Leveraging Computationally Generated Descriptions of Audio Features to Enrich Qualitative Examinations of Sustained Uncertainty

Christina (Stina) Krist, University of Illinois Urbana-Champaign, ckrist@illinois.edu
Elizabeth B. Dyer, University of Tennessee Knoxville, edyer@utk.edu
Joshua Rosenberg, University of Tennessee Knoxville, jmrosenberg@utk.edu
Chris Palaguachi, University of Illinois Urbana-Champaign, cwp5@illinois.edu
Eugene Cox, University of Illinois Urbana-Champaign, emcox3@illinois.edu

Abstract: Prosodic features of speech, such as pitch and loudness, are important aspects of the social dimensions of learning. In particular, these features are likely related to sustained disciplinary uncertainty in collaborative STEM learning contexts. We present a case conducting an exploratory, descriptive analysis of sustained uncertainty in groupwork in a secondary mathematics lesson integrating computational and qualitative methods with audiovisual data. Results of computational audio feature extraction of loudness and pitch, combined with a transcript, were used to identify potential patterns between laughter and uncertainty.

Introduction

Qualitative research in the learning sciences has long relied on audio data as an important data source. This research has recognized the role that non-lexical features of utterances, such as intonation, pitch, and loudness in informing qualitative interpretations. However, the methods for capturing them in the transcript (see Hepburn & Boldin, 2013 for a range of examples) are time-consuming and typically useful for micro-analyses of single events. Computational tools such as voice activity detection (VAD) can be useful for quantifying features of audio data that transcript-based representations do not easily attend to but are often used for goals around prediction or automation (Slyman et al., 2021). This project seeks to explore potential uses of computational tools such as VAD for descriptive purposes within a qualitative methodological paradigm. In this paper, we present an in-the-weeds example exploring sustained uncertainty in collaborative student problem-solving in mathematics.

Background

Sustained uncertainty in problem solving in STEM

Exploration, inquiry, and modeling — important interdisciplinary processes that support student learning (NRC, 2012; NGA & CCSSO, 2010) — provide opportunities to make sense of disciplinary questions or problems. Productive engagement in these processes likely involves sustained uncertainty negotiated over time (Rosenberg et al., 2022; Watkins et al., 2018). However, uncertainty is often risky, devalued, and discouraged in schools, especially in science and mathematics classrooms (Archer et al., 2017). Previous research exploring disciplinary uncertainty from an interactional perspective has highlighted its complex and contextually-situated nature. For example, Watkins et al. (2018) identified a student repeatedly bringing up an idea until it evolved into a question that others picked up. Similarly, explicit expression of uncertainty may be a poor indicator of epistemic stance. For example, “I don’t know” might signal uncertainty in knowledge, but also “I don’t want to talk about this anymore,” or simply, “I am disengaged.” (Tsui, 1991). Attending to the prosodic features of speech could help tease apart nuances in these uses that are not apparent in the lexical (e.g., words used) features of speech alone.

Prosodic features related to uncertainty

The features of discourse relevant to communicating uncertainty have been examined both qualitatively and using computational methods. In both cases, it is common to look at the linguistic and semantic features of discourse. Qualitatively, these analyses have emphasized the social and rhetorical functions of uncertainty, such as how various linguistic features communicate epistemic stance to readers (e.g., Hyland, 2005); and how scientists’ discussions about different claims about data include a gradual softening of assertions (i.e., decreasing the certainty of claims) until they come into alignment with one another (Lynch, 1985). Computationally, these analyses have been used to detect when questions are asked (e.g., Hirsh, 2019) or what intonation speakers use (Hübscher et al., 2017). In addition to identifying questions, there are other linguistic features of semantic uncertainty, including adjectives/adverbs such as “probable, likely, unsure, perhaps”; auxiliaries such as “may, might, can, would, should”; conjunctions such as “if, whether”; and specific verbs and related nouns, such as "propose/proposal; question; investigate; consider;" etc. (Szarvas et al., 2012). These features have been used to detect uncertainty in speech automatically using probabilistic models (e.g., Jean et al., 2016).

