Face-to-face			
strangers	8	1334	127.8
friends	2	580	44.4
combined	10	1914	172.2
Remote			
strangers	6	1212	111.7
friends	4	1761	104.8
combined	10	2973	216.5

Table 4.1: Distribution of Turns and Duration in Minutes for Different Face-to-face and Remote Settings.

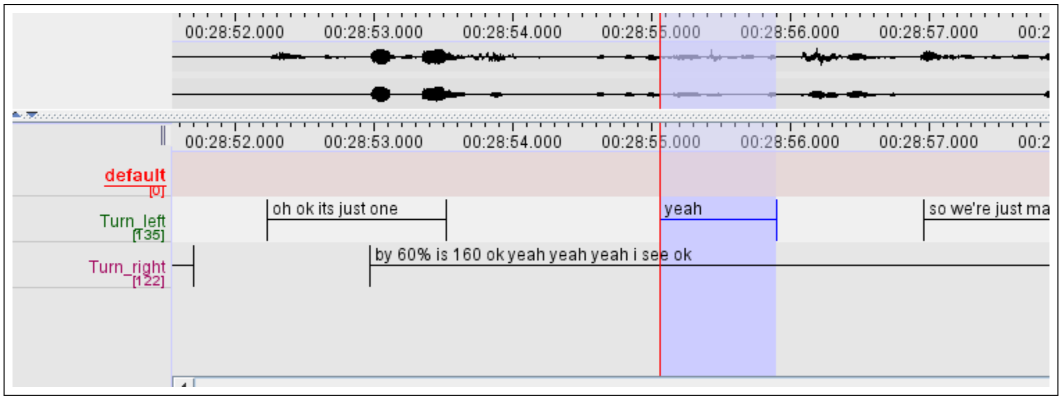


Figure 4.1: A Screenshot of Turn Markings in Elan

4.2 Annotation for Transactivity

Each turn is coded for transactivity and dialogue initiative. Before coding for transactivity, all the turns are coded for reasoning. For a turn to be transactive it has to be a reasoning statement. Reasoning statements can either be at a topic level or surface level. When students converse about a math problem by discussing formulae or the math concepts required to solve the problem, their turns are coded as topic level reasoning statements. Those turns in which only the content of the problem is discussed are marked as surface level reasoning statements. All those turns that are marked as topic level or surface level reasoning are then marked as transactive if there is any knowledge co-construction. If the previous turn is a reasoning turn and is related to the current turn then both the turns are marked as transactive. Appendix A shows the annotation coding manual with detailed examples.

A Transactivity example

(A) Tom is walking away from home.

(B) Lets say jogging, because it said a slightly lower rate when he heads back for home.

(A) Okay then he realizes he lost something and walks back to find it.

(B) He dropped something.

(A) And then runs back.

4.3 Annotation for Dialogue Initiative

The turns are coded for dialogue initiative utilizing utterance-based allocation of control rules (Walker and Whittaker, 1990). Each turn in the dialogue is tagged as either: an assertion, a directive, a question or an acknowledgement. Acknowledgement is usually a turn where no propositional content is expressed. An expert carefully annotates all turns based on transcribed text and listening to the turn audio. A software program allocates the Control based on the following rules:

1. Assertion: Control is allocated to the speaker unless it's a response to a question.

2. Directive: Control is allocated to the speaker.

3. Question: Control is allocated to the speaker, unless it is a response to question or directive.

4. Acknowledgement: Control is allocated to the hearer.

   The following is a dialogue initiative example. The alphabet at end of each sentence in brackets denotes the participant with conversational focus.

   Dialogue initiative example

   (A) I think we should go probably to 10 [A]

   (B) No, because it increases a lot. It think 5 [B]

   (A) Okay, got it. [B]

   (B) Then 2. Is it 15? . [B]

   (A) Hmm, yeah. [B]

The author annotated for transactivity and dialogue initiative for the entire 20 experiments. To validate the annotations, another annotator had annotated four experiments, two face-to-face and two remote experiments. The inter annotator agreement is calculated with Cohen's kappa, the score is 0.89 for reasoning annotation, 0.85 for transactivity annotation and 0.76 for dialogue initiative annotation.

Chapter 5

FACE-TO-FACE AND REMOTE SPOKEN DIALOGUE ANALYSIS

In Chapter 4, speech data is annotated for reasoning, transactivity and initiative shifts. This chapter analyzed the collaborative learning in face-to-face and remote settings based on conversational features and annotation labels.

5.1 Conversational Features

The conversational features are high level spoken dialogue features calculated from the turn information. Table 5.1 shows the list of conversational features used for analysis. Average values are calculated for each of these features by summing up the values for all the sessions and dividing it by number of sessions. Balance measure represents the distribution of these features between the participants in the dialogue.

Features	**Functionals**
Overlap duration	
Turns	Average
Turn length	Balance
Words per turn	

Table 5.1: List of Conversational Features and Functionals Used to Analyze the Spoken Dialogue in Face-to-face and Remote Settings.

Table 5.2 shows the results for average measure of the conversational features. The average number of turns per experiment is higher in the remote setting than the face-to-face setting. In remote setting, it is observed that the number of acknowledgements per experiment is twice to that of face-to-face setting. Acknowledgement usually is a turn where no propositional content is expressed. The participants in the remote settings lacked gestures and dont have the freedom to look at what other person is doing. So they had to verbally communicate to attain grounding and confirmation. This led to increase in number of turns in the remote setting.

Results show that the amount of overlap in the face-to-face setting is 6.6% and in the remote setting it is 4%. The amount of overlap is calculated by dividing the overlap duration to the total duration of the experiment. From figure 5.1, the higher amount of overlap in the face-to-face setting supports the hypothesis that turns are handled more formally in remote spoken dialogue.

Average	**Face-to-face**	**Remote**
Turns per experiment	206.10(118.30)	350.40(181.20)
Overlap duration	0.06(0.03)	0.04(0.024)
Turn length	5.94(2.55)	3.78(1.04)
No of words per turn	9.24(2.36)	7.28(1.54)

Table 5.2: Comparison of Spoken Dialogue Based on Conversational Features.

