Giving Doctors and Patients a Voice - facilitating clinical communication and understanding through augmented reality

LEA BEYLOUNE, University of California, San Diego, USA
RYAN CHAMBERS, University of California, San Diego, USA
UTKRISHT RAJKUMAR, University of California, San Diego, USA
STEVEN RICK, University of California, San Diego, USA
ANISH SINHA, University of California, San Diego, USA

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; *Collaborative and social computing*; Interaction design; Accessibility;

Additional Key Words and Phrases: Computer-supported communication, augmented reality

ACM Reference Format:
Lea Beyloune, Ryan Chambers, Utkrisht Rajkumar, Steven Rick, and Anish Sinha. 2018. Giving Doctors and Patients a Voice - facilitating clinical communication and understanding through augmented reality. *ACM Trans. Comput.-Hum. Interact.* 0, 0, Article 0 (December 2018), 11 pages.

1 INTRODUCTION

Healthcare today is more technology driven than ever. From the widespread adoption of Electronic Medical Records over paper charts thanks to Meaningful Use, to today's increasing usage of machine learning and artificial intelligence to solve problems around increasingly large amount of complex medical data, patients and their providers are surrounded by technology. This has brought about a wealth of benefits. Machines can triage patients when staff are busy, making sure those who need to be seen sooner are prioritized. Deep neural networks can review imaging results, highlighting the cases where a radiologist should look closer, speeding up screenings. And doctors have more information at their fingertips than ever before, with the ability to personalize the medical care they provide for each patient based on their individual history, genetics, and symptoms. Technology has brought about a new revolution in the way healthcare happens. But with increased information, personalization, and automation, this specialized knowledge can sometimes escape the understanding of the patient, and care providers may not have the time or expertise to re-articulate everything. For maximal compliance, it is important that patients not only understand their care, but also trust it, knowing that everything they convey to their care team is understood and considered.

With this observation, our work seeks to answer the following question: "How can technology be leveraged to improve patient-provider communication and understanding?" We hypothesize that Augmented Reality devices,

Authors' addresses: Lea Beyloune, lea.beyloune@ipsa.fr, University of California, San Diego, USA; Ryan Chambers, rchamber@ucsd.edu, University of California, San Diego, USA; Utkrisht Rajkumar, urajkuma@eng.ucsd.edu, University of California, San Diego, USA; Steven Rick, srick@eng.ucsd.edu, University of California, San Diego, USA; Anish Sinha, a1sinha@ucsd.edu, University of California, San Diego, USA.

1073-0516/2018/12-ART0 $15.00
https://doi.org/

such as the Microsoft HoloLens, can support and encourage improved patient-provider communication. We further hypothesize that with appropriate design and implementation of this computer-supported communication, improved patient understanding will lead to increased patient trust and ultimately improved patient compliance with their respective care plan.

In this paper we first motivate our work, situating it within the existing literature and medical technology ecosystem. We then elaborate on the iterative and rapid design processes we employed in order to quickly create and assess multiple solutions to address the problem of patient-provider communication. Through the design thinking double diamond paradigm, we generated a variety of diverse low-fidelity solutions before selecting one to iteratively refine. Engaging medical professions in interviews and user testing generated insights which informed our final system design. Finally, we share the final design and implementation of our solution - AugmentedConnection - elaborating on what did and did not work and driving discussion forward on what future work may come next.

2 MOTIVATION AND BACKGROUND

Significant gaps in information exchange can occur between patient and provider for many reasons, and since 1986 it has been shown that medical terminology and jargon is largely at fault [2]. Wilson et al show that patients with limited English proficiency are even more severely impacted by the specialized vocabulary of physicians [3]. If a patient does not fully comprehend or correctly understand a diagnosis or treatment plan it can lead to issues ranging from lack of care plan compliance to improperly following of care instructions. Furthermore, if a patient does not understand what side-effects or complications they may need to watch out for, providers may not be kept in the loop until a patient destabilizes.

This, of course, assumes that the patient speaks the same language as their providers. In an increasingly global health economy, this increasingly not the case. Problems with communicating a patient's case can be further exacerbated due to language barriers. A patient might not speak the same language as their provider or more likely, the patient's native tongue is different than their doctors' and they communicate through broken phrases in a common language. Consequently, they might not be able to express their symptoms as completely as if they were communicating in their native tongue. As such, some solutions currently exist in the domain of language translation within the medical domain. There are applications and services which provide targeted translation services to clinics serving multilingual populations. CanopyApps provides an application for medical translation which also offers the ability to connect patients and providers to a live translator [1]. Mediglotte provides an application for medical term translation across many languages [2].