Other studies have examined paralinguistic features such as intonation, as well as descriptive features of audio data such as the number and duration of turns. For example, Berger and colleagues (Berger & Calabrese, 1975) showed a correlational link between the amount of verbal communication in an interaction and the degree of uncertainty in the talk: as amount of talk increases, uncertainty decreases. Intonation is also an important marker for differentiating between uses of different types of question words (tag questions, wh- questions, inverted questions, and repetition questions) between speakers of different languages (e.g., Farais, 2013).

Methods

In this paper, we present an in-the-weeds example of how we are "layering on" automatically-detected non-lexical features of audio data in order to increase depth/complexity of descriptive qualitative analysis. We use a methodological approach integrating computational and qualitative techniques within a qualitative methodological paradigm focused on rich description. To do so, we use a single illustrative case or data episode, which allows us to deeply explore the potential features and characteristics of sustained uncertainty. We position this approach in contrast to prediction-oriented computational work, aiming to leverage large data to identify relationships among variables. Our goals and aim are to layer complementary perspectives about a common data episode to develop a descriptive account that incorporates a multiplicity of analytical perspectives spanning both computational and qualitative findings. Additionally, this analytical process involves placing these perspectives in conversation with one another to re-interpret findings with the addition and revision of layers of analyses, similar to the iterative hypothesis generation and testing used in qualitative analysis of video (Engle et al., 2007). This integration is guided towards a goal of rich multi-faceted description, rather than convergence on a single or simple characterization or label.

We selected an approximately 20-minute segment of video/audio of groupwork from a high school mathematics classroom, focusing on a single group. This data comes from a larger study, including video/audio recorded classroom lessons in high school mathematics from 10 teachers. The segment comes from a lesson on solving trigonometric equations as part of a math course for grade 10 and 11 students. This course and teacher, Mrs. Perry, was selected because the teacher used significant amounts of groupwork and previous research has documented that Mrs. Perry's teaching practice is responsive to student thinking (Dyer & Sherin, 2016).

The specific lesson and segment were selected to include a 4-minute episode of collaborative mathematical exploration among the teacher and students identified from previous qualitative analysis (Dyer et al., 2021). This analysis identified shifts in epistemic agency and authority among the participants, as well as sustained disciplinary exploration over several minutes. We selected this larger episode because we expected that exploration would involve frequent instances of disciplinary uncertainty and/or other forms of uncertainty that were not immediately resolved. The larger segment corresponded to the beginning and end of the first portion of groupwork in the lesson and thus included portions with the teacher present and not present. We hypothesized this would provide variation in the interactional patterns and structures of the group interactions over time.

To analyze the episode computationally, we use prosodic feature extraction audio analytics techniques from *openSMILE*, an open-source audio processing program (Eyben et al., 2010) in conjunction with a time-coded transcript of the episode. For prosodic feature extraction, we focused on pitch and loudness as two features we hypothesized would be related to uncertainty based on prior work. Both outputs are provided at the frame level, and thus, we created aggregate measures and displays for each turn of talk as a unit of analysis. These include the maximum value, minimum value, mean value, and smoothed line graph of values over time.

Findings

To facilitate the interpretation of our results, we first provide a short summary of the episode. The episode comes from a small group of four students working together on a task that asked students to solve the equation $-15=20cos(30x)$. Across the episode, the students consider graphical and equation-based approaches to find different solutions. The first approximately 13.5 minutes of the episode involved the students discussing and working with one another, followed by the teacher visiting the group for around 12.2 minutes, and concluding with less than one minute of the group talking before the class transitions to whole-class student presentations (not included in the episode).