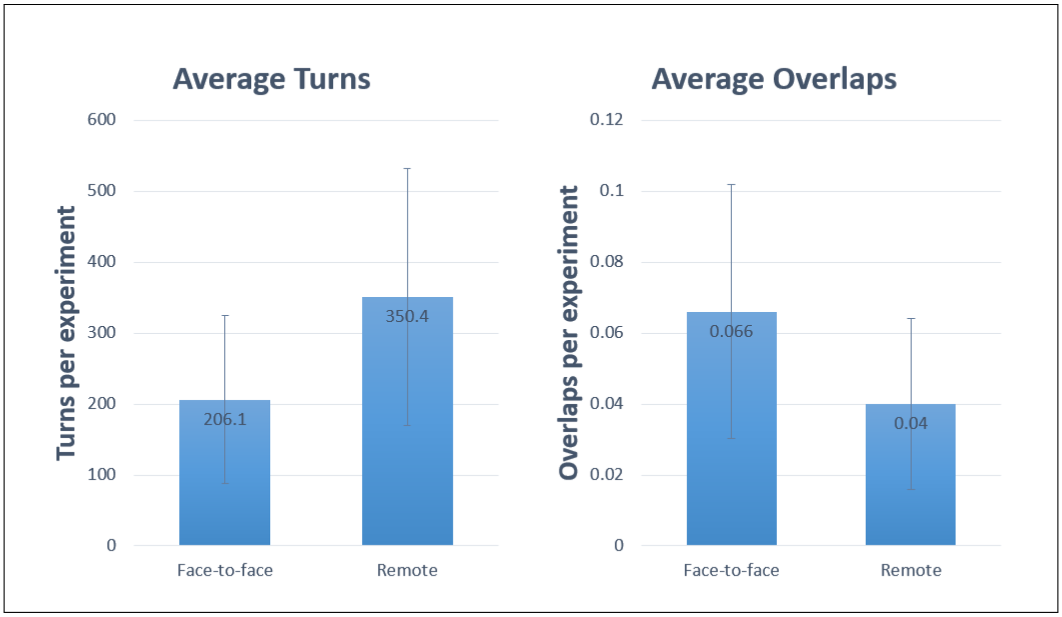


Figure 5.1: Comparison of Average Turns and Overlaps in Face-to-face and Remote Settings.

From figure 5.2, the average turn length per experiment is higher in face-to-face setting. It is observed that acknowledgements per turn is higher in remote setting and also the number of turns is higher in remote setting. As acknowledgements usually have shorter turn length. When averaged across the experiments they led to smaller average turn length in remote setting.

The average number of words per turn is higher in face-to-face setting. Turn by definition is a continuous speech utterance by a single speaker. The larger the turn length, larger are the number of words in the turn. As the average turn length is higher in face-to-face settings. The average number of words per turn shows similar pattern of average turn length.

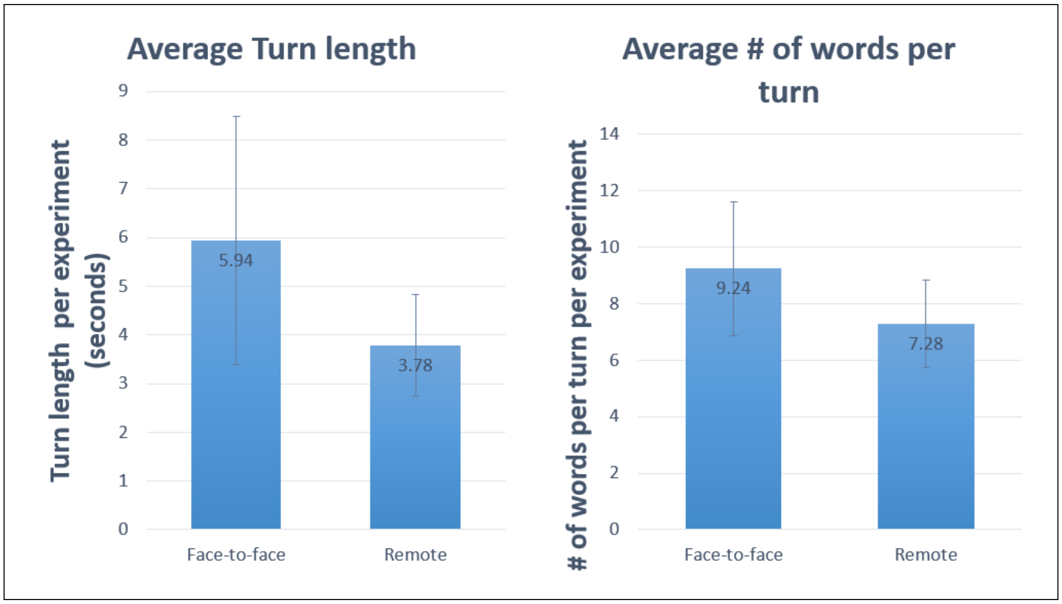


Figure 5.2: Comparison of Average Turn Length and Average Number of Words Per Turn in Face-to-face and Remote Settings.

Table 5.3 shows the results for balance measure of the conversational features. Balance measure captures the amount of contribution by participants in a dyad.

Balance	Face-to-face	Remote
Turns per experiment	0.06(0.07)	0.05(0.05)
Turn length	0.28(0.26)	0.23(0.13)
No of words per turn	0.26(0.24)	0.17(0.16)

Table 5.3: Comparison of Spoken Dialogue Based on Balance in Conversational Features.

Based on analysis from different measures, spoken dialogue in remote settings is more balanced than face-to-face settings. Figures 5.3 and 5.4 show graphs comparing balance in conversational features. It is noticed that in a few face-to-face experiments one participant amongst the pair takes the lead and solves the problem. The other participant helps him. The participant solving the problem doesnt speak out what he is doing as the other participant can see it directly. In the case of remote as the other participant can't see him, they have to verbally explain to each other. So in a face-to-face setting, few participants contributed more to spoken dialogue whereas the other participant is just solving the problem. This led to an imbalance in the contribution.

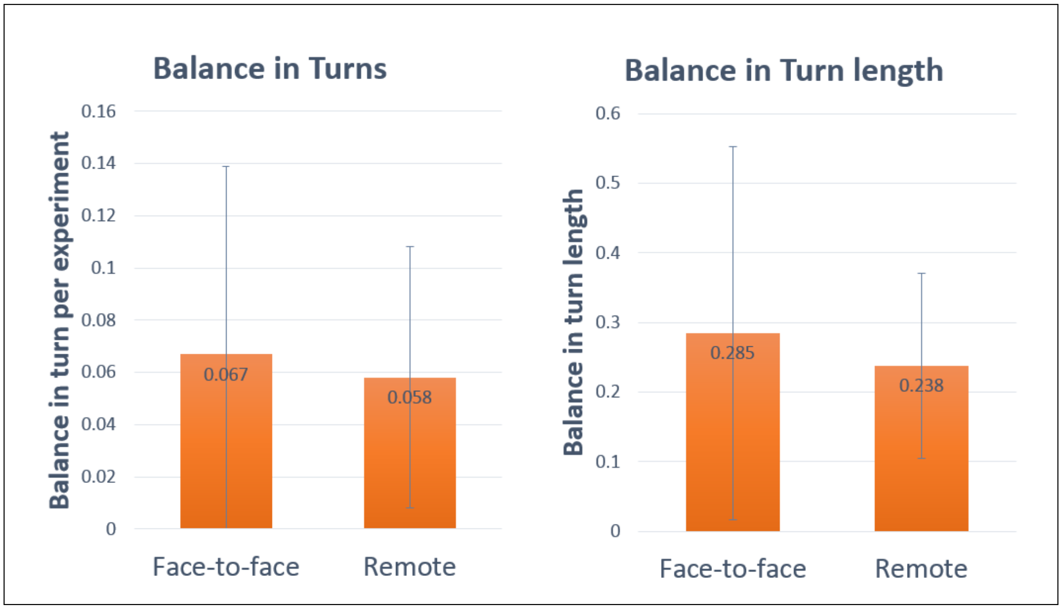


Figure 5.3: Comparison of Balance Measure of Number of Turns and Turn Length in Face-to-face and Remote Settings.

