International Society of the Learning Sciences

Audio Analysis of Teacher Interactions with Small Groups in Classrooms

Chris Palaguachi, University of Illinois at Urbana-Champaign, cwp5@illinois.edu
Eugene M. Cox, University of Illinois at Urbana-Champaign, emcox3@illinois.edu
Cynthia M. D'Angelo, University of Illinois at Urbana-Champaign, cdangelo@illinois.edu

Abstract: This paper presents exploratory work in combining computational methods with qualitative approaches in order to better understand teacher interactions with small groups of students. Classroom audio data is typically difficult to work with, due to background noise and challenging acoustics, and needs customization of algorithm parameters when using audio processing tools for speech detection. This secondary data analysis study looked at patterns in small group discussions of students over multiple class sessions and multiple teachers, especially focusing on times when teachers interacted with the groups. This type of approach can augment and extend the capabilities of qualitative researchers, who could use these computationally-derived analytics and patterns to aid them in better understanding teacher/student interactions and collaborative learning.

Introduction and Theoretical Background

Conducting video-content analysis on classroom learning can be a long and challenging task for qualitative researchers. However, the information gleaned from video-content analysis can help provide meaningful data on classroom discussion and teacher-group interactions. Computational methods can scale up analyses and process data much quicker than humans can, but do not always attend to the important nuances and context that qualitative work can focus on. However, there may be ways to integrate computational methodologies along with the commitments and questions that qualitative work can value. One such methodological framework proposed is computational grounded theory (Nelson, 2016). Computational grounded theory focuses on three steps: 1) pattern detection using computational exploratory analysis, 2) hypothesis refinement using human-conducted interpretive analysis, and 3) pattern confirmation (Nelson, 2016). This short paper will present results from our first steps of using a computational grounded theory framework on audio and qualitative data to better understand classroom discussion patterns and teacher-group interactions that occur in secondary math classrooms.

Investigating how classroom audio patterns relate to groupwork and teacher-group interactions are important because managing and accessing group and collaborative learning tasks can be difficult for teachers to monitor on their own (D'Angelo et al., 2019). By creating tools for teachers through computational audio-data analysis, researchers can help teachers better understand discourse in groups. More importantly, teachers need tools to help them identify groups of collaborating students that are doing well and those that need help (and what kind of targeted help students might need). Ideally, teachers would listen to peer interactions in each group for long enough to understand how discourse is proceeding. However, teachers cannot listen to more than one group at a time. Therefore, this in-depth listening, evaluation, and feedback teachers provide to students cannot be done at scale, even within a single classroom with more than a few groups (Kaendler et al., 2015). Students talking to one another in face-to-face environments is the natural setting for classroom collaboration and group work, and so speech-based analytics and methods are needed to address this problem.

This study, part of a larger methodology-focused project, aims to investigate ways in which the audio data present in classrooms could provide an additional view on classroom discussions and teacher interactions with small groups. The audio analysis can be done on a larger scale than the deep-dive that qualitative researchers typically focus on for smaller segments of interactions. Combining the qualitatively focused research questions and the power of computational analyses, this study reports on the beginning of a larger aim to produce grounded and human-interpretable outputs of computational analyses in order to better understand collaboration, teacher interactions, and small group discussion patterns in classroom spaces.

Methods

Data Set

The study presented here is part of a larger project doing a secondary data analysis of classroom audiovisual data collected to study mathematics teaching and learning in the United States (Dyer, 2016). In this study we focused on small group audio data and classroom activity codes (e.g., whole classroom discussion, small group work, individual work, and teacher interactions) derived from watching videos. The full audio data set includes 106 recorded high school classes from 10 instructors. Students sat at their group's table during each class session. Each class session was approximately 90 minutes long. Audio was collected from Zoom H1 microphones that were stationed at the center of each group's table. As a part of this study's secondary analysis, we selected three class sessions worth of data to process and analyze. In each of these classes, there was approximately 20 students, split into five groups in each class. For the exploratory study presented here, we selected three class sessions that had decent audio quality, lots of small group work, and a minimum number of class interruptions.

Student and Teacher Interaction Codes

Two qualitative coding schemes were used to supplement the audio data analysis: activity formats and group interaction codes. The activity formats noted whether students were working in groups (or other instructional formats such as teacher presentation or individual work occurring). Group interaction codes (Hudson, Parr, & Dyer 2021) were used to classify instances of teacher/group interactions where the teachers offered supportive statements, questions, or directions to the group to better sustain group activity.

Voice Activity Detection

To process the audio data, we used openSMILE – an open-source audio processing program (Eyben, Wollmner & Schuller, 2010). For this study, openSMILE's Voice Activity Detection (VAD) algorithms were used to detect where in the audio files human speech was occurring. The VAD algorithms include a variety of functions, including those to segment the frames of human-detected speech into turns of speech.