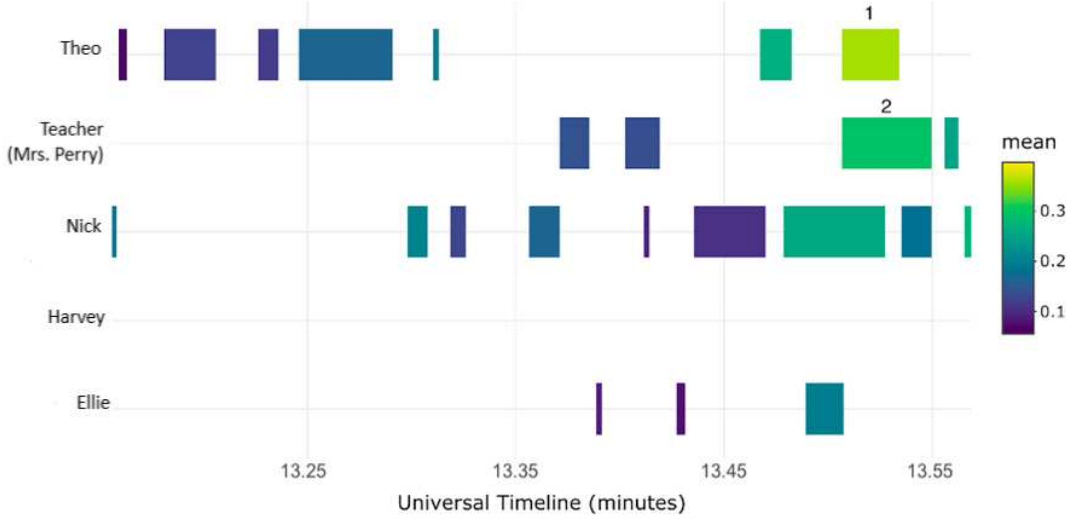
Case: Examining anomalous moments of loudness

We present here a case example that emerged from our explorations of the loudness of turns during the focal episode. Figure 1 shows the mean normalized loudness of speech by segment, with each row representing an individual speaker. Note the most yellow-appearing segment, located in the top-right of the graph and annotated with the number "1.". We can see that this utterance from Theo's had the highest average loudness. It also had the highest maximum loudness. The other speakers' turns during this time were also louder than the turns preceding it.

In examining the transcript, during segment 1, Theo said, "I know *((LAUGHTER))*! I still don't understand, though." During segment 2, the teacher said, "*((LAUGHTER))* And you still didn't get it *((LAUGHTER))*!"–the segment with the second-highest mean loudness. Examining the frame-by-frame measures of loudness within each turn shows that the loudest portions of each segment were not when the individuals were laughing. Instead, although laughter occurred *during* the segments with the highest mean loudness, the laughter was not registered as the loudest part of the segment.

Figure 1

The Mean Loudness Per Utterance During the Period of Sustained Uncertainty



These patterns suggest two things. First, laughter may accompany (or indirectly be the cause of) louder utterances. Perhaps students became louder when they laughed because their laughter represented an expression of relief or a breaking of tension caused by uncertainty. The words they spoke may have been louder in order to match the tension-breaking tone of the laughter. Alternatively, perhaps these segments were not exceptionally louder than average but were instead louder than the unusually quiet turns that preceded them—turns that were quiet because they were tentative or embarrassed about whatever it was that they were *still* not understanding (Theo).

Discussion and future work

We have presented an analytic case of leveraging computational tools in service of qualitative methods. Specifically, we used prosodic feature extraction (loudness) to initiate grounded hypotheses for future exploration. Rather than serving as an endpoint for prediction, the computational analysis instead functioned as an early pattern-detection tool. In our ongoing analytic work, we are expanding our analysis beyond the 20-min segment described to explore hypotheses generated within the segment. For example, how often does loudness correlate with laughter? How do moments of laughter (and/or loudness) reflect uncertainty, or other epistemic markers?

This example provides an initial demonstration for how prosodic features are helpful as accompaniments to intensive qualitative analysis—and not a replacement for such methods. Specifically, we found that the prosodic features of pitch and loudness are most helpful in a descriptive, rather than a highly inferential, manner. For example, the transcript allowed our team to identify the association between segments of high average loudness and laughter. Similarly, examining the pitch within segments led our team to interpret the role of laughter as a co-occurrence with loud speech rather than the primary driver of high average loudness.