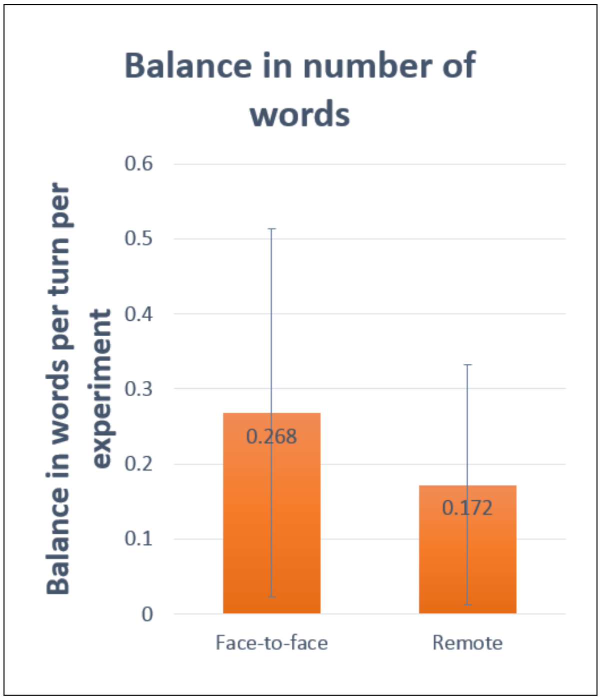


Figure 5.4: Comparison of Balance in Number of Words per Turn in Face-to-face and Remote Settings.

5.2 Transactivity

Both the settings are individually compared for transactivity based on number of transactive turns. The amount of transactivity is calculated by the total number of transactive turns per total number of turns in the experiment. The values from each experiment are averaged for face-to-face and remote settings. Table 5.4 shows the results for transactivity in face-to-face and remote settings. The percentage of transactive turns for the combined set of strangers and friends in face-to-face is 37% whereas its 32% in remote setting. A one-way ANOVA was used to test for differences in amount of transactivity among strangers and friends in face-to-face and remote settings. The amount of transactivity differed significantly across the four groups, F(3,16)=3.336, p=0.0459. Testing the difference between groups, the transactivity amongst strangers varied significantly in face-to-face and remote settings, two-tailed t(11.8)=2.43, p-value=0.03164. From figure 5.5, the amount of transactivity amongst strangers is high in face-to-face settings than remote settings.

Scenarios	Avg transactivity (STDEV)
Face-to-face	
strangers	0.40 (0.09)
friends	0.27(0.04)
combined	0.37 (0.10)
Remote	
strangers	0.30 (0.07)
friends	0.34 (0.04)
combined	0.31 (0.06)

Table 5.4: Average Amount of Transactivity for Face-to-face and Remote Settings.

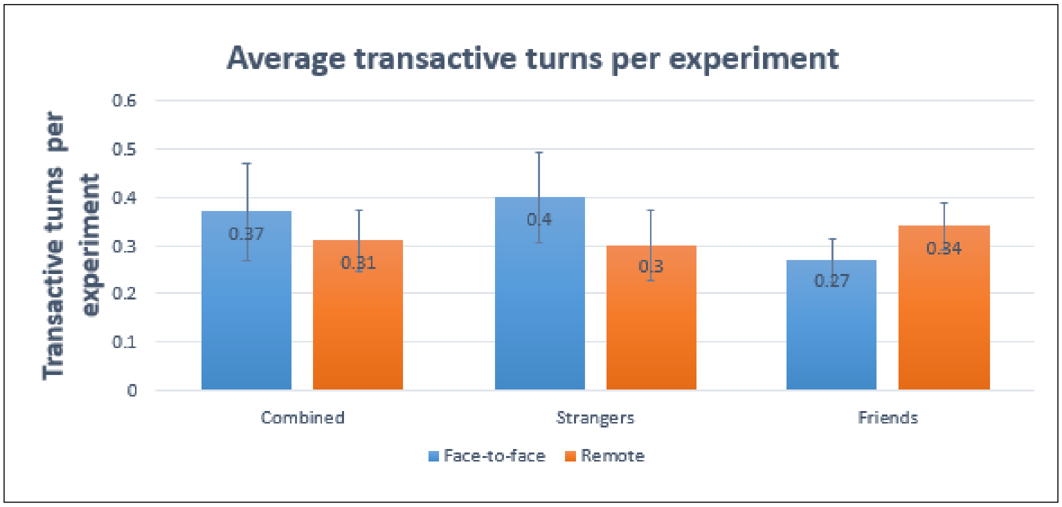


Figure 5.5: Average Transactivity in Face-to-face and Remote Settings.

5.3 Dialogue Initiative

In order to have a good collaboration, partcipants in the dialogue have to contribute equally to the problem solving task. In this thesis dialogue initiative is used as a measure of collaboration. Analysis is carried out in two steps. First step is to look at the balance in number of turns where a participant holds control of a dialogue. The second step is to analyze the number of control shifts in dialogue initiative from one participant to the other. Dyads who collaborated well and contributed equally will have low balance measures. Similarly for the control shifts the dyads with good collaboration will have high control shift scores. Table 5.5 show the results for both of these approaches. Figures 5.6 and 5.7 are graphs depicting the distribution for balance measures and control shifts.

Scenarios	Avg Balance (STDEV)	Avg Initiative Shift (STDEV)
Face-to-face		
strangers	0.26 (0.15)	0.48 (0.09)
friends	0.04 (0.001)	0.51 (0.10)
combined	0.22 (0.16)	0.49 (0.08)
Remote		
strangers	0.14 (0.05)	0.48 (0.07)
friends	0.09 (0.03)	0.48 (0.08)
combined	0.12 (0.04)	0.48 (0.07)

Table 5.5: Average Dialogue Initiative Balance Measures and Control Shift Initiative Measures per Turn for Face-to-face and Remote Settings.

A one-way ANOVA was used to test for differences in balance measures and initiative shifts among strangers and friends in face-to-face and remote settings. The balance measures and initiative shifts showed no significant difference. ANOVA for balance measures is F(3,16)=1.766, p=0.194. ANOVA for initiative shifts is F(3,16)=0.067, p=0.976.