While applications like these attempt to resolve the gap in common understanding, they do not allow for a one-on-one conversation between patient and provider. Attention is diverted from the person who is being spoken to towards the technology which is mediating conversation. In some clinics if a patient does not speak the same language as their care team, they may ask for a health care interpreter. Such an individual would either work at the clinic in person or engage with the patient and providers via the phone or online. Unfortunately, interpreter services are not easily scaled and can be quite expensive, regardless of if in person or remote. Average costs can range as follows:

- $0.75 to $2.50 per minute for in-person interpreters
- $1.25 to $3.00 per minute for interpreters via telephone
- $1.95 to $3.49 per minute for remote interpreting via video conference call

Like when using an application, communication when using an interpreter/translator leaves communication between patient and provider indirect. Because discussion is routed through a third person much of the natural

[1] https://withcanopy.com/speak/
[2] http://mediglotte.e-monsite.com/

conversational flow is disrupted. While it is true that patients will occasionally bring an acquaintance with them who can translate, such as a relative or a friend, there is still the potential problem of the third person not properly understanding medical terminology or failing to communicate everything that a patient is experiencing or understanding. One such case occurred in Switzerland - a woman who was feeling ill came to the hospital with her husband to receive medical care. After discussion with the doctor, the husband informed the wife that she had the flu. However, treatment received by the patient was much more elaborate than a typical influenza case. It was later discovered that the patient had lymphoma, but was unaware due to a failure in translation and understanding by the patient's partner [1].

Given the latest advances in artificial intelligence related to automatic speech recognition and natural language processing [4], we believe a properly designed solution can help improve communication between a patient and their care team. We propose using a system that takes as input a user's speech, translates their words to text, and then facilitates medical language translation and provides further clarification of terms for shared interaction on wearable displays. Such a solution enables patient and doctor to maintain eye-contact and hold an engaged conversation while also being able to properly understand each other and share an externalized common ground.

3 DESIGN

When designing the user experience we had envisioned, it was important to keep the volume of user interface elements low as things in the medical world are already stressful, cumbersome, and overwhelming. We did not want to distract or confuse our users and wanted to make the application as intuitive and non-invasive as possible.

3.1 Interviews and Need finding

Interviews were conducted with two Navy Nurse Corps. nurses, ages fifty and fifty-eight, as well as a Skilled Nursing Facility admissions director, age twenty-one, with over eighty years of combined experience. Initial suggestions were most concentrated along the follow categories: language barriers, whether that be through native language differences or language deficits; agitated, confused, or ignorant patients; unresponsive patients; and lastly, confirmation between patients and health care providers that the proper service has been provided.

With these insights we began to ideate on possible solutions. Designs centered on an augmented reality display that could provide information like translations and definitions for patients and doctors, promoting the ability to maintain eye contact while also allowing for a wealth of resources to be brought right into the wearer's field of view.

3.2 Low Fidelity Prototypes

We began prototyping out Augmented Reality Keyword Simplification and Translation features via paper first. In combination with Wizard-of-Oz techniques, the team was able to role-playing and evaluate the experience from different perspectives: the Doctor, Patient, or the application.

When a physician user selected "Keyword Simplification" by pointed at the button, the "patient" would start talking about their symptoms, at which moment paper Thought Clouds would appear behind the patient with the keywords of what the patient has said, redefining them with simpler language, as shown in Fig 1.

If the user pointed at "Full Translation Mode" instead they would see everything that the patient had said, translated into a specified language.

