Towards Automated Measurement of Active Listening Skills in Collaborative Problem Solving Using Multimodal Learning Analytics and Large Language Models

Xiaomeng Huang, Xavier Ochoa, Shiyu Zhang, Madhumitha Gopalakrishnan
New York University
xiaomeng.huang@nyu.edu, xavier.ochoa@nyu.edu, sz3561@nyu.edu, mg7584@nyu.edu

ABSTRACT: Active listening is a critical social–emotional skill that supports effective and inclusive communication in collaborative learning. Although well-studied in dyadic communication, it remains underexplored in small-group collaborative problem solving (CPS). However, measuring active listening at scale is difficult: teachers rarely have the capacity to observe each learner, in real time, across multiple groups. Existing measurement approaches rely on self-report or manual coding, both of which are labor-intensive and prone to bias. In this study, we propose an automated pipeline for measuring students' active listening skills in CPS using multimodal learning analytics and large language models (LLMs). Guided by a theoretical model of active listening, we (1) extract layered collaboration contexts from transcripts using LLMs, (2) automatically detect behavioral features from multimodal data (e.g., audio, video), and (3) fuse contextual and behavioral features into a transformer-based predictive model to estimate active listening skill levels along the dimensions of the theoretical model. We outline the methodology and discuss implications.

Keywords: Active Listening; Collaborative Problem Solving; Multimodal Learning Analytics; Large Language Model

1 INTRODUCTION

Effective communication starts with listening (Brownell, 2023). In small-group collaborative problem solving (CPS), one of the most common breakdowns occurs when learners fail to establish and maintain a shared problem space (Barron, 2003). Active listening—attending to, making sense of, and responding to others' ideas appropriately—is central to this process (Wise et al., 2013), yet remains underexamined in CPS. Existing measurement approaches largely rely on self-report surveys or manual coding (Worthington & Bodie, 2017), which are labor-intensive, prone to bias, and not scalable for classroom use. Two core challenges make active listening difficult to measure in CPS. First, active listening is **multimodal**: learners signal listening through speech, gesture, gaze, and other nonverbal behaviors. Multimodal Learning Analytics (MmLA) offers a promising way to automatically detect and analyze such behaviors (Ochoa, 2022). Second, active listening is **contextual**: the meaning of a behavior depends on what is happening in the interaction. For example, looking at a speaker may indicate attentiveness when ideas are being presented, but could indicate disengagement when one only looks without contributing during negotiation phases. Because of this contextual variability, any automated approach must integrate context information into the measurement pipeline.

Building on our prior work on Idea Tracing Analysis (ITA) (Huang et al., under review), which captures layered collaboration contexts in CPS through idea evolution analysis, we propose an automated measurement pipeline that (1) derives layered collaboration contexts from transcripts using large language models (LLMs), (2) situates multimodal behaviors within those detected contexts using MmLA, and (3) fuses contextual and behavioral features in a transformer-based predictive model to estimate students' active listening skill levels aligned with a theoretical model. By implementing this pipeline, our goal is to make theory-aligned measurement scalable and lay the foundation for delivering targeted feedback to support learning of this critical skill.

2 METHODOLOGY

As illustrated in Figure 1, our pipeline operationalizes the HURIER model of active listening (Brownell, 2023) through four automated components: data preparation, context extraction, behavioral feature extraction, and predictive modeling. The data for this study was collected in the Summer of 2024 from 30 graduate students who worked in groups of 3 or 4 to tackle a design challenge.