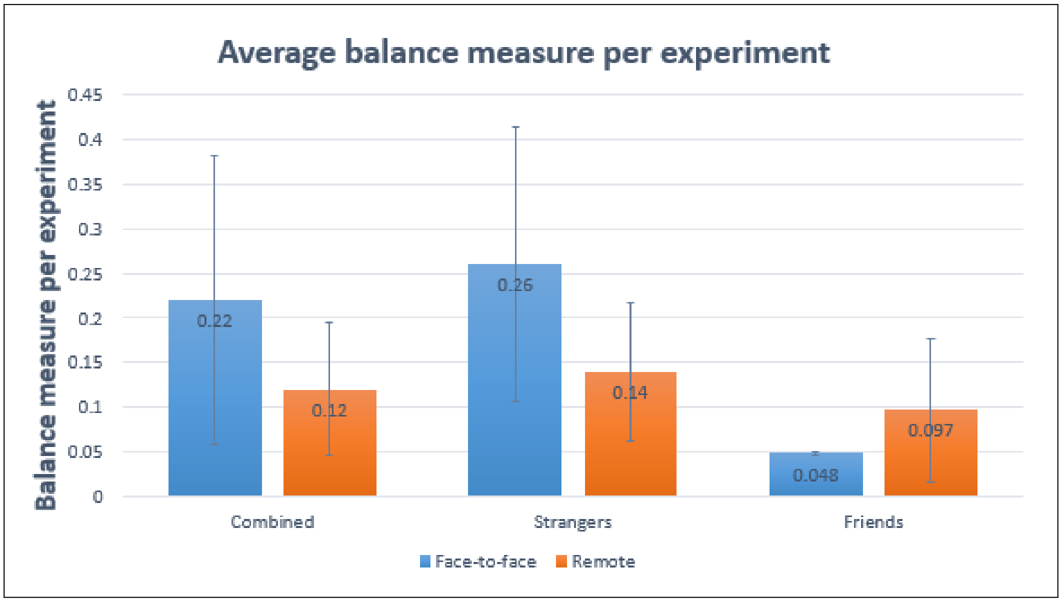


Figure 5.6: Comparison of The Balance Measure in Initiative for Face-to-face and Remote Settings.

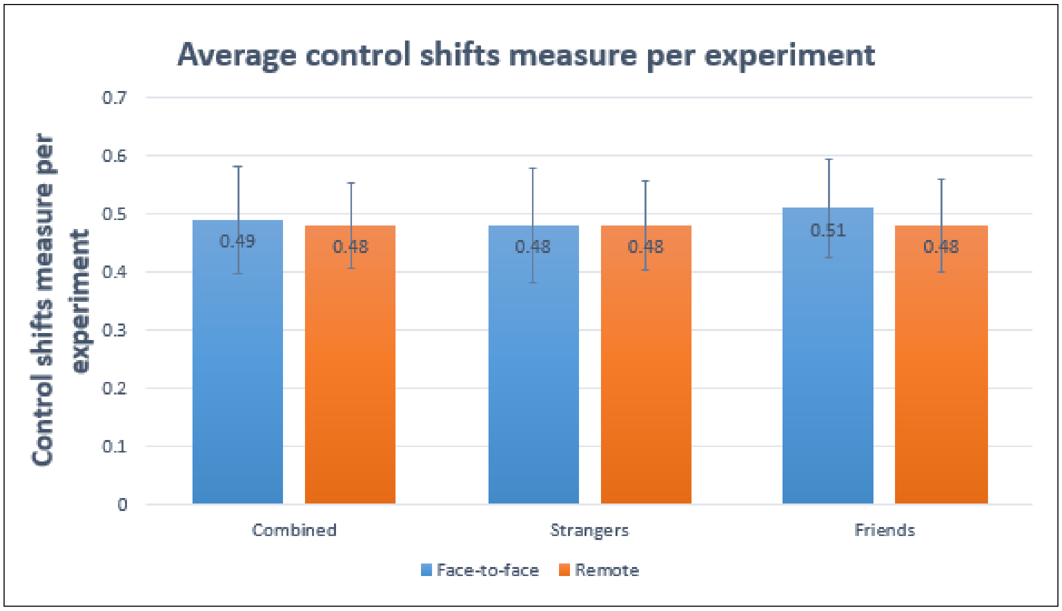


Figure 5.7: Comparison of Control Shifts in Initiative for Face-to-face and Remote Settings.

5.4 Transactivity and Dialogue Initiative

In my research transactivity is considered to be a measure of collaborative learning. And initiative shifts as a measure of collaboration. A higher number of control shifts indicate good collaboration. A positive correlation between transactivity and initiative shifts implies good collaborative learning. Pearson's correlation coefficients are calculated for face-to-face and remote settings. The face-to-face setting showed a positive correlation r = 0.152, t(8) = 0.886, p-value = 0.051. The remote setting showed a non-significant positive correlation r = 0.628, t(8) = 0.573, p-value = 0.051.

5.5 Summary

In comparing spoken dialogue based on conversational features, the average number of turns per experiment is higher is remote settings. There is a higher amount of overlap in the face-to-face setting compared to the remote supporting that the turns are more formally handed over in remote spoken dialogue. The strangers in the face-to-face setting are significantly more transactive compared to the ones in remote, which shows that strangers feel more comfortable in a face-to-face conversation than in a remote online conversation. The average turn length and average number of words per turn is also higher in face-to-face settings. Remote settings are more balanced compared to face-to-face settings.

Analysis based on distribution of annotation labels showed that amount of transactivity significantly varies in face-to-face and remote settings. Transactivity is higher amongst strangers in face-to-face settings compared to remote settings. Balance and control shifts in dialogue initiative are used as a measure of collaboration. Collaboration is compared in both the settings and the results showed no significant difference in collaboration for balance measures and amount of control shifts for face-to-face and remote settings.

Chapter 6

MACHINE LEARNING CLASSIFICATION EXPERIMENTS

In Chapter 5, face-to-face and remote settings are analyzed based on conversational features and annotation labels for transactivity and dialogue initiative. This chapter analyzed the collaborative learning in face-to-face and remote settings using the acoustic and prosodic features extracted from the speech data. Machine learning classifiers are built to analyze the spoken dialogue based on reasoning and transactivity. Section 6.1 explains the set of acoustic and prosodic features extracted, Section 6.2 explains various machine learning models used to build classifiers. Sections 6.3, 6.4 and 6.5 shows different machine learning models and their results. In section 6.6, the results are summarized.

6.1 Acoustic Feature Extraction

The unit of analysis for extracting features is a turn. Pitch, intensity, duration, voice quality, Mel-frequency cepstral coefficients and linear spectral coefficient are extracted from the speech data. The duration features model the temporal aspects of the spoken dialogue. The intensity features model the loudness energy of a sound as perceived by the human ear. Jitter and shimmer are associated with subtle voice qualities. The spectrum characterizes the spoken content. The cepstrum, the inverse spectral transform of the logarithm of the spectrum, represents changes in periodicity in the spectrum and it is relatively robust against noise. Using the openSMILE toolkit (Eyben et al., 2010), a set of 1562 low level acoustic/prosodic features are extracted using emobase configuration. Table 6.1 shows the features extracted.

6.2 Machine Learning Models

Machine learning classifiers are built to predict transactivity, initiative and face-to-face vs remote settings using the Weka toolkit (Hall et al., 2009). The Adaptive Boosting machine leaning algorithm (Freund and Schapire, 1995) with base classifier as JRip or SMO are used. The Adaptive Boosting algorithm was designed to be resilient to noisy data and outliers because of the way it trains a model over multiple iterations, and the instances that are misclassified in early iterations receive more attention in the subsequent rounds through re-weighting mechanism. JRip is a Weka implementation of RIPPER, a rule based learner. SMO is a Weka implementation of the Support Vector Machines (SVM) algorithm. As the dataset is smaller in size, cross validation is done to calculate accuracy. Sufficient care is taken to ensure the instances from a same person are not part of training and test sets during cross validation.