3.3 PrototipAR Prototypes

After the paper prototypes were refined, we began to prototype the application in Augmented Reality. We used an application developed in UCSD's Weibel Lab, PrototipAR. This application allows the users to design

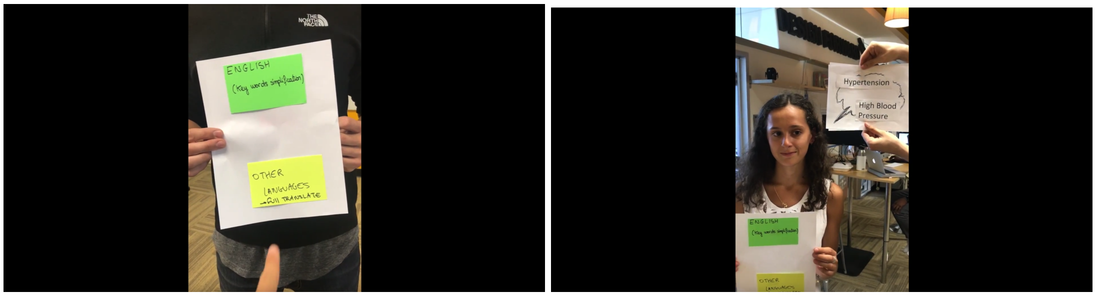

Fig. 1. On the left we have our paper prototype user interface, on the right we show how the application simulates a user's speech as a bubble next to their head.

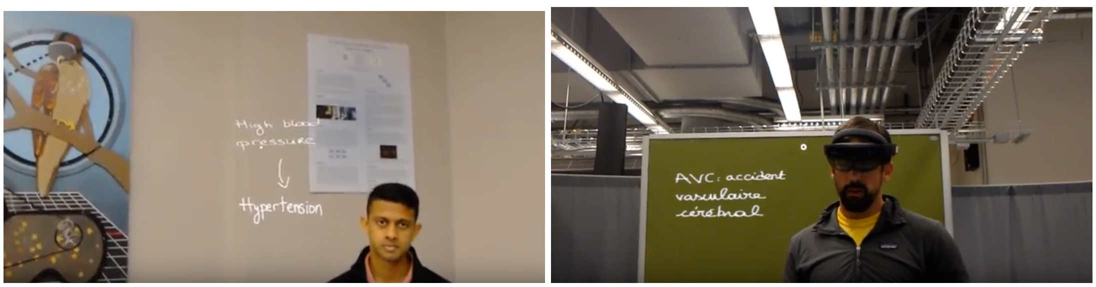

Fig. 2. Here we show what it looked like in PrototipAR to prototype speech as text holograms.

interface elements on an iPad and send those elements to a Microsoft HoloLens to simulate what a fully developed application might be like.

With PrototipAR we established a base user interface and tested the idea of live translations and keyword descriptions via usage of multiple frames within the environment. We found that it was rather non-intrusive to have text in the field of view, and thus continued with the idea of peripheral transnational aides, as shown in Fig 2. Further in development, we moved prototype design and evaluation into a mock operating room inside of The Design Lab at UCSD. This environment allowed for a much more realistic prototyping session, in which we realized it would be useful to have descriptions of some of the medical artifacts available, such as device or imaging result descriptions.

3.4 High fidelity User Interface Mock-ups

As testing advanced and prototyping continued a higher fidelity prototype was produced. Using Figma as a wire-framing tool we built out a "minimum viable product" based on our research, proposed solution, and evaluation so far. This model included all necessary features, specifically 1) a shared transcript viewable in the native languages of the user, 2) an emphasis on medical keywords derived from the conversation, 3) overlain descriptions of medical devices and scans, 4) the ability to call for a nurse, and 5) general descriptions of location, time, etc, all as show in Fig 3 and Fig 4

This model provided points-of-view for both the doctor and the patient. The style is most similar to a heads-up-display, in which text bubbles appear in the upper left, and the transcript and nurse button are available along the bottom of the field of view. Each emulated Microsoft HoloLens can support its own language, native to

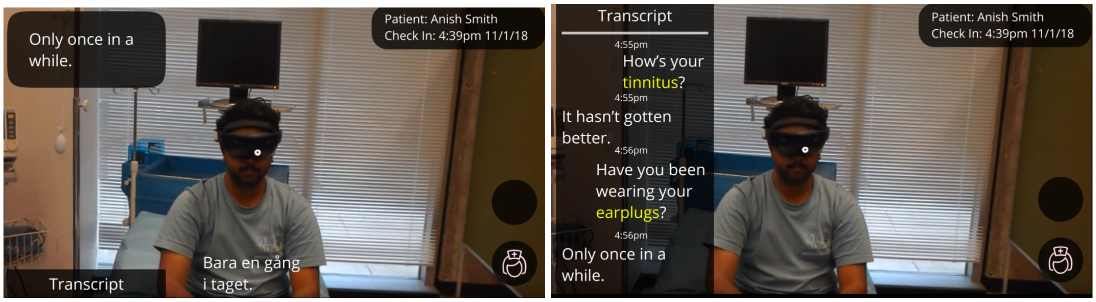

Fig. 3. A high fidelity prototype of the doctor's view. On the left is the general view and on the right is the shared board with translations and keywords.

