John Edwards

Signal Processing Comes to the Senses

Signal processing is making an impact in taste, sentiment, and touch research

Senses allow us to gather information about our surroundings, playing an important role in everyday life. Research teams worldwide are now turning to signal processing to help replicate, augment, or interpret human senses.

A matter of taste

University of Cambridge researchers have trained a robot chef to assess the saltiness of food at different stages of the chewing process, imitating a similar practice in humans. The project promises to contribute to the development of automated or semiautomated food preparation by helping robots learn what tastes good and what doesn't.

"We usually refer to it as a *robotic chef*," says Cambridge graduate student Grzegorz Sochacki, who works with Arsen Abdulali, a research associate in the university's engineering department, and Muhammad W. Chughtai, a senior scientist at appliance manufacturer Beko.

Collaborating in the university's Bio-Inspired Robotics Laboratory, the researchers have created a custom-made mobile tasting platform. "The bird's-eye view of our setup is a robotic arm with a salinity probe attached to it," Sochacki says. The robot is a Universal Robot (UR5) arm, an off-the-shelf device manufactured by Denmark-based Universal Robots. "We usually fit [the arm] with sensors and cooking utensils," he says (Figure 1).

In the case of this study, the probe was a conductance sensor created out of two platinum electrodes, Sochacki says. AC voltage is set across the electrodes with a control circuit and the resulting current is measured. The final value is dependent on several factors, including the sampled food, seasoning (the amount of salt), humidity, and texture.

While a variety of sensors are available, factors such as the type of food, its size, and its composition require the use of devices that are easy to clean and are least dependent on the food's consistency and contact area. "Therefore, we went for a conductance sensor that can give us enough consistency when all of these conditions vary," Sochacki says. He notes that it's easy to clean a conductance sensor due to its sturdy construction and adaptability to various food types.

Research teams worldwide are now turning to signal processing to help replicate, augment, or interpret human senses.

Signal processing is used at the point of sensing to measure the current and compute the conductance between the electrodes. "A single tasting with our

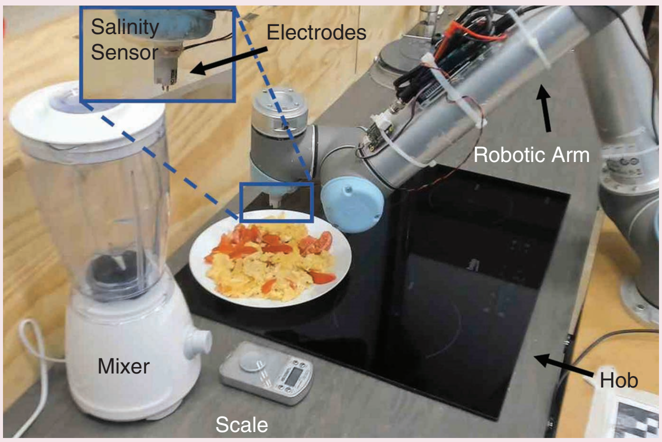

FIGURE 1. A UR5 robot is fitted with a conductance sensor for saltiness tasting. An induction hob is used for cooking. Food is presented for tasting on a ceramic plate. The entire installation is controlled by a program running on a laptop computer. (Source: University of Cambridge; used with permission.)

Digital Object Identifier 10.1109/MSP.2022.3202327
Date of current version: 29 December 2022

robot consists of 400 samples spread across the dish on a square grid and then blended," Sochacki explains. The samples are typically presented as an image. "These images usually differ significantly from each other, with the blended ones being much more homogeneous and with many fewer outliers," he notes.

Sochacki says that all of the information collected from the individual tastings needs to be condensed into a single number—something that can create a goal value. "Therefore, we usually compute some statistics of the computed values, like mean, median, variance. or entropy," he says. "A bunch of these metrics are then used to describe what a good dish is, and the robot tries to adjust the recipe parameters to produce such a dish."

Various methods, other than computing statistics, can be used to extract features. "For example, convolutional neural networks work well on images," Sochacki says. Unfortunately, most of these methods require a significant amount training data, which are expensive to produce when working with food. "This is because producing every sample [requires] cooking another meal, and, also, these meals need to be cooked very precisely to make the dataset accurate," he explains.

Sochacki believes that the technology will eventually be used to create robotic chefs possessing built-in common sense. Such a chef would, presumably, understand the outcome of its work better and, therefore, need much less training and supervision than a human.

Although the current system is simply a proof of concept, the researchers believe that, by imitating the human processes of chewing and tasting, robots will eventually be able to produce food that humans will enjoy and could be adjusted to match individual tastes.

"We acknowledge that the overhead required to implement a robotic chef in a restaurant is quite big," Sochacki says. "Therefore, we aim for places that can make the most of them." The team is hoping to target food preparation sites that operate 24/7 or in high-labor-cost areas. "Perhaps airport restaurants could be a good place for the first implementation," he notes. Dark kitchens—facilities with no humans present—could also be a viable application area, Sochacki adds.

