Multi-Sensor Modeling of Teacher Instructional Segments in Live Classrooms

Patrick Donnelly[1], Nathan Blanchard[1], Borhan Samei[2], Andrew M. Olney[2], Xiaoyi Sun[3], Brooke Ward[3], Sean Kelly[4], Martin Nystrand[3] & Sidney K. D'Mello[1]

[1] University of Notre Dame, [2] University of Memphis, [3] University of Wisconsin-Madison, [4] University of Pittsburgh

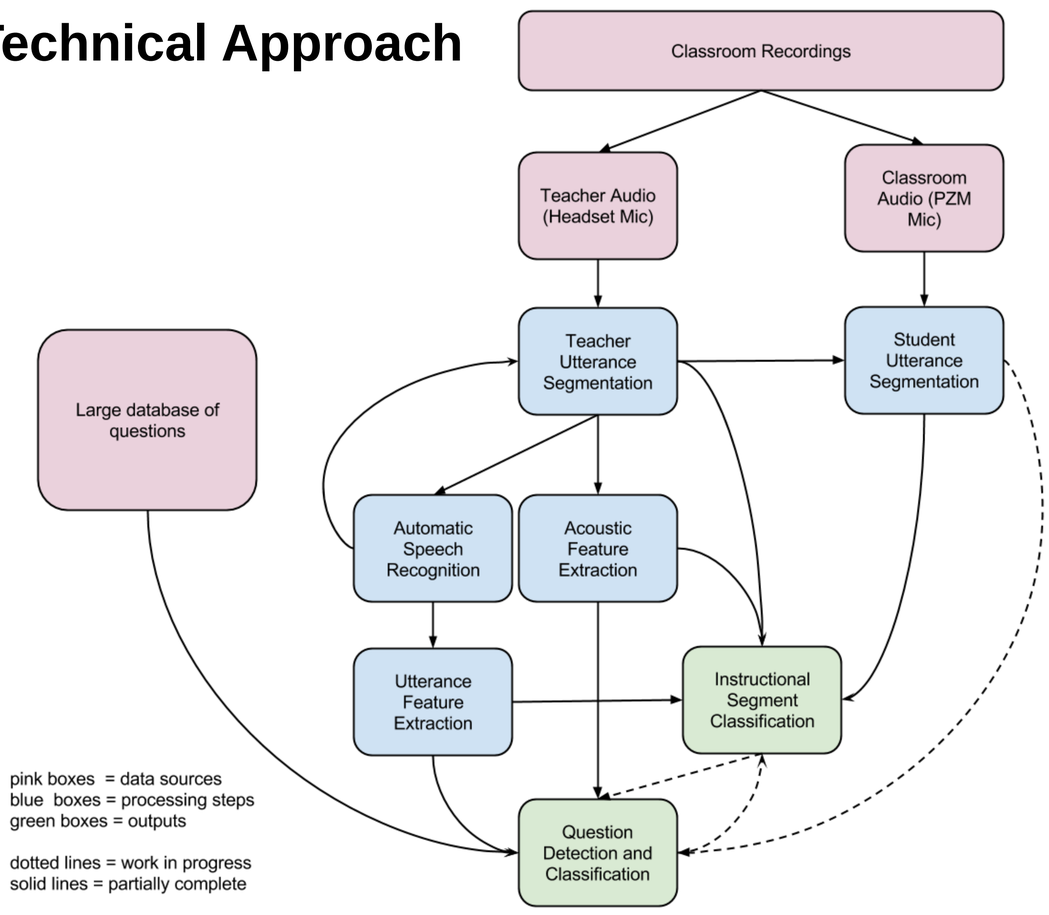
Abstract

We investigate multi-sensor modeling of teachers' instructional segments (e.g., lecture, group work) from audio recordings collected in 56 classes from eight teachers across five middle schools. Our approach fuses two sensors: a unidirectional microphone for teacher audio and a pressure zone microphone for general classroom audio. We segment and analyze the audio streams with respect to discourse timing, linguistic, and paralinguistic features. We train supervised classifiers to identify the five instructional segments that collectively comprised a majority of the data, achieving teacher-independent F_1 scores ranging from 0.49 to 0.60. With respect to individual segments, the individual sensor models and the fused model were on par for Question & Answer and Procedures & Directions segments. For Supervised Seatwork, Small Group Work, and Lecture segments, the classroom model outperformed both the teacher and fusion models. Across all segments, a multi-sensor approach led to an average 8% improvement over the state of the art approach that only analyzed teacher audio. We discuss implications of our findings for the emerging field of multimodal learning analytics.