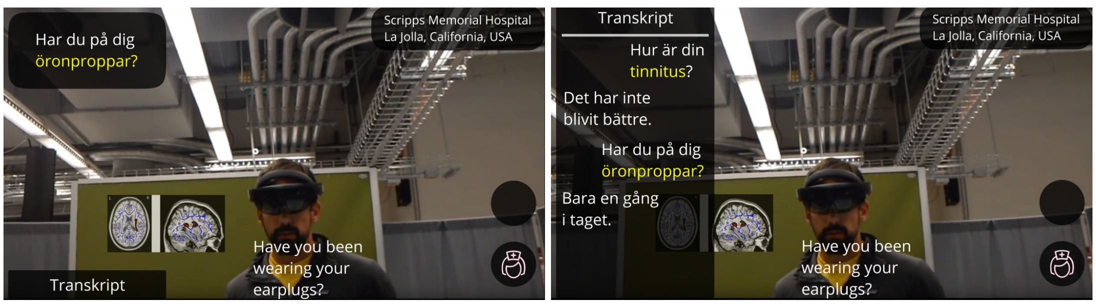

Fig. 4. A high fidelity prototype of the patient's view, showing two view types similarly to how we do so with the doctor view in Fig 3.

the user. This prototype was presented to user for testing, and was subsequently displayed alongside the final demonstration.

3.5 User Testing

With the application fully prototyped in PrototipAR and the High Fidelity user interfaces flowing smoothly, we began user testing our application. By evaluating both potential user perspectives we were able to adjust for differences between how a patient or provider might want to interact with such technology. We further got input regarding how speech bubbles might appear when a Patient and Doctor are conversing. If there is only 1 doctor and 1 patient, then it was found to be alright to have a subtitle feature that stays fixed in one screen position, regardless of where the user looks. However, if there were more than 1 patient or more than 1 doctor or nurse, then the text should be closer to their the speaker.

3.6 Final Design

Final designs drew from the results of our testing, simplifying the user interface on the HoloLens itself to a shared billboard experience. By giving both users a shared list of keywords extracted from the conversation all users are able to see things discussed such as symptoms, conditions, prescriptions, dosages, and diagnoses, among others, shown in Fig 5. This interface was found to be non-disruptive, allowing for natural communication between the doctor and patient, while allowing for a common ground to be established. The interface acts as a reference, and

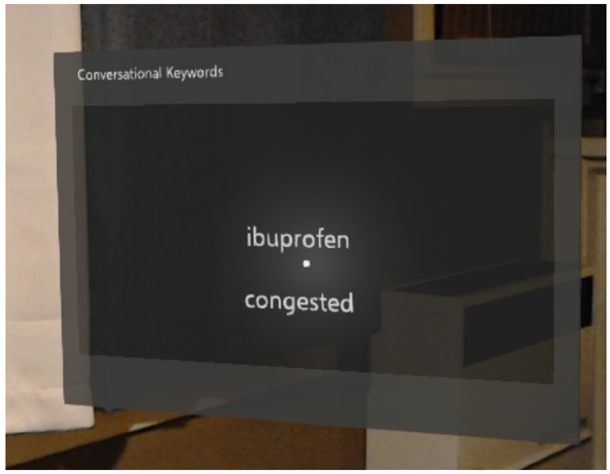

Fig. 5. Screenshot of Shared Keyword Detection in the final application design

subsequently a record of the visit, available to both the patient and the provider. This valuable information can prevent further misdiagnoses or failed treatment enforcement.

4 SYSTEM DEVELOPMENT

Our system utilizes a modular architecture (shown in Fig 6 in order to support distributed and iterative development, but also to accommodate the asynchronous nature of our application. By combining independently running processes with a UDP messaging framework the system is able to be run on a single machine, or distributed across as many devices as is necessary. Information is passed from module to module gracefully, so should one component fail the entire system does not cease to function. Generally, we stream audio of speech from our users to a central server. The central server performs speech-to-text conversion, language translation, and parsing of natural language to identify the medical terms used and then pushed that information to the patient and doctor's HoloLens.