Sochacki states that finding a theory behind taste is the most interesting and important part of his team's work. "We believe that formalizing a theory of robotic taste can lead to its proliferation in the future."

Determining sentiment

Making artificial intelligence (AI) "emotionally intelligent" promises to open the door to more natural human–machine interactions. To achieve this goal, it will be necessary to determine a user's sentiment during a human–machine dialogue, says Shogo Okada, an associate professor in the School of Information Science of the Japan Advanced Institute of Science and Technology (JAIST).

Speech and language recognition technology is becoming increasingly commonplace, as Amazon's Alexa, Apple's Siri, and similar technologies are incorporated into a growing number of stationary and mobile devices. A significant milestone in the advancement of highly accurate AI dialogue systems will be the addition of emotional intelligence to speech and language recognition. A system that's capable of recognizing users' emotional states would create a deeper empathetic response, leading to a fuller, more immersive user experience, Okada says.

A system that's capable of recognizing users' emotional states would create a deeper empathetic response, leading to a fuller, more immersive user experience, Okada says.

Multimodal sentiment analysis is a collection of methods that can automatically analyze a person's psychological state from their speech, inflection, posture, and facial expression, all of which are essential cues for human-centered AI systems, Okada notes. The technique could potentially lead to an emotionally intelligent type of AI that can understand its user's sentiment and generate an appropriate response.

To accomplish this feat, it's necessary to detect the user's sentiment during a human–machine dialogue. Physiological signals could provide a direct route to such sentiments, Okada says. Existing emotion estimation methods concentrate entirely on observable information. What's not detected is critical information embedded in unobservable signals, particularly physiological signals.

Working with Prof. Kazunori Komatani of the Institute of Scientific and Industrial Research at Osaka University, Okada has added physiological signals to multimodal sentiment analysis.

An individual's internal emotional state is not always accurately reflected within dialogue content, but it's difficult for people to intentionally manage their biological signals, such as breathing or heart rate. Okada believes, however, that it could be useful to use such involuntary signals for estimating a subject's emotional state.

Project participants communicated with a virtual agent presented on a display. The vocal utterances of each participant were recorded with a Microsoft Kinect V2 sensor. Participants' facial expressions were recorded with a video camera, while motion data were acquired via the Kinect sensor.

Electrodermal activity (EDA) data were collected in the form of physiological signals during the dialogues using a physiological sensor—an Empatica E4 wristband. "Since the E4 device is wireless and worn like a wristwatch, it causes neither disturbance nor discomfort, which is a top priority for naturalistic dialogue," Okada says.

The EDA measured skin conductance (SC), which reflects sweat gland activity through the sympathetic nervous system. The SC level was calculated using polynomial fitting, and the galvanic skin response was detected using PeakUtils, a utilities software package.

The researchers analyzed 2,468 exchanges from 26 volunteers to estimate the level of enjoyment experienced by the user during the conversation. Each individual was asked to assess how enjoyable or unenjoyable the conversation was. The team used a multimodal dialogue dataset, “Hazumi1911,” that uniquely combined speech recognition, voice color sensors, facial expression, and posture detection with skin potential (Figure 2).

The research suggests that the discovery of physiological signals in humans, generally hidden from external view, could open the door to a new generation of emotionally intelligent AI-based dialog systems, allowing for more natural and relevant human–machine interactions.

“The aim of our study was to clarify the effects of physiological signals in multimodal sentiment analysis by comparing other modalities, such as text and audiovisual signals,” Okada says. “Therefore, we decided on . . . signal processing, which is often used in affective computing.”

The network is then able to create new texture models following an evolutionary approach that relies only on user input on how a generated model feels in comparison to the real texture.

Feeling touched

Although often overlooked, tactile sensation is a critical sense—one that allows humans to perceive reality. Haptic devices can produce extremely specific vibrations that mimic touch, yet people are very particular about whether or not something feels quite “right.” Unfortunately, virtual textures don’t always hit the mark.

Researchers at the University of Southern California (USC) Viterbi School of Engineering have developed a more accurate haptic texture detection method. The framework takes advantage of humans’ ability to distinguish between certain texture details, using this natural attribute as a tool to bring virtual counterparts closer to accurate sensations.

Heather Culbertson, an assistant professor of computer science at the USC Viterbi School of Engineering, says she began developing realistic haptic texture models in 2010. Her early work involved creating data-driven texture models, recording the vibrations and forces that a user feels when dragging a pen across a textured surface. Those models required a costly data recording installation to document data for each new texture. “The benefit of this new preference-driven work is that we can create models for new textures without needing to record data,” she says.

The current technology uses a machine learning network, a generative adversarial network, that’s trained by a set of data-driven models developed for 100 different textures. “This training tells the network how a texture model should behave in terms of its features and response to the user’s motions,” Culbertson says. The network is then able to create new texture models following an evolutionary approach that relies only on user input on how a generated model feels in comparison to the real texture.