6.3 Machine Learning Models to Predict Face-to-face and Remote Settings

Face-to-face and remote collaborative learning settings are compared based on acoustic and prosodic features. Classification models are learnt for the entire data set and transactive turns to predict collaboration setting type (face-to-face or remote). Adaptive boosting algorithm with support vector machine base classifier is used. The results of the machine learning experiments are shown in Table 6.2. Both the classifiers are able to classify the instances with very high accuracy. Based on the Decision trees learnt for both the settings Mel-frequency cepstral coefficients and log Mel-frequency features are significantly different in face-to-face and remote settings. The model for entire dataset was able to predict the collaborative learning setting type with 93.19% accuracy. While the accuracy for transactive turns model is 94.25%.

Acoustic/Prosodic feature	**# Feature**	**Deltas**	**Functionals**	**Total**
MFCC	15	15	21	630
Mel Frequency Band	8	8	21	336
Linear Spectral Coefficient	8	8	21	336
loudness	1	1	21	42
F0-envelope	1	1	21	42
voicing	1	1	21	42
Pitch	1	1	19	38
jitter	1	1	19	38
jitter(ddp)	1	1	19	38
Shimmer	1	1	19	38
Turn duration in seconds				1
# pitch onsets (pseudo syllables)				1
			Sum	1582

Table 6.1: Acoustic and Prosodic Features Extracted for Each Turn.

Figure 6.1 shows the accuracy of face-to-face versus remote classifier.

Scenarios	#Turns	F	R	Baseline(%)	Acc(%)
All turns					
strangers	2546	1334	1212	52.39	93.19
friends	2341	580	1761	75.22	90.52
combined	4887	1914	2973	60.83	90.27
Transactive turns					
strangers	1089	628	461	57.6	94.25
friends	868	192	676	77.8	92.54
combined	1957	820	1137	58	91.47

Table 6.2: Comparing Face-to-face Versus Remote Scenarios Based on Topic Level and Surface Level Reasoning Statements That Contribute to Transactivity. F: Face-to-face, R: Remote.

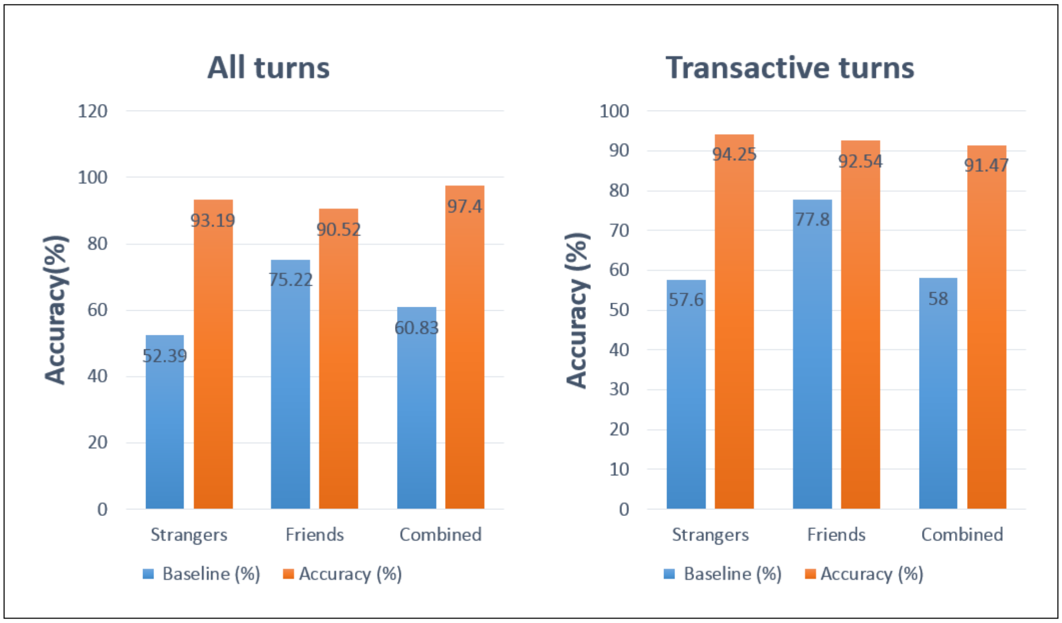


Figure 6.1: Classifier Prediction Accuracy of Face-to-face and Remote Classifier.

6.4 Machine Learning Models to Predict Reasoning

Adaptive boosting machine learning algorithm with support vector machines base classifier is used to build classification models to predict different reasoning types in face-to-face and remote settings. Table 6.3 shows the results of the classifier. Different possibilities with strangers only and with friends are also considered. Figure 6.2 shows the accuracy of multiclass reasoning classifier.

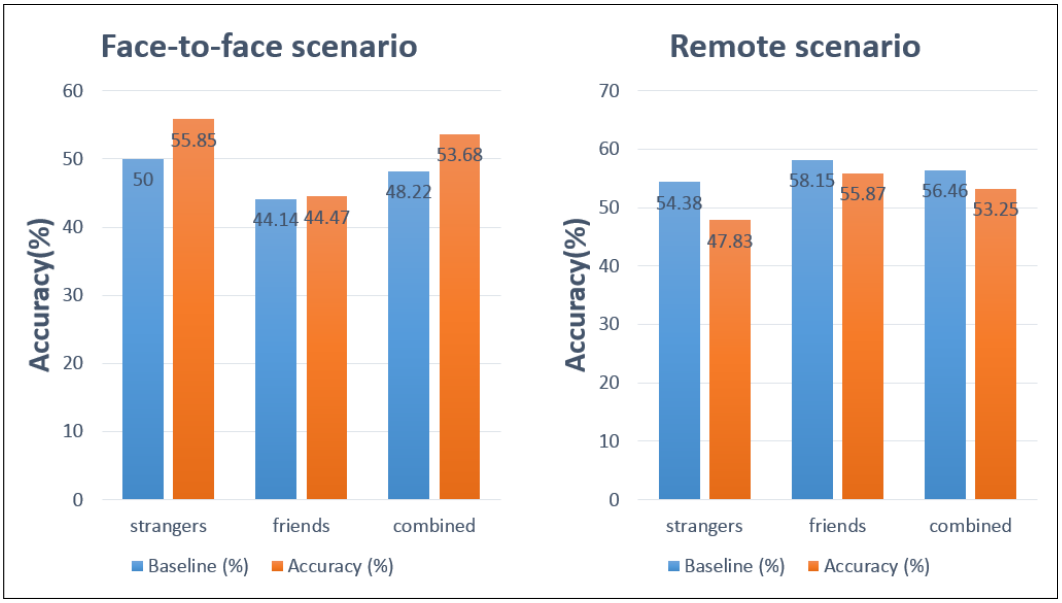


Figure 6.2: Classifier Prediction Accuracy of Multiclass Reasoning Classifier.