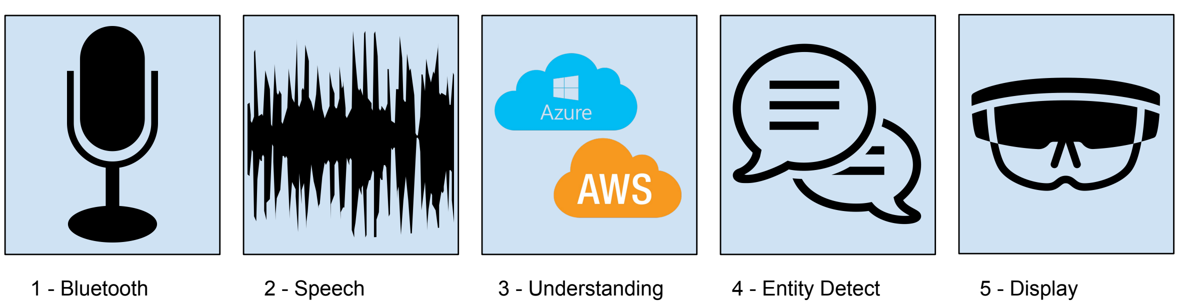

Fig. 6. The modules of our system: 1) Bluetooth audio capture, 2) speech recognition, 3) language understanding, 4) medical entity detection, and 5) information display.

4.1 Architecture

Our application is composed of 5 main modules: Bluetooth audio capture, speech to text, language understanding, medical entity detection, and display of shared information on HoloLens.

4.1.1 Module 1: Audio capture. We use two Bluetooth mics that each user independently wears. This enables us to more easily segment out the speaker when two individuals are standing near each other as compared to using the built-in microphone on the HoloLens. Audio is captured when the users speak and then sent to our central server, tracking where the audio came from and dynamically setting speech energy thresholds to differentiate between user speech and ambient noise.

4.1.2 Module 2: Speech to text. The central server runs a python script based on the SpeechRecognition [5] library to then send data to Microsoft's Azure Cognitive Services, returning a text data object of the recognized speech, shown in Fig 8.

