Westat

Improving Lives Through Research®

# Applying Machine-Learning to Interviewer Monitoring and Question Assessment


**Hanyu Sun, Ting Yan, Anil Battalahalli, *Westat***

# Background

- Survey organizations has long used Computer-Assisted Recorded Interviewing (CARI) as a quality control tool:
  - Interviewer performance
  - Question Assessment
- Conventionally, a human coder needs to first listen to the recording and then evaluate and code features of the question-answer sequence using a pre-specified coding scheme:
  - Labor intensive and time consuming

# Question-Answer Sequence

- Example:

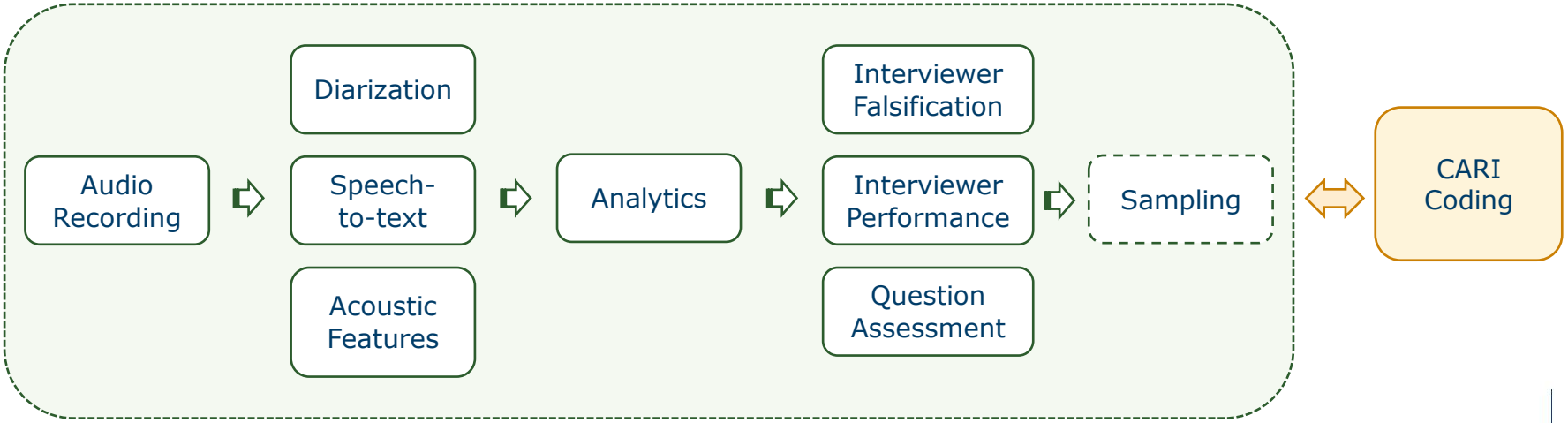
  **Interviewer:** In general, would you say your mental health is excellent, very good, good, fair, or poor?

  **Respondent:** Good.

  **Interviewer:** Thank you.

# Audio Pipeline: Overview

- Westat developed a proprietary Machine Learning Audio Pipeline to process and analyze audio recordings of survey interviews to facilitate quality control (Sun and Yan, 2023; Yan, Sun, and Battalahalli, Accepted)



# Audio Pipeline: Components

- Speaker diarization:
  - Detect who spoke at which turn in a question-answer sequence to output the number of different speakers (pyannote)
    - Produce turn-level measures (e.g., number of interviewer turns, turn duration)
- Speech-to-text:
  - Transcribe question-answer sequence at turn level (OpenAI Whisper) and then compare turn-level transcript to the question wording by distance scores:
    - OCR text recognition (EasyOCR) to extract the exact question wording from the Blaise screenshot
    - Jaccard distance and Cosine distance
- Acoustic feature extraction:
  - Extract acoustic parameters at turn level (OpenSMILE)
    - Train models to predict response difficulty with human coded results

# Utilize Audio Pipeline in Quality Control

- Interviewer performance (Sun and Yan, 2023):
  - Flag recordings with 0 or only 1 speaker as high risk of falsification
  - Use distance scores to determine whether FI read the question verbatim and whether FI maintained question meaning
  - Prioritize/Sample cases or interviewers for conventional CARI coding
- Question assessment (Yan, Sun, and Battalahalli, Accepted):
  - Rank survey questions on performance measures:
    - Number of turns, Duration of respondent 1st turn after the interviewer finishes reading the question, total duration with speech, long pauses, positive emotions
  - Prioritize/Sample survey questions for conventional CARI coding

# Discussion

- The quality of the audio recording affects the performance of the pipeline
  - How to identify audio recordings of poor quality before sent to the pipeline for processing?
  - How to account for the quality of audio recordings during the pipeline processing?
  - Interviewer training?
- Utilize human coded results from the conventional CARI coding to improve the performance of the pipeline
- Communication with stakeholders

# Thank you

Hanyu Sun

hanyusun@westat.com

westat.com

# References

Sun, H. and Yan, T. (2023). “Applying Machine Learning to the Evaluation of Interviewer Performance.” *Survey Practice* 16 (1). https://doi.org/10.29115/SP-2023-0007.

Yan, T., Sun, H., and Battalahalli, A. (Accepted). “Applying Machine Learning to Survey Question Assessment.” *Survey Practice*.