Avatar

github.com/2emoore4/avatar

Miles Steele
Massachusetts Institute of Technology
miles@milessteele.com

Evan Moore
New York University
2emoore4@gmail.com

ABSTRACT
The goal of this paper is to present a device that is designed to enhance direct communication between two people using ambient media. Our hope is that this device will allow for the transmission of various auditory, visual, and even physical cues without requiring users to explicitly convey this information. These cues will be presented using a lit fountain which will subtly output sound and light, while remaining in the background and not requiring a person's full attention.

INTRODUCTION
This project is meant to explore human to human communication. There are plenty of relevant forms of active communication, including text and speech. Text is typically used remotely in email or SMS, while speech can either be local or remote (in-person, phone, video call, etc). For remote conversations, there tends to be a lot of missing information between the multiple parties involved. With text, we are missing volume, pitch, and inflection of the other person's voice. Even speech is missing physical sensations and facial expressions. For decades, these forms of remote communication have been improving, one small feature at a time. Emoticons and profile pictures have been an attempt to add emotions to typed language. However, these are static and tend not to be interpreted very seriously.

An attempt to improve real time remote communication is video calling, which lets one see the face or body of someone else. This allows people to judge emotions and activities simply using vision. This form of communication requires someone's full attention. Not only do they have to keep an eye on their partner's video stream, but they are also tasked with presenting themselves to their own camera. Despite video calling's ability to simulate an in-person conversation, it feels like users are tied down.

Real in-person communication can easily happen in the background. Multiple people can be speaking to each other while working and not necessarily devoting their full attention to the conversation. Still, there is a lot of information to be gleaned from various sources, like ambient noise of your partner moving around or breathing, or facial expressions, or gestures. Lots of data can be transferred between two parties without necessarily moving the conversation into the foreground. Even text communication can happen asynchronously and in the background, although it does not convey nearly as much useful information.

With our project, we hope to investigate a method of performing active remote communication that happens in the background. During a conversation, people do not want to be tied down to a device, even if it is mobile; they want the ability to walk around or make gestures. Additionally, we want to be able to transmit more than just our voice. Presence, emotion, and level of activity are just a few examples that add a lot of relevant information to a conversation. Currently, including this data in a remote conversation requires announcing to your partner that you are busy, or making a certain facial expression in front of a camera. However, for in-person conversations, these details remain in the background of our awareness.

We are investigating the possibility of keeping this relevant data in the background, even during remote communication.

RELATED WORKS
Several projects from the Tangible Media Group at the MIT Media Laboratory have explored both active and passive communication, as well as presenting ambient media to individuals.

ambientROOM
This project intended to seamlessly present certain data to humans using ambient media. Many different ambient output types were explored that would allow users to process information without devoting their full attention to one specific source. These output types include video displays, light, audio, air flow, and tangible bits in the form of graspable bottles. Some discoveries revolve around the difference between background and foreground awareness. It is important to note that an ambient media source would be considered significant if it can seamlessly make this background/foreground transition and if it can still convey information while remaining in one's background awareness state.

ClearBoard
The goal of ClearBoard is to enhance direct person to person communication in situations where this is typically difficult. It succeeded in bridging the distance gap for remote collaboration. The hardware includes cameras and displays, which together create a direct video portal into another user's board. The transmission of data requires very few abstractions because it is intended to simulate the presence of two collaborators working at the same

whiteboard. Using techniques like gaze tracking to simulate eye contact, the technology fades into the background to allow user interactions to be highlighted.

OVERVIEW

We envision the final product as a bidirectional medium, but for this prototype we implemented only one instance of each of the listening and presenting ends. The listening end is a collection of sensors in a room and the presenting end is a fountain in another room which actively represents the state of the listening room. A person near the presenting end can be made aware of the state of the listening end room.

Implementation

Listening End

The listening end is ideally hidden from view of the user except for the capability to signal to it using gestures. The listening end draws information from three sensor inputs located in the listening room:

- Sound to communicate level of presence through vocal and other activity.
- Light to communicate whether the person is likely to be present at all.
- Gestures as a means for the users to actively communicate.