Our future work will continue to leverage additional computational tools, such as automatic speech recognition, as an additional layer and visualize select components from that output (e.g., use of hedging words; use of question words) in conjunction with prosodic features such as loudness. Another feature under development is a way of representing the absence of speech, including pauses within turns, using voice activity detection (VAD) algorithms. Though unconventional and atypical, we encourage other researchers to creatively and critically leverage these tools for computational tools such as VAD for descriptive purposes within a qualitative methodological paradigm, rather than only for automated prediction.

References

Archer, L., Dawson, E., DeWitt, J., Godec, S., King, H., Mau, A., ... & Seakins, A. (2017). Killing curiosity? An analysis of celebrated identity performances among teachers and students in nine London secondary science classrooms. *Science Education*, *101*(5), 741-764.

Berger, C., & Calabrese, R. (1975). Some explorations in initial interactions and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research, 1*, 99-112.

Dyer, E. B., Parr, E. D., Machaka, N., & Krist, C. (2021). Understanding joint exploration: The epistemic positioning underlying collaborative activity in a secondary mathematics classroom. Research Report. *43rd Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education*. PME-NA 2021.

Dyer, E. B., & Sherin, M. G. (2016). Instructional Reasoning about Interpretations of Student Thinking that Supports Responsive Teaching in Secondary Mathematics. *ZDM, 48*(1–2), 69–82. https://doi.org/10.1007/s11858-015-0740-1

Engle, R. A., Conant, F. R., & Greeno, J. G. (2007). Progressive refinement of hypotheses in video-supported research. In Goldman, R., Pea, R., Barron, B., & Derry, S. J. (Eds.). (2014). *Video research in the learning sciences*. Routledge.

Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1459-1462).

Farías, M. G. V. (2013). A comparative analysis of intonation between Spanish and English speakers in tag questions, wh-questions, inverted questions, and repetition questions. *Revista Brasileira de Linguística Aplicada*, *13*, 1061-1083.

Hepburn, A., & Bolden, G. B. (2013). The conversation analytic approach to transcription. *The handbook of conversation analysis*, *1*, 57-76.

Hirsch, R. V. (2019). *Automatic Question Detection From Prosodic Speech Analysis* (Doctoral dissertation, Colorado State University).

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language, 37*(1), 24-41.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173-192.

Jean, P. A., Harispe, S., Ranwez, S., Bellot, P., & Montmain, J. (2016, June). Uncertainty detection in natural language: A probabilistic model. In *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics* (pp. 1-10).

Lynch, M. (1985). Discipline and the material form of images: An analysis of scientific visibility. *Social Studies of Science*, *15*(1), 37-66.

National Governors Association Center for Best Practices and Council of Chief State School Officers. 2009. Forty-nine states and territories join common core standards initiative. Washington, DC: NGA Center and CCSSO. Available at: www. corestandards.org.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

Rosenberg, J. M., Kubsch, M., Wagenmakers, E. J., & Dogucu, M. (2022). Making sense of uncertainty in the science classroom. *Science & Education, 31*(5), 1239-1262.

Slyman, E., Daw, C., Skrabut, M., Usenko, A., & Hutchinson, B. (2021). Fine-Grained Classroom Activity Detection from Audio with Neural Networks. *arXiv preprint arXiv:2107.14369.*

Szarvas, G., Vincze, V., Farkas, R., Móra, G., & Gurevych, I. (2012). Cross-genre and cross-domain detection of semantic uncertainty. *Computational Linguistics*, *38*(2), 335-367.

Tsui, A. B. (1991). The pragmatic functions of I don't know. *Text-Interdisciplinary Journal for the Study of Discourse*, *11*(4), 607-622.

Watkins, J., Hammer, D., Radoff, J., Jaber, L. Z., & Phillips, A. M. (2018). Positioning as not-understanding: The value of showing uncertainty for engaging in science. *Journal of Research in Science Teaching*, *55*(4), 573-599.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1920796.