Mayo Clinic - ASU Synapse
10/21/2025

Touchpoint Meeting - An AI-Powered Communication Coach for Healthcare Professionals

Tanner Hochberg, Elijah Don Alex Roussas, Ian Marcon, Ethan Vanderpool

Charter Outline

Background/Objective

Background:

1. Lack of scalable, effective, and objective ways to practice and refine communication skills.
2. The emotional toll of difficult conversations contributes to **practitioner burnout** and empathy erosion
3. A lack of empathetic communication can amplify a patient's fear and grief, leading to a poorer emotional state, lower patient satisfaction, and decreased trust in their providers.

Objective:

Design and validate an AI/ML-based tool aimed at delivering timely, objective, and continuous feedback on conversational skills.

Goals	Technical Indictors
1. Explore skill development and training 2. Validate objective feedback and assessment 3. Improve well-being and professional development	1. Data input and processing (audio, video) 2. Feedback delivery method, effectiveness 3. Feedback Latency 4. Reduction in faculty coaching time

Team Members	Stakeholders	Business Value & KPIs	Project Milestones
Tanner Hochberg Elijah Don Alexander Roussas Ian Marcon Ethan Vanderpool	1. Healthcare Professionals & Medical Trainees 2. Patients & Families 3. Medical Educators & Training Institutions 4. Healthcare Administrators & Institutions	• Improved Patient-Healthcare provider relationship • Improved Patient Satisfaction • Mitigate Healthcare professional Burnout • Improved Healthcare Professional/Trainee training and outcomes	1. Project Definition and Alignment 2. Needs & Model Research 3. Model & Workflow Design 4. Prototype or Pilot Development 5. Implementation & Validation 6. Prototype or Pilot Development 2 7. Implementation & Validation 2 8. Scale & Sustainability Planning

Current Design Concept:

An AI-Powered Communication Coach for Healthcare Professionals

Core Concept

Standalone app that helps practitioners improve communication

The tool *records doctor-patient conversations and leverages AI* to provide immediate, actionable feedback on empathy and communication quality.

How it Works: Simple Workflow

Record: Practitioner uses app to securely record patient conversation

Analyze: Immediately after, AI performs audio analysis on the content (***what was said***) and the vocal tone (***how it was said***)

Receive Insights:
- immediate, ***concise summary*** for quick reflection
- ***comprehensive dashboard*** for deeper learning

Key Design Features:

Dual Feedback: Provides quick, instant summary (text/email) + detailed dashboard

Non-Intrusive: Analyzes conversations post-encounter

Holistic Analysis: Evaluates what was said + how it was said for a complete picture

Scalable Insights: Secure data for individual growth and organizational-level insights

Overall Architecture: A 4-Step Pipeline

Our software follows a modular pipeline to analyze recorded conversations and generate feedback:

- Secure Transcription: Convert audio to text, identifying speakers (using STT APIs)
- Multi-Modal Analysis Engine: Analyze both text and vocal tone: 2A. Textual Analysis: Use a Large Language Model (LLM API – e.g., Gemini) to assess content (empathy, clarity, questions answered). 2B. Vocal Analysis: Use Speech Emotion Recognition (SER) & feature extraction tools to analyze tone, pace, pitch for both doctor and patient.
- Synthesis & Scoring: Custom logic combines text and vocal analyses into a unified report with scores and insights.
- User-Facing Dashboard: Present feedback clearly via immediate summaries and a detailed dashboard.

Vocal Analysis Module: Core Architecture

Input: Conversation Audio

- The process starts with the raw audio file of the doctor-patient conversation.

Step 1: Vocal Feature Extraction

- The audio is processed by two specialized Python libraries simultaneously to extract a comprehensive set of vocal features.
- Parselmouth (Praat): Used to extract precise, clinically-validated voice quality metrics (e.g., jitter, shimmer) to ensure scientific accuracy.
- Librosa: Used to efficiently extract standard audio features (e.g., MFCCs) that are essential for machine learning models.