Sound is collected through the microphone of the computer on the listening end running a Javascript app. Light and gesture data are collected by an LED and infrared distance sensor attached to the Arduino which feeds values to a Python script on the listening computer. The Javascript app and Python script then stream the sensor values over HTTP Websocket connections to a computer on the presenting end.

A gesture signal is sent when the user waves their hand near the distance sensor.

Presenting End

The presenting end's visible facade is a fountain in a fishbowl capable of varying its water activity as well as displaying a colored light in its basin. The other parts of the presenting end, meant to be hidden from the user, include a tupperware container that houses the control electronics as well as a computer to receive sensor inputs and control output parameters.

Figure 1. The presenting end, with fishbowl & controller

The presenting end computer runs a Python server which receives sensor data over Websockets and translates the sensor input into an appropriate output state which it commands the Arduino to take on.

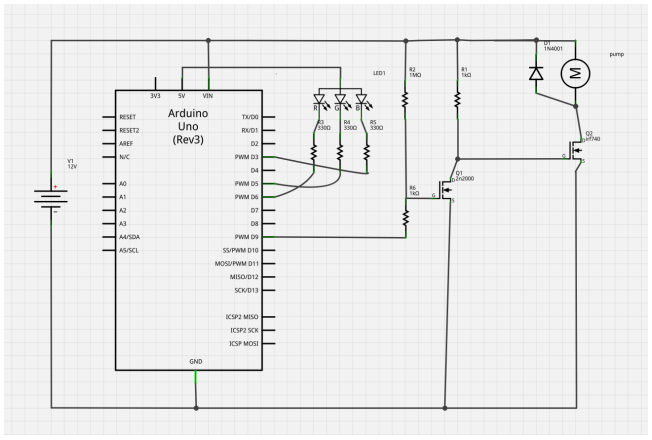

Figure 2. The controller for the fountain

Mapping

We experimented to a small degree with how to map sensor inputs to visible or audible outputs. This is an area which would benefit from further experiment. Our current scheme is to map listening volume proportionally and immediately to pump power. We do the same with light, mapping the received light to the output LED brightness. When the user on the listening end makes a gesture, the presenting end LED and pump turn to their maximum values for a half second in order to signal the intention to the person at the presenting end.

CONCLUSION

The goal of this project was to add useful cues and feedback to active remote conversations. Using the LEDs and water pump in the fishbowl, this was successful. The colored lights can easily indicate presence or time of day. Likewise, the sound and visuals of the fountain can simulate level of activity. This was effective in addressing the problem because this output does not take away a user's attention. Because of this, users are available to focus on their speech conversation or whatever they were previously

doing, and they don't have to spend any time conveying any extra bits of information.

Future Work

With only two outputs in our device, it is difficult to discern what the fountain is trying to say. For example, a yellow light could mean that the user is busy, but his partner will have no idea how to decipher the light color without reading a manual. It is very clear that the water flow rate is tied to audio, but still the lights are ambiguous. Since you can't see your partner, it would be hard to learn what different signals mean. The issue of mapping input sensors to output is very common, and has been explored with TMG's ambientROOM. With such a wealth of input data, it is difficult to present it in an easy to learn fashion that a human can not only read, but also interpret into useful information. Future research in this area will involve finding ways to transmit (in real time) machine readable sensor data to a human who can only process a few sources at a time. Additionally, this would also happen in the background of this user's awareness.

REFERENCES

1. Perlin, K. and Ishii, H. Shared Tangible Augmented Reality (MIT, Spring 2014)
2. Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B., Yarin, P., (1998). ambientROOM: Integrating Ambient Media with Architectural Space. *Conference Summary of CHI '98* (April 1998).
3. Ishii, H., Kobayashi, M., (1992). ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. *Conference Summary of CHI '92* (May 1992).