Overview

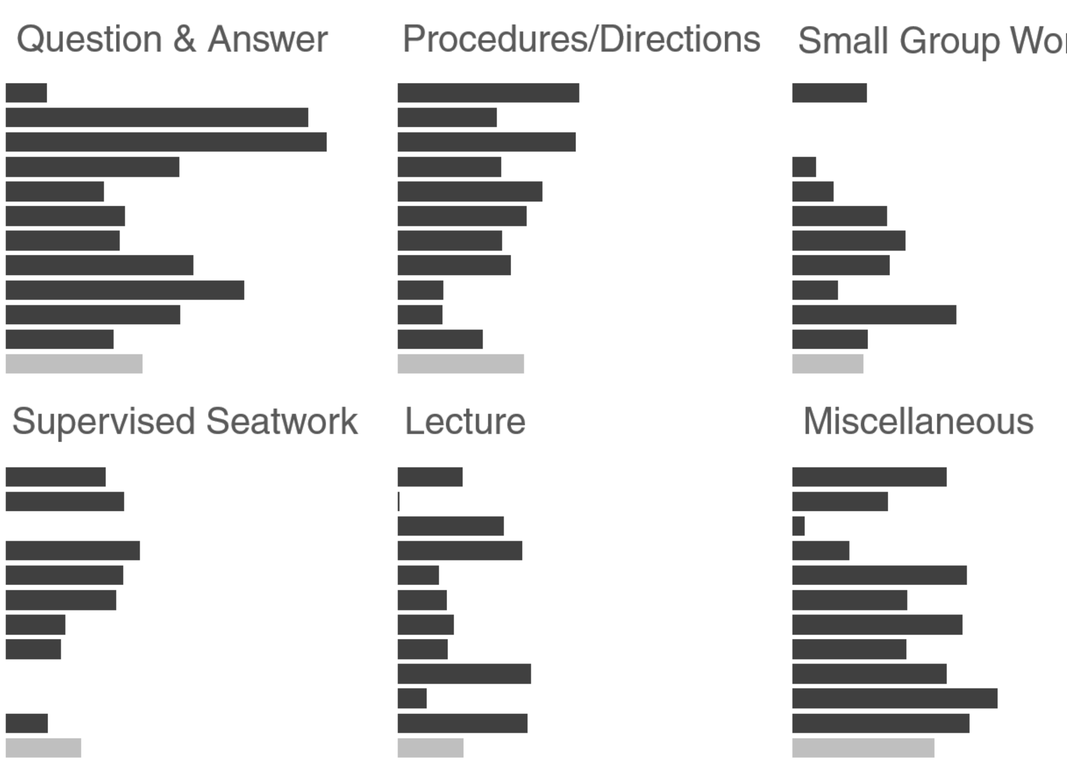
- Dialogic instruction is a powerful predictor of student engagement and achievement. How to measure the teacher's use of class time?
- This project focuses on automatic analysis of types of classroom instruction solely from audio recordings.

Technical Approach



Data Collection

- Teachers wore a portable headset microphone to capture high fidelity speech. General classroom activity was also recorded.
- Dialogic instruction was annotated by trained observers.
- 76 sessions collected, totaling 76 hours, averaging 60 ± 20 minutes each session, from 11 teachers.



Individual teachers appropriate class time differently.

Features

Teacher Channel

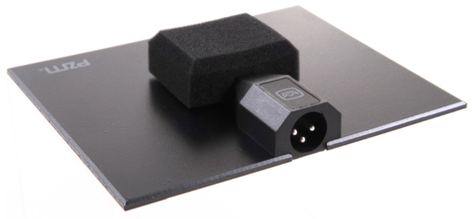
o Utterance timing	(n=13)
o Natural language	(n=37)
o Acoustic for channel	(n=30)
o Acoustic by utterance	(n=25)

Classroom Channel

o Utterance timing	(n=13)
o Acoustic by channel	(n=30)
o Acoustic by utterance	(n=25)