6.5 Machine Learning Models to Predict Transactivity

Adaptive boosting machine learning algorithm with RepTree base classifier is used to build classification models to predict transactivity in face-to-face and remote settings. Table 6.4 shows the results of the classifier for each setting individually and also for combined settings. Different possibilities with strangers only and with friends are also considered. From the decision tree classifiers built for same set of settings, the turn duration is the top most feature that predicts transactive and non-

Scenarios	#Turns	R1	R2	R3	NR	Baseline(%)	Acc(%)
Face-to-face							
strangers	1334	135	493	39	667	50	55.85
friends	580	20	172	132	256	44.14	44.47
combined	1914	155	665	171	923	48.22	53.68
Remote							
strangers	1438	96	440	120	782	54.38	47.83
friends	1761	52	624	61	1024	58.15	55.87
combined	3199	148	1064	181	1806	56.46	53.25

Table 6.3: Comparing Different Reasoning Types in Face-to-face and Remote Settings. R1: Conceptual Relevance, R2: Surface Relevance, R3: Activity Relevance.

transactive instances. Transactive instances have a higher turn duration compared to non-transactive turns. In the case of face-to-face most of the turns with duration greater than 1.59 seconds are marked transactive, for remote it is 1.69 seconds and for the combined the turns with duration greater than 1.76 seconds are marked transactive. Figure 6.3 shows the accuracy of binary transactivity classifier.

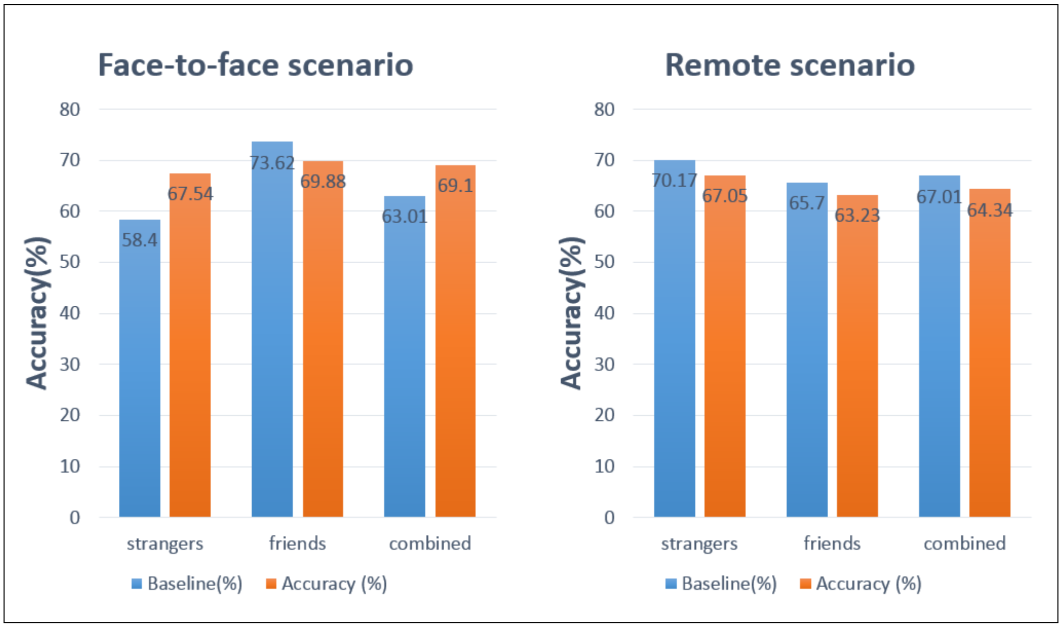


Figure 6.3: Classifier Prediction Accuracy of Binary Transactivity Classifier.

Scenarios	#Instances	T	NT	Baseline(%)	Acc(%)
Face-to-face					
strangers	1334	555	779	58.39	67.54
friends	580	153	427	73.6	69.88
combined	1914	708	1206	63	69.1
Remote					
strangers	1438	429	1009	70.17	67.05
friends	1761	604	1157	65.7	63.23
Combined	3131	1033	2098	67.01	64.34

Table 6.4: Comparing Transactivity in Face-to-face and Remote Scenarios. T: Transactive, Nt: Non-transactive.

Transactivity at a session level in both the settings is also analyzed. Instead of considering turn as a basic unit of analysis, features averaged at a turn level for the experiment is considered as a basic unit of analysis. Adaptive boosting machine learning algorithm with J48 base classifier is used to build classification models to predict transactivity in face-to-face and remote settings. A total of 10 instances each for face-to-face and remote settings is considered for analysis. The input features for the machine learning algorithm includes acoustic and prosodic features averaged for all the turns in an experiment session and conversational features like number of turns in a session, average turn length, number of words per turn and overlap duration for the session. Table 6.5 and Figure 6.4 shows the accuracy of machine learning classifier.

Scenarios	**#Instances**	**H**	**L**	**Baseline(%)**	**Acc(%)**
Face-to-face	10	6	4	60	80
Remote	10	6	4	60	90

Table 6.5: Comparing Transactivity at a Session Level in Face-to-face and Remote Scenarios. H: High Transactivity, L: Low Transactivity.

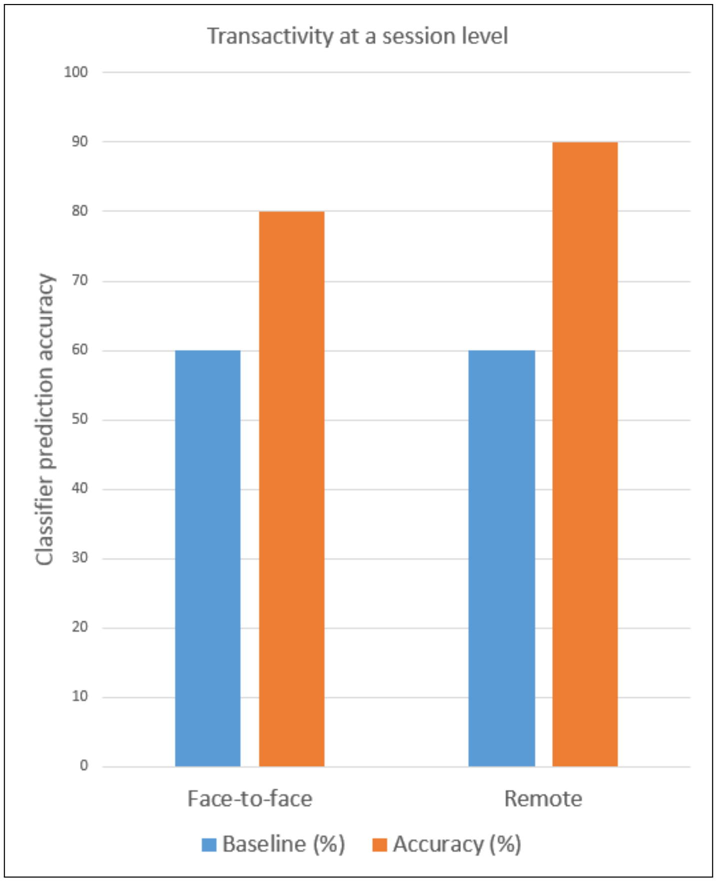


Figure 6.4: Classifier Prediction Accuracy of Binary Transactivity Classifier at a Session Level.