As expected, we found that the default parameters for the VAD and turn detector did not pick up all the human speech utterances in our audio data. Since the VAD default parameters are used to detect speech that is not often representative of a typical classroom environment, we went through an in iterative process and made two changes to the default VAD parameters: (1) to account for the dynamic variation of speech volume, and (2) to detect shorter utterances. The first change was adjusting the Root-Mean Square (RMS) values to an auto-threshold setting that allowed the VAD to automatically adjust the RMS values for dynamic shifts in speech per each turn detected. The second change was adjusting the turn detector parameters. We decreased the nPre value in the turn detector so that the number of frames of speech needed to detect the beginning of a new turn was shortened. We also decreased the nPost value in the turn detector so that the number of frames of silence needed to detect the end of a turn was shortened. These changes resulted in VAD's ability to detect shorter turns and helped segment turns that were not being picked up using default configuration settings. This however increased the probability of false positives in the form of non-speech utterances such as classroom noise. To help control for this and in line with existing research (Donnelly et al., 2017), speech utterances shorter than 500 milliseconds were removed from the data as they were unlikely to contain meaningful speech information.

Validation Process

To evaluate that the VAD tool was working correctly, we conducted a validation check on a randomized 10% subset of speech utterances to see if the turns included human speech and if the turns were segmented correctly. Through this validation process, we created categories to note instances of human speech, non-human speech, background noise, background speech, clean audio turns, and cut-off audio (turns that cut speech towards the end of a turn). In finalizing VAD's thresholds and parameters, audio for each of the five groups in each class were processed through openSMILE and then labeled and merged into datasets using R. Groups were labeled by color (blue, green, orange, purple, and red) corresponding to the group table naming convention and a script in R was then created to process and label the human-coded content logs (student interaction, group interaction, and activity formats) into a merged data set.

As a result of validating our parameters during audio processing, results showed our VAD configuration was able to detect speech utterances from different classroom sessions (see Table 1). For this data, we found that the VAD did the correct job of detecting human speech in clean turns (48% of the turns), background speech (34%), and cut-off audio (13%).

Table 1
Frequency of validation categories (and percentages within each session) for VAD performance

Session	Background Noise	Background Speech	Clean Turns	Cut-off Audio
Teacher 102 Session 1	9 (5%)	82 (46%)	59 (33%)	25 (14%)
Teacher 102 Session 2	7 (4%)	35 (20%)	92 (54%)	34 (20%)
Teacher 103 Session 1	12 (6%)	61 (33%)	100 (55%)	8 (4%)

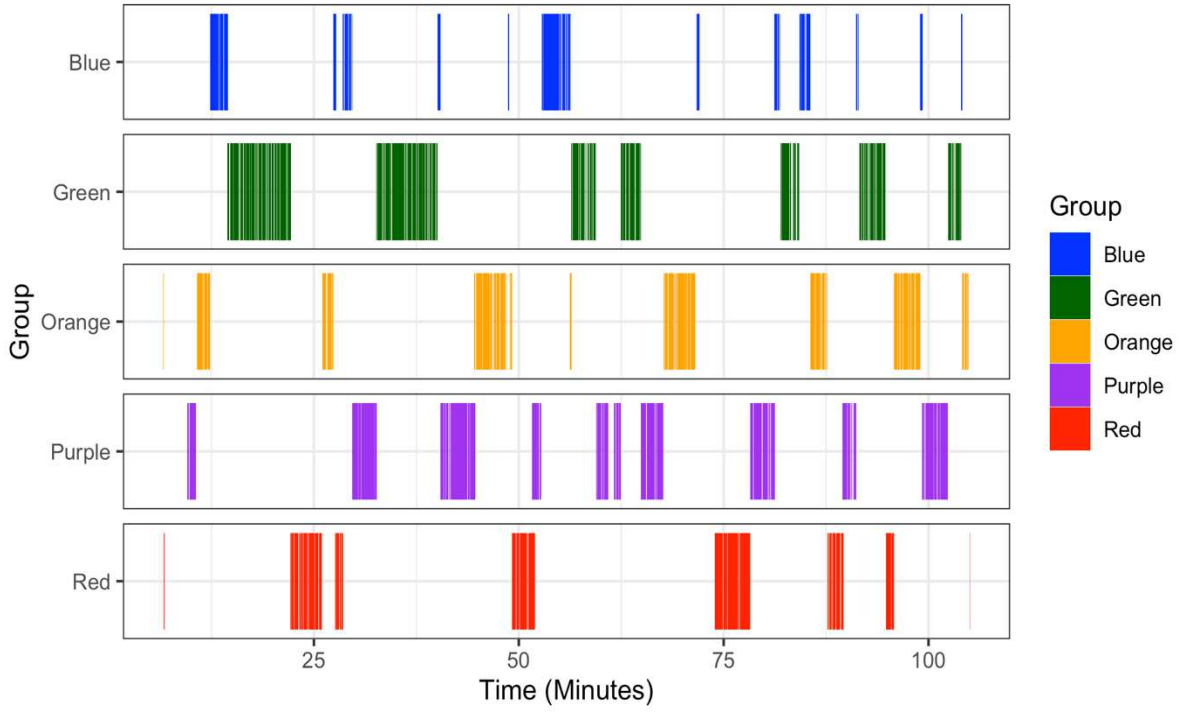
Although the VAD and turn detector functions accurately picked up speech and segmented turns, there were instances of background speech where the equipment picked up speech from other groups. For this study, we focused on comparing variation of speech activity from each session with the same classroom of students and teacher, and the variation of speech activity from different classes of students and teachers.

