

Wearable AI system can detect a conversation's tone

February 1 2017, by Adam Conner-Simons



Mohammad Ghassemi and Tuka Alhanai converse with the wearable. Credit: Jason Dorfman, MIT CSAIL

It's a fact of nature that a single conversation can be interpreted in very different ways. For people with anxiety or conditions such as Asperger's,

this can make social situations extremely stressful. But what if there was a more objective way to measure and understand our interactions?

Researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and Institute of Medical Engineering and Science (IMES) say that they've gotten closer to a potential solution: an artificially intelligent, wearable system that can predict if a conversation is happy, sad, or neutral based on a person's speech patterns and vitals.

"Imagine if, at the end of a conversation, you could rewind it and see the moments when the people around you felt the most anxious," says graduate student Tuka Alhanai, who co-authored a related paper with PhD candidate Mohammad Ghassemi that they will present at next week's Association for the Advancement of Artificial Intelligence (AAAI) conference in San Francisco. "Our work is a step in this direction, suggesting that we may not be that far away from a world where people can have an AI social coach right in their pocket."

As a participant tells a story, the system can analyze audio, text transcriptions, and physiological signals to determine the overall tone of the story with 83 percent accuracy. Using deep-learning techniques, the system can also provide a "sentiment score" for specific five-second intervals within a conversation.

"As far as we know, this is the first experiment that collects both physical data and speech data in a passive but robust way, even while subjects are having natural, unstructured interactions," says Ghassemi. "Our results show that it's possible to classify the emotional tone of conversations in real-time."

The researchers say that the system's performance would be further improved by having multiple people in a conversation use it on their smartwatches, creating more data to be analyzed by their algorithms. The

team is keen to point out that they developed the system with privacy strongly in mind: The algorithm runs locally on a user's device as a way of protecting personal information. (Alhanai says that a consumer version would obviously need clear protocols for getting consent from the people involved in the conversations.)

How it works

Many emotion-detection studies show participants "happy" and "sad" videos, or ask them to artificially act out specific emotive states. But in an effort to elicit more organic emotions, the team instead asked subjects to tell a happy or sad story of their own choosing.

Subjects wore a Samsung Simband, a research device that captures high-resolution physiological waveforms to measure features such as movement, heart rate, blood pressure, blood flow, and skin temperature. The system also captured audio data and text transcripts to analyze the speaker's tone, pitch, energy, and vocabulary.

"The team's usage of consumer market devices for collecting physiological data and speech data shows how close we are to having such tools in everyday devices," says Björn Schuller, professor and chair of Complex and Intelligent Systems at the University of Passau in Germany, who was not involved in the research. "Technology could soon feel much more emotionally intelligent, or even 'emotional' itself."

After capturing 31 different conversations of several minutes each, the team trained two algorithms on the data: One classified the overall nature of a conversation as either happy or sad, while the second classified each five-second block of every [conversation](#) as positive, negative, or neutral.

Alhanai notes that, in traditional neural networks, all features about the

data are provided to the algorithm at the base of the network. In contrast, her team found that they could improve performance by organizing different features at the various layers of the network.

"The system picks up on how, for example, the sentiment in the text transcription was more abstract than the raw accelerometer data," says Alhanai. "It's quite remarkable that a machine could approximate how we humans perceive these interactions, without significant input from us as researchers."

Results

Indeed, the algorithm's findings align well with what we humans might expect to observe. For instance, long pauses and monotonous vocal tones were associated with sadder stories, while more energetic, varied speech patterns were associated with happier ones. In terms of body language, sadder stories were also strongly associated with increased fidgeting and cardiovascular activity, as well as certain postures like putting one's hands on one's face.

On average, the model could classify the mood of each five-second interval with an accuracy that was approximately 18 percent above chance, and a full 7.5 percent better than existing approaches.

The algorithm is not yet reliable enough to be deployed for social coaching, but Alhanai says that they are actively working toward that goal. For future work the team plans to collect data on a much larger scale, potentially using commercial devices such as the Apple Watch that would allow them to more easily implement the system out in the world.

"Our next step is to improve the algorithm's emotional granularity so that it is more accurate at calling out boring, tense, and excited moments, rather than just labeling interactions as 'positive' or 'negative,'" says

Alhanai. "Developing technology that can take the pulse of human emotions has the potential to dramatically improve how we communicate with each other."

Provided by Massachusetts Institute of Technology

Citation: Wearable AI system can detect a conversation's tone (2017, February 1) retrieved 18 April 2024 from <https://techxplore.com/news/2017-02-wearable-ai-conversation-tone.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.