6.6 Summary

In comparing spoken dialogue based on the acoustic and prosodic features of spoken dialogue, the face-to-face and remote settings are different. The classifiers are built to detect both the settings for the complete data set and for the transactive-turns-only dataset. For both these datasets classifiers resulted in very high accuracy. The results also show that the Mel-frequency cepstral coefficients are different in both cases. The classifiers built to predict reasoning and transactivity didn't do well compared to the baseline classifier.

Chapter 7

CONCLUSION AND FUTURE WORK

The research undertaken explored how spoken dialogue varied in face-to-face and remote collaborative learning environments. My thesis contributed towards collecting a speech corpus and performing empirical analysis of the spoken dialogue using machine learning techniques. High quality speech data was collected from participants solving math problems on a tablet. Speech was transcribed and manually annotated for different types of reasoning, transactivity and dialogue initiative as explained in Chapter 4. A new coding scheme was presented to analyze collaborative learning.

Spoken dialogue was analyzed based on conversation, collaboration, and speech prosody features. In comparing spoken dialogue based on conversational features, the average number of turns per experiment is higher is remote settings. There is a higher amount of overlap in the face-to-face setting compared to the remote supporting the findings from the previous research (Sellen, 1995) that the turns are more formally handed over in remote spoken dialogue.

The transactivity in dialogues between strangers was higher in face-to-face settings than remote settings, which may be attributed to strangers feeling more comfortable to collaborate and learn in a face-to-face setting. A one-way ANOVA was used to test for differences in amount of transactivity among strangers and friends in face-to-face and remote settings. The amount of transactivity differed significantly across the four groups, $F(3,16)=3.336$, $p<0.05$. Testing the difference between groups using two-tailed t-tests, the transactivity amongst strangers varied significantly in face-to-face and remote settings, two-tailed $t(11.8)=2.43$, $p<0.05$. The t-tests for the balance and shifts in the initiative as a measure of collaboration showed no significant difference in between face-to-face and remote settings. Both the balance and control shifts in initiative were normalized using the number of turns for that session.

The acoustic and prosodic features of speech, MFCCs and LSPs differed in face-to-face and remote settings and helped in classifying the turns as a face-to-face setting turn or remote setting turn with up to 93% accuracy. The accuracies of classifiers built to predict reasoning and transactivity were better than baseline values in face-to-face settings, whereas the prediction accuracies were less than baseline values in remote settings. The class distribution in reasoning and transactivity is skewed in remote settings and that may be the reason the prediction accuracies were lower than baseline values. From the decision trees built to classify turns into transactive and non-transactive the turn duration and MFCC's were the most significant features that predicted transactive turns.

Learning the differences between face-to-face and remote collaboration will help researchers to build efficient collaborative learning tools. My research showed that there was more co-construction in the face-to-face setting than the remote setting. This brings out a new problem to the research community on why co-construction varies in face-to-face and remote settings. The machine learning classifiers built to predict transactive and non-transactive turns showed that turn duration and MFCCs are features which can differentiate them. So researchers can make use of these features to build tools that automatically predict transactive turns from spoken dialogue.

In the current work most of the experiments comparing face-to-face and remote settings didn't show a significant difference or correlation due to a smaller sample size. This can be improved by considering more test subjects and conducting the research on a larger scale. Most of the test subjects were undergraduate students in computer science with strong foundation in mathematics. They found the elementary math problems easy to solve. Either choosing difficult problems or targeting high school students for whom these problems can be challenging, might yield better results. The present domain of mathematical problems led to high surface level reasoning and low conceptual level reasoning. Choosing reasoning problems from different domains will push participants to collaborate and co-construct enriching solutions. This can help researchers analyze the transactivity better.

Ray Birdwhistell an American anthropologist found that the verbal component of a face-to-face conversation is less than 35 percent and that over 65 percent of communication is done nonverbally (Birdwhistell, 1974). Given that the remote setting lacks a few non-verbal gestures compared to face-to-face setting, it will be interesting to see how facial and hand gestures impact collaborative learning in both the settings.

REFERENCES

Josianne Basque and Béatrice Pudelko. The effect of collaborative knowledge modeling at a distance on performance and on learning. In *Proceedings of the First International Conference on Concept Mapping (CMC 2004), September 14-17, vol. 1*, pages 67–74, 2004.

Ray L Birdwhistell. The language of the body: the natural environment of words. *Human Communication: Theoretical explorations*, pages 27–52, 1974.

Michelene TH Chi, Marguerite Roy, and Robert GM Hausmann. Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2):301–341, 2008.

Jennifer Chu-Carroll and Michael K Brown. Tracking initiative in collaborative dialogue interactions. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 262–270. Association for Computational Linguistics, 1997.

Pierre Dillenbourg. What do you mean by collaborative learning? *Collaborative-learning: Cognitive and computational approaches.*, pages 1–19, 1999.

Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the International Conference on Multimedia*, pages 1459–1462. ACM, 2010.

Yoav Freund and Robert E Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory*, pages 23–37. Springer, 1995.

Gahgene Gweon. *Assessment and support of the idea co-construction process that influences collaboration.* PhD thesis, Carnegie Mellon University, 2012.

Gahgene Gweon, Rohit Kumar, Soojin Jun, and Carolyn Penstein Rosé. Towards automatic assessment for project based learning groups. In *AIED*, pages 349–356, 2009.

Gahgene Gweon, Mahaveer Jain, John McDonough, Bhiksha Raj, and Carolyn P Rosé. Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. *International Journal of Computer-Supported Collaborative Learning*, 8(2):245–265, 2013.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

RG Hausmann, Michelene TH Chi, and Marguerite Roy. Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. In *26th annual Conference of the Cognitive Science Society*, pages 547–552, 2004.

Mahaveer Jain, John McDonough, Gahgene Gweon, Bhiksha Raj, and Carolyn Penstein Rosé. An unsupervised dynamic bayesian network approach to measuring speech style accommodation. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 787–797. Association for Computational Linguistics, 2012.

David W Johnson and Roger T Johnson. Learning together and alone: Overview and meta-analysis. *Asia Pacific Journal of Education*, 22(1):95–105, 2002.

Cynthia Kersey, Barbara Di Eugenio, Pamela Jordan, and Sandra Katz. KSC-PaL: A peer learning agent that encourages students to take the initiative. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 55–63. Association for Computational Linguistics, 2009.

Roberto Martinez-Maldonado, Judy Kay, and Kalina Yacef. An automatic approach for mining patterns of collaboration around an interactive tabletop. In *Artificial Intelligence in Education*, pages 101–110. Springer, 2013.

Bruce M McLaren, Oliver Scheuer, Maarten De Laat, Rakheli Hever, Reuma De Groot, and Carolyn Penstein Rosé. Using machine learning techniques to analyze and support mediation of student e-discussions. *Frontiers in Artificial Intelligence and Applications*, 158:331, 2007.

Diana Oblinger, James L Oblinger, and Joan K Lippincott. *Educating the net generation*. Brockport Bookshelf, 2005.

Brid O'Conaill, Steve Whittaker, and Sylvia Wilbur. Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-computer interaction*, 8(4):389–428, 1993.

Paul Resta and Thérèse Laferrière. Technology in support of collaborative learning. *Educational Psychology Review*, 19(1):65–83, 2007.