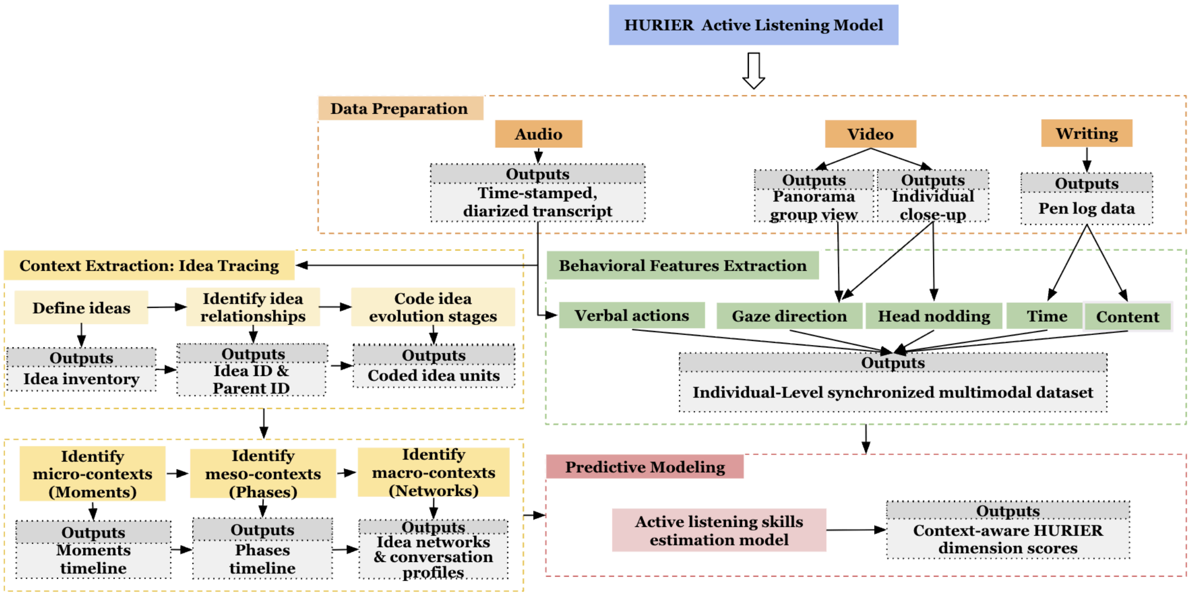

Figure 1: Automated Active Listening Measurement Pipeline

Data Preparation. Guided by the HURIER model, which decomposes active listening into six sub-skills (hearing, understanding, remembering, interpreting, evaluating, responding), we mapped behavioral indicators from self-reported, behaviorally anchored survey items into collaboration contexts and identified the required modalities: audio, video, and writing. We used a 360-degree conference system to record audio and video, and smart pens to record writing traces. Audio was transcribed and speaker-diarized using WhisperX (Bain et al., 2023), yielding 3,204 utterances in total. Videos were rendered with both a panoramic group view and individual close-up views, and pen log data were stored as JSON files and converted to writing timestamps and content per participant.

Behavioral Features Extraction. From the HURIER model, we identified 19 verbal actions associated with the six active listening sub-skills. We first developed a codebook: three coders adapted initial codes from the HURIER model, coded 20% of the data, and met for more than 40 hours to refine and finalize the coding scheme. We then iteratively optimized LLM-based coding (ChatGPT-5 API) through codebook refinement, prompt iteration, and context engineering, and applied the final prompts to the remaining data. For nonverbal behaviors, we extracted gaze direction, head nodding, and writing activity. Gaze direction was detected from video using computer vision (OpenCV), aligned with speaker information, and head nodding was detected using the CCDb-HG model (Vuillecard et al., 2024). Writing traces were extracted from pen log data.

Context Extraction. We employ Idea Tracing Analysis (ITA) (Huang et al., under review), which traces how group ideas evolve (e.g., when ideas are introduced, expanded, evaluated, accepted, or refuted) to characterize layered collaboration contexts. These evolution stages are then used to derive contextual indicators that situate multimodal behaviors within meaningful interaction moments. The extraction process is implemented using LLM-based coding (ChatGPT-5 API), with iterative refinement of codebook definitions, prompts, and context engineering.

Predictive Modeling. We fuse contextual indicators and multimodal behavioral features and use a transformer-based model to estimate context-aware active listening skill levels aligned with the HURIER dimensions. Transformers are selected because their attention mechanisms allow the model to jointly weigh contextual information and behavioral signals when generating predictions.

3 IMPLICATIONS AND FUTURE WORK

The pipeline we introduce offers a scalable pathway for measuring active listening in authentic CPS settings, reducing reliance on manual coding. Automated, context-aware indicators can support formative feedback and make listening behaviors more visible to teachers and learners. Future work will validate model performance across tasks and populations, refine mapping between behaviors and HURIER dimensions, and explore real-time deployment in feedback systems.

REFERENCES

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). *WhisperX: Time-Accurate Speech Transcription of Long-Form Audio* (No. arXiv:2303.00747). arXiv. https://doi.org/10.48550/arXiv.2303.00747

Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Brownell, J. (2023). *Listening: Attitudes, Principles, and Skills* (7th ed.). Routledge. https://doi.org/10.4324/9781003316794

Huang, X. , Ochoa, X., Gopalakrishnan, M., Zhang, S.. (Under Review). Idea Tracing Analysis (ITA): A Multimodal Micro-Ecological Methodology to Contextualize Learners' Behaviors, Patterns and Profiles in Collaborative Problem Solving.

Ochoa, X. (2022). Chapter 6: Multimodal Learning Analytics—Rationale, Process, Examples, and Direction. In *The Handbook of Learning Analytics* (pp. 54–65). Soc. Learn. Analytics Res.

Vuillecard, P., Farkhondeh, A., Villamizar, M., & Odobez, J.-M. (2024). CCDb-HG: Novel Annotations and Gaze-Aware Representations for Head Gesture Recognition. *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, 1–9. https://doi.org/10.1109/FG59268.2024.10581954

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, *41*(2), 323–343. https://doi.org/10.1007/s11251-012-9230-9

Worthington, D. L., & Bodie, G. (Eds.). (2017). *The sourcebook of listening research: Methodology and measures*. John Wiley & Sons, Inc.