Findings

When comparing the average number of turns for each group in each classroom session, we found that Teacher 103 Class Session 1 (2,196 turns) contained more average turns of speech per group than either Teacher 102 session 1 (1,793 turns) or Teacher 102 Class Session 2 (1,728 turns). Furthermore, when comparing the average total duration of turns for each group in each class session, Teacher 102 Class Session 1 had longer average total duration (84.1 minutes) than Teacher 103 Class Session 1 (74.6 minutes) or Teacher 102 Class Session 2 (57.2 minutes). This information helps to illustrate that classroom speech can vary in both the number of turns and total duration, and that differences in student activity and teacher-group interactions can lead to variation in turns and durations for class sessions.

We also wanted to focus in on times when the teacher was interacting directly with groups during their group work. We selected one of the teacher sessions (Teacher 102 Class Session 2) for further inquiry because it had the highest amount of student activity consisting of groupwork and group interactions throughout the entire class session. Figure 1 visualizes the discussion patterns when the teacher was interacting with each small group throughout the whole class session.

Figure 1
Group speech activity during teacher-group interactions (Teacher 102 Class Session 2). Note: Each slice is a turn, and the width of each slice is related to the duration of the turn.



As seen in Figure 1, we explored discussion patterns when the teacher was interacting with students working in small groups. In this visualization we can observe the number of times the teacher moved from one group to another and the length of time the teacher/students spent interacting and discussing with each other. For instance, the blue group had very short interactions with the teacher, with very few turns per interaction. Whereas the green group had longer sustained interactions with the teachers with many turns during each interaction. In Figure 1, we can see that the green group had a large number of short turns, especially compared to the purple group which had longer turns when the teacher/group interaction occurred. Visualizations like these and, in general, analytics using audio data, can help researchers better understand small group and teacher discussion patterns at a larger scale than previously possible.

Conclusion and Implications

In this paper we explored the integration of segmented speech data with student-teacher interaction indicators to better map speech activity during teacher interventions with small groups. By combining computational audio features in the future, researchers can gain insights from small group discussions and visualize how teachers orchestrate and interact during a collaborative learning environment. This builds on extant literature on combining computational methods as a means of extracting relevant features and discerning patterns, with the intention of combining this with qualitative coding. Audio feature extraction relies on automatic turn segmentation and as a result, often has imperfect speech segmentation when using dynamic voice thresholds & fixed frame limits on speech detection. However, when combined with human coded time segments of group activity we can make assumptions about the collaborative quality among a group of students when a teacher intervenes during student groupwork by overlaying with speech data. Future work will explore new ways of modeling audio data with observed student and teacher interactions.

References

D'Angelo, C. M., Smith, J., Alozie, N., Tsiartas, A., Richey, C., & Bratt, H. (2019). Mapping individual to group level collaboration indicators using speech data. In Lund, K., Niccolai, G., Lavoué, E., Hmelo-Silver, C., Gweon, G., & Baker, M. (Eds). *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings: 13th International Conference on Computer Supported Collaborative Learning.* Lyon, France: International Society of the Learning Sciences

Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D'Mello, S. K. (2017). Words matter: Automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 218–227. https://doi.org/10.1145/3027385.3027417

Dyer, E. B. (2016). Learning through Teaching: An Exploration of Teachers' Use of Everyday Classroom Experiences as Feedback to Develop Responsive Teaching in Mathematics [Dissertation, Northwestern University]. http://gradworks.umi.com/10/16/10160667.html

Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor. *Proceedings of the International Conference on Multimedia - MM '10*, 1459. https://doi.org/10.1145/1873951.1874246

Hudson, H., Parr, E. D., Dyer, E. B. (2021). *Factors influencing teachers' use of groupwork in secondary math classrooms*. Poster presentation at TN STEM Education Research Conference.

Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher competencies for the implementation of collaborative learning in the classroom: A framework and research review. *Educational Psychology Review*, *27*(3), 505-536.

McIntoish, T., Wagner, A., Hudson, H., & Parr, E. D. (2021, January). *Math Instruction Across Secondary Classrooms: An Analysis of Teachers' Use of Activity Format*. Poster presentation at TN STEM Education Research Conference. Virtual.

Nelson, L. K. (2017). Computational Grounded Theory: A Methodological Framework. Sociological Methods & Research, 0049124117729703. https://doi.org/10.1177/0049124117729703

Acknowledgments

This material is based upon work supported by the National Science Foundation (DRL-1920796). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.