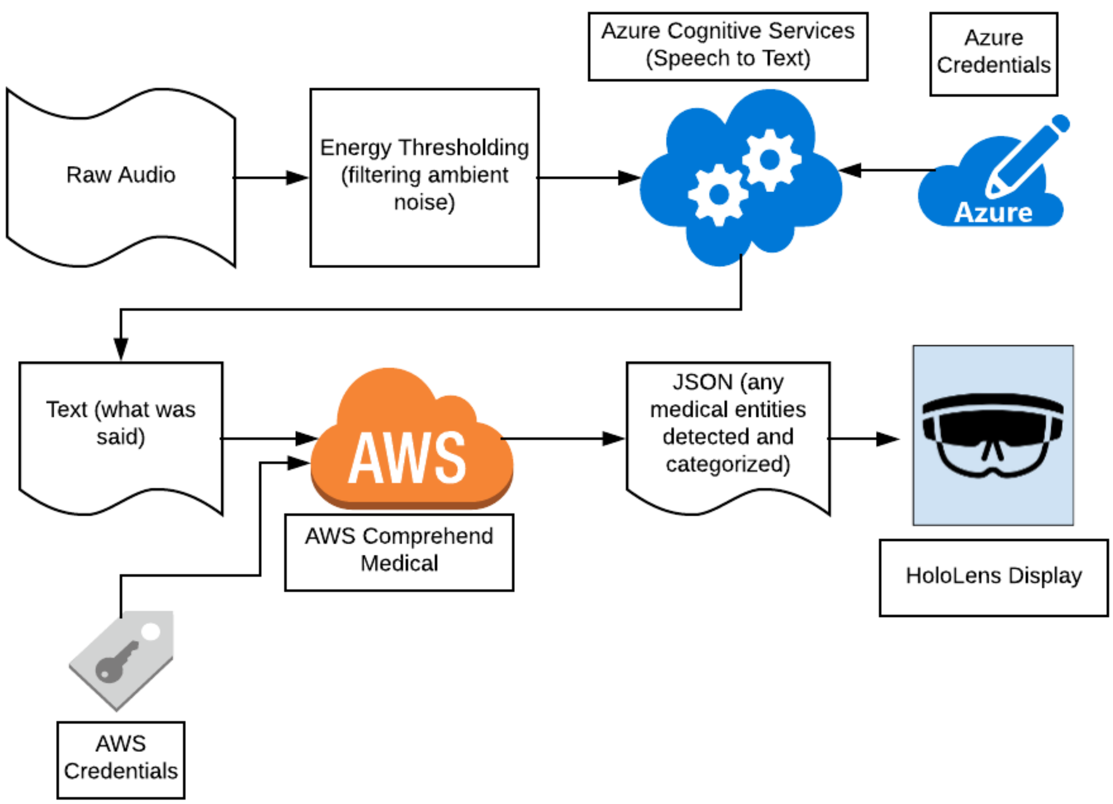

Fig. 7. Our general system data pipeline. Credentials for Azure and AWS are used as data moves from audio capture to text analysis and finally to JSON parsing and transmission to HoloLens. Every module shown passes data along and receives data via a UDP socket connection, allowing for asynchronous behavior that flows as naturally as human conversation.

4.1.3 Module 3: Language understanding. Using Azure Cognitive Services to analyze the text we generated from speech and then translate that text between languages. We take the patient's language and convert it to the doctor's language and vice-versa. We perform full language translation and defer displaying data until after keyword segmentation. This allows for the patient and doctor to speak as much as they choose in their own language before arriving at a common ground.

4.1.4 Module 4: Parsing of unstructured medical data. Using Amazon Comprehend Medical entity detection, we then pass the text which came back from speech to text and translation in order to identify medically salient terms. These come back from AWS with categories which allow us to defer or display content, ignoring Personal Health Information (PHI) but not ignoring medications, diagnoses, or symptoms. This is shown in Fig 9.

4.1.5 Module 4: Display of shared information. Finally the keywords which return from Amazon Comprehend are checked against those terms which have already been heard and displayed and all new terms are added to the shared billboard. The HoloLens of each user operates independently of their partner's device but display the same shared content. This was accomplished through use of the Microsoft Mixed Reality Toolkit [3]. By building a Unity [4] application that is spatially aware, we can anchor the shared billboard that the users see into a specific location in the world.

4.2 Technology Used

As discussed we predominantly used Python, but also heavily leveraged cloud computing resources such as Azure and Amazon Web Services (AWS). The system manages communication and the passing of information via UDP sockets and runs with each component operating asynchronously.

[3]https://github.com/Microsoft/MixedRealityToolkit-Unity
[4]https://unity3d.com/

```
Press ENTER to EXIT
A moment to adjust for ambient noise...
Ready!
You said: So tell me how you been feeling.
You said: I've had a bit of a fever and a bit of migraine all week.
```

Fig. 8. Speech to Text, accomplished using Azure Cognitive Services

```
I've had a bit of a fever and a bit of migraine all week.
Comprehend Medical
{'Id': 0, 'BeginOffset': 20, 'EndOffset': 25, 'Score': 0.9737816452980042, 'Text
': 'fever', 'Category': 'MEDICAL_CONDITION', 'Type': 'DX_NAME', 'Traits': [{'Nam
e': 'SYMPTOM', 'Score': 0.6168150305747986}]}
{'Id': 1, 'BeginOffset': 39, 'EndOffset': 47, 'Score': 0.9662578105926514, 'Text
': 'migraine', 'Category': 'MEDICAL_CONDITION', 'Type': 'DX_NAME', 'Traits': [{'
Name': 'DIAGNOSIS', 'Score': 0.6773641705513}]}
sending:
fever
migraine
```

Fig. 9. Medical Entity Detection, accomplished using AWS Comprehend Medical

5 TESTING AND EVALUATION

5.1 Testing and Evaluation

We conducted numerous levels of user testing to produce our final prototype. In our first iteration, we used the Wizard of Oz (WoZ) method to prototype our application. We then tested among our team by having members play the role of doctor or patient. We created a script to ease new users into application evaluation, giving them lines to follow along with before adding their own. Lastly, we tested our prototype with a medical in-take professional and two nurses in order to receive professional feedback on the viability and usability of our system. They suggested that having full translations might be a bit too overwhelming and patients more often speak at least broken English and only have trouble with certain words that they would rather say in their native tongue. We incorporated their feedback into our final design, which employed translations of only keywords with the option for full translation, and parsing and sharing medical keywords exchanged between both doctor and patient.