Abigail J Sellen. Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, 10(4):401–444, 1995.

Amy Soller and Alan Lesgold. A computational approach to analyzing online knowledge sharing interaction. In *Proceedings of Artificial Intelligence in Education*, pages 253–260, 2003.

David R Traum and Peter A Heeman. Utterance units in spoken dialogue. In *Dialogue Processing in Spoken Language Systems*, pages 125–140. Springer, 1997.

Jeremy I Tutty and James D Klein. Computer-mediated instruction: A comparison of online and face-to-face collaboration. *Educational Technology Research and Development*, 56(2):101–124, 2008.

Marilyn Walker and Steve Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics*, pages 70–78. Association for Computational Linguistics, 1990.

APPENDIX A

CODING MANUAL

A.1 CODING SCHEME

In order to analyze the collaborative learning dialogue, it is required to have instruments that measure collaboration and learning separately. For the FACT corpus the balance and control shifts in the dialogue initiative is used as a measure of collaboration. The knowledge co-construction a form of transactivity in the field of learning sciences is used as a measure of learning. The following subsections explain the coding scheme to annotate each turn based on its content.

A.1.1 RELEVANCE

Relevance is marked at a turn level and it is broadly classified into three types, conceptual relevance, surface relevance and activity relevance. A turn can be marked R1,R2 or R3 based on the action type by following conditions given below.

1. **Conceptual Relevance**: Contains mathematical terms like Speed, distance, graph, angle. Mark conceptual relevance as [R1].
   - but doesn't the skateboarding one fit it better than it fits this one because he **slows down** and this **graph is speeding up** in the middle
   - then the walking part is not **slowly on the top** so that the **part angle should be steeper** on the middle like it is
2. **Surface relevance**: If its related to the problem that is being worked on based on the graphs or text of the problem. Mark surface relevance as [R2].
   - yeah ok opposite toms home is a hill tom climbs slowly up the hill walked across the top then ran quickly down the other side so he is far away from his house the whole time
   - Now that, it would be **eight boomerangs** right, **eight large** ones
3. The activity or application specific tasks like moving the cards, matching the values should be marked [R3].
   - I feel like it would be best if like we would like make sure the **table has the actual graph**

   But not when discussing about the functionality of the application, like drawing, erasing and tools.
   - Oops **highlighters** too
   - im just **unplugged** so it disappeared.
   - i cant even like pinch or zoom in and out.

All those turns which doesn't follow the above conditions should be marked not relelvant [NR].

A.1.2 TRANSACTIVE CONTRIBUTIONS

A statement to be marked transactive contribution [T] it should have a clear reference to relevant content in the previous contribution/turn. A clear reference can be either a **repeated content word(s)** or a **pronoun**. To mark a turn as a transactive contribution, the turn being considered and its previous turn should be a relevant statement marked [R].

Examples:

1. **Repeated content word(s)** in the current turn to previous turn. A content word conveys information in a text or speech act. It is also known as lexical word. Content words include nouns, lexical verbs, adjectives and adverbs.
   - (A) but doesnt the **skateboarding** one fit it better than it fits this one because he slows down and this graph is speeding up in the middle
(B) **skateboarding** from his house gradually building up speed slows down and he speeds up. Hmm
   - (A) Now that ,it would be **eight boomerangs** right, eight large ones
(B) I think yeah what they are asking for ah they just say how many how many large small and large **boomerangs** should they make. So I am I am thinking that its asking how many they supposed to make most money for charity
   - (A) ok so from hundred and fifty to two hundred its **going up** by (B) **going up** by fifty dollars which is thirty three and one third (A) ya when it **goes down** its down by (B) twenty five (A) ok
2. **Pronoun** referring to relevant noun phrase in the previous turn.
   - (A) Tom ran from his home and then stopped
(B) so he is running he stops and then he walks backhome
(A) so that would be G
   - (A) I feel like there is nothing that goes with H
(B) We probably have to make that one up
(A) ok, would you do R with F (B) Ahhmmm, I wouldnt because around the middle they are not really moving, and you see around three and four there is increasing in speed.

A.1.3 INITIATIVE

Initiative annotation takes place in multiple steps. First step is to classify the utterances into one of the four categories given below. Then based on the categories and control rules mark who holds the control.

Utterance Type

1. **Assertion**: Declarative utterances used to state facts. Mark an assertion as [S], The given below are few examples for assertion. Utterances with propositional content

   (a) Tom ran from his home and then stopped.
   (b) You probably moved on my N like C
   (c) Slowed down to avoid some rough ground, but sped up again.

2. **Question**: Utterances which are intended to elicit information. Questions can be direct, the utterances that begin with keywords like *What*, *How*, *Where* etc.. Mark the question as [Q], the given below are few examples for question.

   (a) Do you wanna start with stories ?
   (b) wait! How did you do the distance for that oh it would be like the same time but a different distance
   (c) Which one did you move ?
   (d) What do you mean by graph steadily rising ?

3. **Question-indirect**: Utterances which are intended to elicit information indirectly. Indirect questions like *I was wondering whether I should....* It is likely that utterance can be either a question or a statement. While annotating pay attention to the intonation. Usually questions end with rising intonation. Mark the question as [QI], the given below are few examples for Indirect question.

   (a) I doubt if we should consider the straight line as constant speed with skateboarding down the hill ?
   (b) Hey no shouldnt that be two thirds ?

4. **Directives**: Utterances intended to instigate action. Generally imperative form. Mark the utterance as [D].

   (a) Just click on the pin button to unpin the card.
   (b) Can you do it I dont think I know how to do this

5. **Directives-indirect**: Utterances intended to instigate action, but not directly. For example *My suggestion would be that you do ....* Here the speaker is not asking or commanding the user to directly do something. Mark such directives as [DI]

   (a) I think the graph fits right into M, but am unable to move it. (Here M is a table that matches the graph. Expects the other participant to take an action and move it.)

6. **Acknowledgement**: Any utterance that do not express propositional content. Mark the utterance [A] if its an acknowledgement.

   (a) yeah
   (b) Okay
   (c) uhmmm, hmmm
   (d) ya right

Examples

In the below example, the utterance type and control are marked in brackets.

1. Example 1
   (A) *Do you wanna start with the stories* [Q] [A control]
   (B) *Yeah* [S][A control] (Here "yeah" is considered assertion)
   (A) *ok* [A][B - Control] (A acknowledges and control shifts to B)
   (B) *tom ran from his home and then stopped, so he is running he stops and then he walks back home so that would be G* [S][B-control]
   (A) *yeah,how do i move it* [Q][A-control]
   (B) *just click it*[S][B-control]

2. Example 2
   (A) so what i think would be better first would be to match the time graphs with the charts [Q] [A-Control]
   (B) yeah me too [A] [A-Control]
   (A) so looks like P matches with like E down there [S] [A-control]
   (B) so hard to see all of them one time. ahhm, i would say T is better ah better match for E [S] [B-control]
   (A) yeah thats true [A] [B-control]
   (B) i think one that would look just like a reflection of E would be better for P, may be G [B-control]
   (A) yeah I think that will work [A] [B-control]
   (B) Hmmm Yeah [A] [A-control]