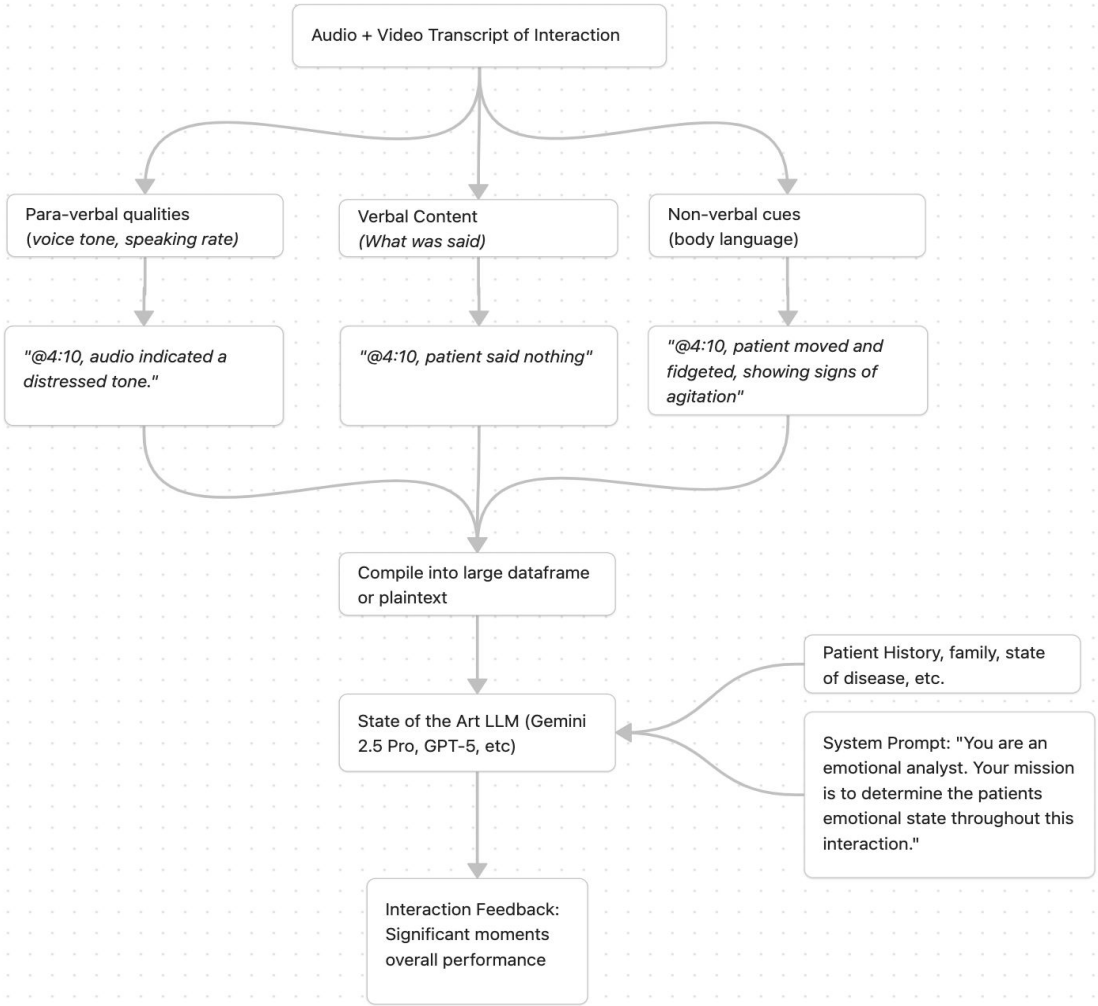
Step 2: Emotion Recognition Model

- The combined vocal features are fed into a custom-trained emotion recognition model.
- This model is built by taking a state-of-the-art, pre-trained foundation model from Hugging Face and fine-tuning it using the SpeechBrain toolkit. This allows us to adapt a powerful general model to the specific nuances of clinical conversations.

Output: Structured Emotion Analysis

- The model outputs the final analysis in a structured JSON format, ready for the next module.

FUTURE DIRECTION



Year/Quarter/Month

Thank you