Chemistry Pods: A Mutlimodal Real Time and Retrospective Tool for the Classroom

Khalil Anderson
Northwestern University
Evanston, Illinois
khalilanderson2023@u.northwestern.edu

Theodore Dubiel
Sony
theodore.dubiel@sony.com

Kenji Tanaka
Sony
Kenji.K.Tanaka@sony.com

Marcelo Worsley
Northwestern University
Evanston, Illinois
marcelo.worsley@northwestern.edu

ABSTRACT

Instructors are often multitasking in the classroom. This makes it increasingly difficult for them to pay attention to each individual's engagement especially during activities where students are working in groups. In this paper, we describe a system that aids instructors in supporting group collaboration by utilizing a centralized, easy-to-navigate dashboard connected to multiple pods dispersed among groups of students in a classroom or laboratory. This allows instructors to check multiple qualities of the discussion such as: the usage of instructor specified keywords, relative participation of each individual, the speech acts students are using and different emotional characteristics of group language.

CCS CONCEPTS

• Human-centered computing → Collaborative and social computing devices; **• Applied computing →** *Learning management systems.*

KEYWORDS

learning analytics, collaboration, audio processing

ACM Reference Format:
Khalil Anderson, Theodore Dubiel, Kenji Tanaka, and Marcelo Worsley. 2019. Chemistry Pods: A Mutlimodal Real Time and Retrospective Tool for the Classroom. In *2019 International Conference on Multimodal Interaction (ICMI '19), October 14–18, 2019, Suzhou, China.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3340555.3358662

ICMI '19, October 14–18, 2019, Suzhou, China

ACM ISBN 978-1-4503-6860-5/19/10.
https://doi.org/10.1145/3340555.3358662

1 INTRODUCTION

Collaboration is increasingly becoming a central component of the many learning experiences. Group collaboration and discussion has long been the norm in classes labelled as seminars, but even within more traditional lecture-based classes, professors are incorporating group collaboration more and more. Within many institutions, this approach is called active learning, which ultimately aims to engage students in more interactive in-class learning experiences. Student collaboration helps promote ideas formulation and reformulation based on the opportunity to voice ideas and receive feedback from others.

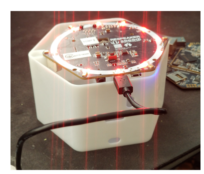

Figure 1: A chemistry pod out of the case, recording

Furthermore, several of the active learning strategies also have the goal of helping faculty members be more aware of challenges that students might be having with the current course content. With the new ways for engagement and interactivity come several new challenges for faculty members and students. While previous work has involved allowing instructors to see related information, most of these analyses are purely retrospective[4], [3], [6].

2 DESIGN

The professor or user can access the dashboard from their own mobile device, or directly on the dashboard server. The user can view analysis of different segments of the audio here. Within each time window, the dashboard displays a timeline of the conversation with annotations of discussion, silence and questions. Just below the timeline is a direction of

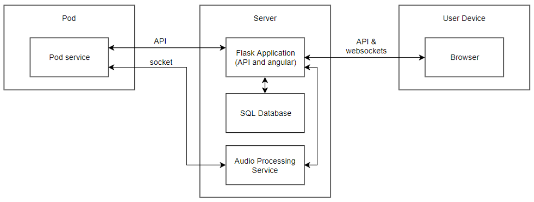

Figure 2: Architecture design of the dashboard of the chemistry system

arrival heatmap. This representation shows how many verbal contributions have been generated from different angles around the microphone. Below that heatmap is the keyword analysis box. Prior to data recording, instructors can specify which keywords they expect students to utilize during the discussion for that data or sessions. At the bottom of the page a a collection of sentiment categories.

The Chemistry pod system's dashboard collector service is the central system that each pod's audio is sent to. At this part in the system, the audio for each pod is independently processed for the features: direction of arrival (DoA), speech recognition, keyword detection, question detection, sentiment analysis, user interactivity, and speaker diarization.

While the Chemistry pod system allows for bring your own device, the system was designed to be utilized with the ReSpeaker Core v2's 6-mic array. The microphone array is what allows the system to perform several of its functions including DoA, which is achieved by calculating the differences in time it takes the audio to reach each microphone.

For speech recognition, we experimented with IBM Watson, Google Cloud Speech, Speechmatics, DeepSpeech, Sony Cloud ASR, and PocketSphinx. DeepSpeech, Pocket Sphinx, and Google Cloud Speech provide local server implementations which decrease server vulnerability. Google Cloud Speech also provided punctuation detection which is essential for question detection which helps instructors better understand each groups' collaboration process [6].

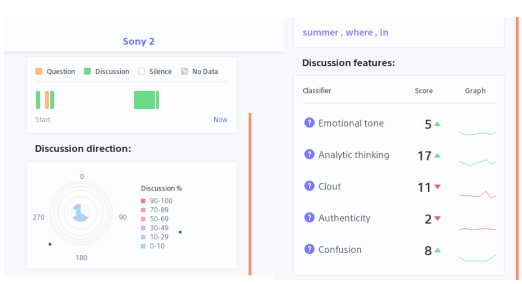

Figure 3: The features UI which show percentage of directional audio, the emotional features, keywords, and more

Following speech-to-text, additional analyses are conducted. As previously noted, users have the option to add custom keywords. Those keywords are expanded using word2vec[2] to do identify similar words. The platform also includes a tool that automatically creates a keyword list based on text from a user provided collection of paper.

Finally, the transcripts from speech-to-text are used for sentiment analysis. We use the Harvard Inquirer and Linguistic Inquiry Word Count (LIWC)[5]. They allow us to proxy for a number of measures related to psychological and/or affect state of a conversation. Beyond speech-to-text, we also process the audio data for basic prosodic features and for extracting speaker labels. The current speaker diarization implementation extends the pyAudioAnalysis [1] library and uses a k-nearest neighbors' model over short-term (200ms) and mid-term Mel-frequency cepstral coefficients (MFCC) features (2s).

3 FUTURE WORK

We have a number of additional features that are in the process of being merged into the main branch. First, we are implementing beamforming and source separation to improve speech-to-text accuracy and speaker diarization. We also have a branch that incorporates body pose estimation through 360 degree cameras. Along with this, our next steps include more user tests with instructors and students to further iterate on the system design and user interface.

REFERENCES

[1] Theodoros Giannakopoulos. 2015. pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. *PLOS ONE* 10, 12 (12 2015), 1–17. https://doi.org/10.1371/journal.pone.0144610

[2] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. arXiv:cs.CL/1301.3781

[3] Xavier Ochoa, Katherine Chiluiza, Gonzalo Méndez, Gonzalo Luzardo, Bruno Guamán, and James Castells. 2013. Expertise Estimation Based on Simple Multimodal Features. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI '13)*. ACM, New York, NY, USA, 583–590. https://doi.org/10.1145/2522848.2533789

[4] Sharon Oviatt, Adrienne Cohen, and Nadir Weibel. 2013. Multimodal Learning Analytics: Description of Math Data Corpus for ICMI Grand Challenge Workshop. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI '13)*. ACM, New York, NY, USA, 563–568. https://doi.org/10.1145/2522848.2533790 Workshop.

[5] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54. https://doi.org/10.1177/0261927X09351676

[6] Marcelo Worsley and Paulo Blikstein. 2015. Leveraging Multimodal Learning Analytics to Differentiate Student Learning Strategies. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (LAK '15)*. ACM, New York, NY, USA, 360–367. https://doi.org/10.1145/2723576.2723624