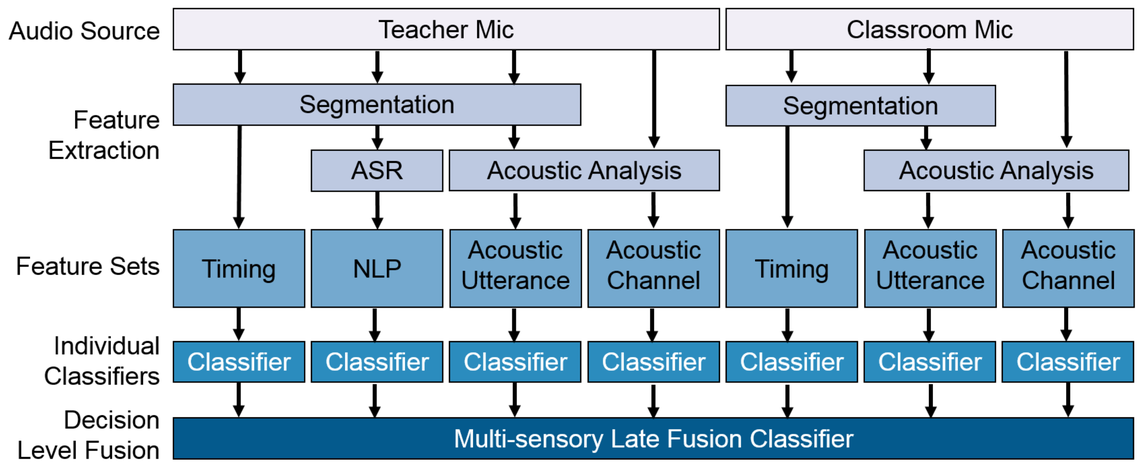


Teacher's Microphone



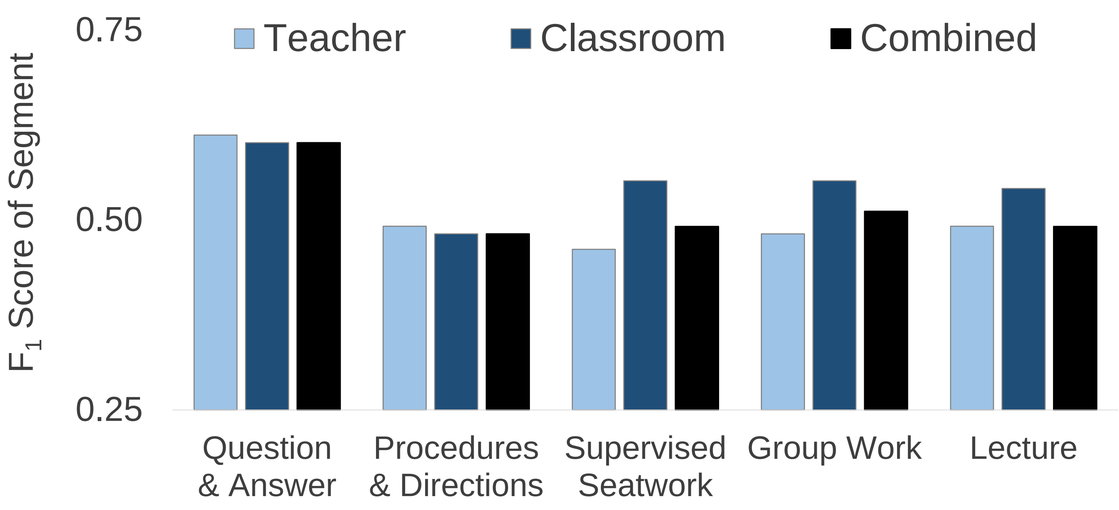
Classroom Microphone

Decision Level Fusion

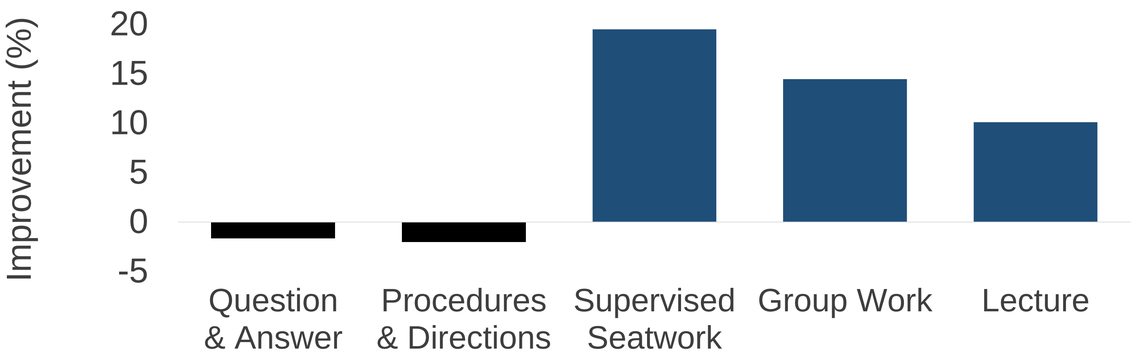


Results

- Classroom recordings are analyzed for temporal, acoustic, spectral, and natural language features.
- Machine learning models are trained to identify instructional segments and are validated independently of the teacher.



F_1 score balances precision (relevance) and recall (specificity).



Addition of the classroom channel aids in identification of student-focused segments (Supervised Seatwork, Group Work) as well as Lecture.

Next Steps

- Collection of additional class sessions
- Modeling use of instructional segments over time
- Dissemination of software/hardware to teachers, researchers, and professional development specialists

This research was supported by the Institute of Education Sciences (R305A130030)

UNIVERSITY OF NOTRE DAME
EMOTIVE COMPUTING LAB