In conclusion, user testing allowed for developer assumptions to surface, helping us to be more informative and thorough in our prototyping methods as well as in the final design. It also offered a fresh perspective on the scope and viability of our project, often times simplifying the development and use of our product. Lastly, it gave insight into how clunky the technology can be, while also being super helpful when there are no errors. In our testing we mainly employed WoZ and body storming testing methods.

6 COLLABORATION

This project was undertaken by five students, three undergraduate students from CSE 118: Anish, Lea, and Ryan, and two graduate students from CSE 218: Utkrisht and Steven. We held three weekly stand-up meetings on Monday, Wednesday, and Friday mornings to discuss current progress, next steps, and resolve blocks, if any. We were aided by a TA, Scott Lim, who aided with any software and hardware blocks we had. The other TAs also collaborated with us to help us during our design iterations.

6.1 Structure of team

We divided our 5 person team into 3 subgroups. Group A consisted of Lea and Utkrisht, group B was Steven and Anish, and lastly, Group C was Ryan, who worked independently but coordinated with both teams. Group A worked on syncing the doctor and patient's HoloLens so that both the doctor and patient can visualize the same key points in the discussion. Group B focused on speech-to-text and the parsing of medical terms from the text. Group C worked on the story board and the design of the app itself.

As new problems arose in the later development of the product, all members of the team helped with various other aspects of the project than what was originally assigned to each person. This flexibility in our team allowed us to easily overcome blocks in the project.

6.2 Overall collaboration across 118-218

We made sure to pair one undergrad with one grad student, with the exception of Ryan who worked with all members as he was in charge of user interaction and design. The collaboration between CSE 118 and 218 went very smoothly since both classes knew what was expected of the other class through the class website, piazza, and our stand-up meetings. We also had mandatory discussion sections where both CSE 118 and CSE 218 met to have face-to-face time work on the project. This ensured that the whole team met in person at least once a week to have time to work together. Aside from these discussion sections, we also dedicated time during the week for the subgroups to meet and finish weekly milestones and deliverables. All members were receptive to each others

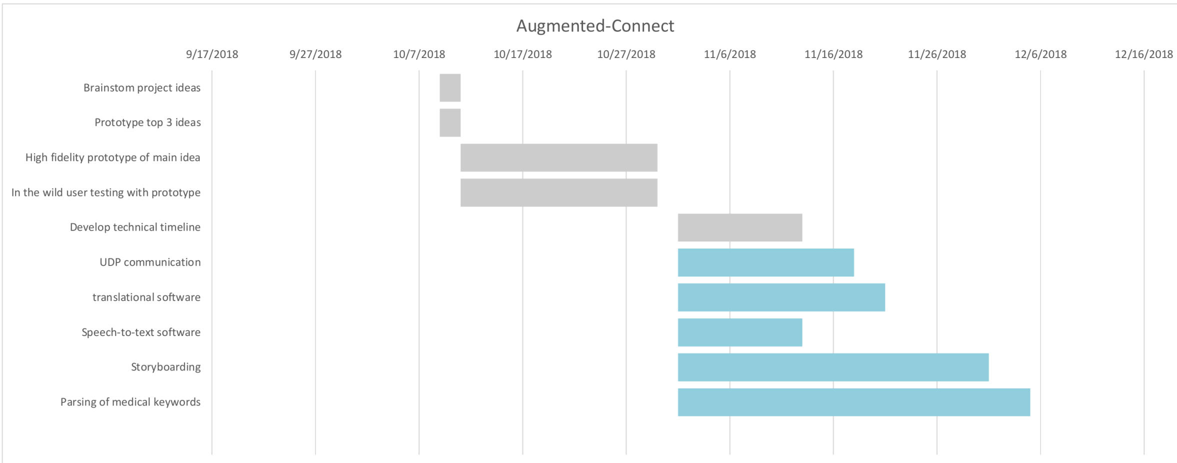

Fig. 10. The project's Gantt chart

needs and communicated accordingly. We tracked our project in GitHub [5] and followed a timeline outlined by our project's Gantt chart shown in Fig 10.

6.3 Problems / Issues and how they were resolved

Throughout the quarter, we had multiple issues that we iteratively resolved. Our first issue was having all hands on deck during our standup meetings. We fixed this problem by talking to individual members to attend the meetings more diligently, but also being cognizant of emergencies during which certain members could not attend.