In the system’s current version (Figure 3), virtual material friction and hardness are created as forces via a force-feedback device. The textures are then displayed as vibrations through a voice-coil actuator that’s attached to the handle of the touch device. The actuator is controlled by the host computer’s audio output. “We have also created a

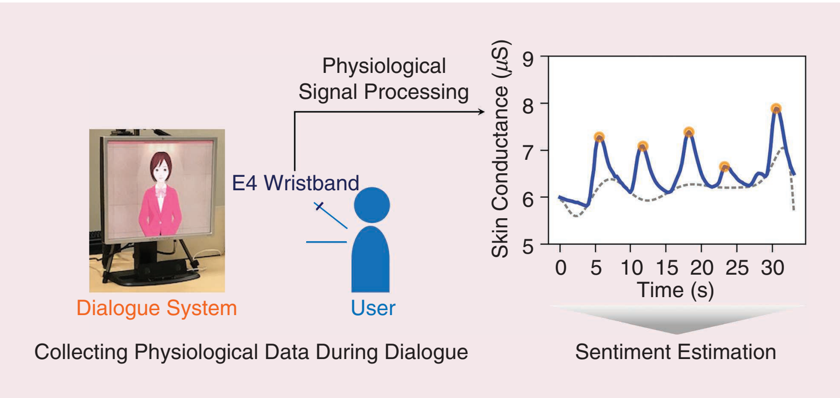

FIGURE 2. A multimodal neural network predicts user sentiment from text, audio, and visual data. (Source: Shogo Okada, JAIST; used with permission.)

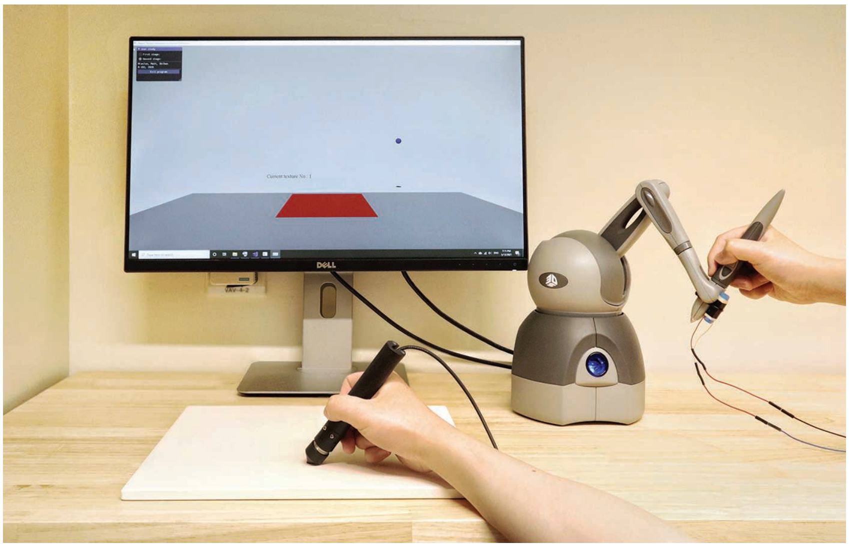

FIGURE 3. Interfaces for interacting with real and virtual textures. A haptic device (right) displays both force and vibration output. (Source: Shihan Lu, USC; used with permission.)

more simplified version of the virtual textures where we only play the vibrations through a voice-coil actuator attached to a stylus used in conjunction with a tablet," Culbertson explains.

The system is currently limited to modeling isotropic and homogenous materials that feel the same in all directions and in all areas. "But even given these limitations, we can still model a wide range of textures, including fabrics, metals, wood, plastics, stone, and carpet," Culbertson says. "The network was trained with 100 data-driven models, and we have used the system to train 10 additional textures."

The researchers record the tool's vibration in the form of three-axis acceleration signals—vibrotactile signals—and the user's interactive motions when dragging the tool across textured surfaces. The tool's vibration and user's normal force are recorded at 10 kHz by an accelerometer and a force/torque transducer, respectively. User position and orientation are recorded at 125 Hz by a magnetic position sensor, which is converted to the user's tangential speed and up-sampled to 10 kHz for matching the sampling rate of the vibration and force.

"For the signal processing, we apply a high-pass filter at 20 Hz to the recorded three-axis acceleration signals to eliminate the effects of gravity and human motions," explains doctoral student Shihan Lu. A low-pass filter at 1,000 Hz is applied to the accelerations to remove the effects of sensor resonance as well as frequencies that are imperceptible to humans. "Furthermore, the filtered three-axis acceleration signals are transformed into a single-axis acceleration signal using the DFT321 algorithm by preserving the perceptual fidelity," he notes. "This single-axis acceleration signal is the base of the texture modeling through the autoregressive process."

Looking ahead, Culbertson envisions online retailers creating haptic models of their products, allowing potential customers to virtually feel the texture of fabrics and other products. "Game designers could model different materials to add some realism to the virtual environments in their games," she adds.

Author

John Edwards (jedwards@johnedwardsmedia.com) is a technology writer based in Gilbert, AZ 85234 USA. Follow him on Twitter @TechJohnEdwards.