We had difficulty setting up meeting times for all members to meet and work on the project. However, this problem was quickly resolved when the class structure allowed for group meetings during the Thursday discussions as well as free-work time during Tuesday lectures. We also shifted our focus from having entire group meetings to having smaller subgroup meetings.

Another major problem was members being unresponsive to blocks they were encountering and requesting help when needed. This led to a great reduction in speed of development, where at times it seemed the project had almost come to a stand still. The problem persisted throughout the quarter and we tried to resolve this by first following up with members multiple times. When this proved to not be very effective, the task was re-assigned to other members so the project can move forward.

7 FUTURE WORK

In the future, we would like to add many features and modify many existing ones. First and foremost, we would like to use light-weight glasses that only have display capabilities as opposed to the heavy computation power of the HoloLens. Since our final design has both HoloLens communicating with a central computer that is doing all the heavy computation, the HoloLens themselves only act as interactive displays. Hence, a small pair of glasses that carries just this feature would heavily reduce the physical burden of wearing the HoloLens.

Currently, we are able to parse medical keywords, but we would like the patient to be able to see further explanation of the parsed keywords. Just the parsed keywords might still be overwhelming or confusing if

[5] https://github.com

what they mean is still a mystery. For this reason, we will explore possibly being able to click on or gaze at the parsed words and having mini-search results appear. If a user stares long enough, the application will give more information about that keyword. This would be similar to hovering on computers, or even Force Touch on the Mac. If people do not understand something, they tend to stare at it for a long time, trying to make sense of it. Hence, this implementation for providing more info would provide a more intuitive user experience as well.

Third, we would also like to improve upon our user interface since currently, the shared bulletin space is not moveable by the user. Having the functionality would greatly improve user experience. Doctors' offices can be a very busy setting with bulletin boards already existing, as well as charts and graphs and patient information being posted all over the room. By allowing the user to move the board, be it doctor or patient, this would allow for better organization of space and less mental load and stress on either user.

Lastly, we would like to be able to overlay information on the medical environment. For instance, if the patient sees a heart monitor, it would be helpful for them to know how they are doing by overlaying information on the monitor that indicates whether their heart rate is at a proper pace. Of course, this has certain concerns as the doctor might wish for the patient to not be made aware of certain details. The exact addition of this feature will require much more user testing with medical professionals.

8 CONCLUSION

We aimed to develop a product that will improve communication between a patient and their care team by facilitating the understanding of technical medical language and the translation of words spoken in the non-common language. Our system takes as input a userâĂŹs speech, translates their words to text, and then facilitates medical language translation and provides further clarification of terms for shared interaction on wearable displays. We currently employ the HoloLens to display this information so that the patient and doctor can maintain eye-contact and hold an engaged conversation while also being able to properly understand each other and share an externalized common ground.

ACKNOWLEDGMENTS

We would like to thank our professor Nadir Weibel and our teacher assistants Danilo Gasques, Janet Johnson, Scott Lim, and Tommy Sharkey for their time, invaluable feedback, and technical help in the development of our project. We would also like to thank our peers and nurses who helped with the testing of our prototype throughout the course of this project. All of their input really truly did change our design and direction we took our project. Their honest feedback is what allowed us to have a better project after every iteration.

REFERENCES

[1] La Croix. 2010. Le poids des mots pour les patients venus d'ailleurs. https://www.la-croix.com/Ethique/Sciences-Ethique/Sciences/Le-poids-des-mots-pour-les-patients-venus-d-ailleurs-_NG_-2010-09-13-578498

[2] Richard D Gibbs. 1986. Patient understanding of commonly used medical vocabulary. (1986).

[3] Elisabeth Wilson, Alice Hm Chen, Kevin Grumbach, Frances Wang, and Alicia Fernandez. 2005. Effects of limited English proficiency and physician language on health care comprehension. *Journal of general internal medicine* 20, 9 (2005), 800–806.

[4] Wayne Xiong, Lingfeng Wu, Fil Alleva, Jasha Droppo, Xuedong Huang, and Andreas Stolcke. 2018. The Microsoft 2017 conversational speech recognition system. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5934–5938.

[5] Anthony Zhang. 2017. SpeechRecognition. https://github.com/Uberi/speech_recognition

Received ; revised ; accepted