# The Good, the Bad, and the Facts: Multimodal Representation of Medical Conversations for Patient Understanding

by
Jaclyn Berry
B.A. Architecture - University of California, Berkeley (2014)

Submitted to the Department of Architecture and
the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science in Architecture Studies
and
Master of Science in Electrical Engineering and Computer Science
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019



Author . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Department of Architecture and
Department of Electrical Engineering and Computer Science
May 23, 2019

Certified by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Terry Knight
Professor of Design and Computation
Thesis Supervisor

Certified by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Randall Davis
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Nasser Rabbat
Aga Khan Professor
Chair of the Department Committee on Graduate Students

Accepted by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

**Thesis Committee**

Terry Knight
Professor of Design and Computation
Thesis Advisor

Randall Davis
Professor of Electrical Engineering and Computer Science
Thesis Advisor

# The Good, the Bad, and the Facts: Multimodal Representation of Medical Conversations for Patient Understanding

by Jaclyn Berry

Submitted to the Department of Architecture and
the Department of Electrical Engineering and Computer Science
on May 23, 2019 in partial fulfillment of the
requirements for the degrees of
Master of Science in Architecture Studies
and
Master of Science in Electrical Engineering and Computer Science

## Abstract

Medical patients face significant challenges for managing their health information. In particular, cancer patients have a uniquely difficult experience where they must endure the physical and emotional effects of their illness while simultaneously navigating overwhelming amounts of medical information. In this thesis, I focus on the challenge of capturing, reviewing and extracting information from medical appointments for patients enduring serious health conditions such as cancer. First, I propose a novel multimodal interface to help patients review and understand information they received from conversations with their doctors. This interface captures medical conversations as text and audio, with important positive and negative information highlighted. I conducted 25 user studies where I enacted fictional conversations between a doctor and a patient to evaluate whether this method of representing information would help patients review and understand their appointments. Results from the user studies show that the web interface serves as a useful tool for reviewing the content of the conversations, however its effect on patient understanding cannot yet be determined. Second, I propose a machine learning algorithm to automatically classify the positive and negative information in medical conversations based on analysis of the text and prosody in speech. The model with the highest performance on my dataset achieved an accuracy of 90.6% and F1-score of 0.888. While I focus on challenges within the medical field, findings from this thesis may be relevant to emotional conversations in any setting such as sportscasting, political debates and more.

Thesis Supervisor: Terry Knight
Title: Professor of Design and Computation

Thesis Supervisor: Randall Davis
Title: Professor of Electrical Engineering and Computer Science

I dedicate this thesis to my dad.

David B. Berry

April 29, 1947 – May 4, 2018

Until we meet again.

# Acknowledgments

Thank you, Terry Knight and Randy Davis, for encouraging me to pursue this thesis and finish my degrees at MIT. Your guidance and support have been invaluable throughout my time here.

Thank you to my friends on Team Armadillo for helping me complete the spring semester in 2018 when my father was in ICU. And thank you to everyone at MIT and back home in California who supported me and my family after my father's passing.

Thank you, Shokofeh, for our walks and jogs along the Charles. Thank you, Anesta, for bringing me ibuprofen, going out for treats and sharing avocado jokes.

Thank you, Kevin, for supporting me through this long-distance endeavor. I am looking forward to spending the rest of our lives together.

Thank you, mom. Your optimism and perseverance inspire me every day. You have been a pillar of stability for my entire life and I know I have your support no matter what I pursue. I could not have done this without you.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

On May 4, 2018, my father died due to a rare complication from a bone marrow transplant for an extremely uncommon form of leukemia. The time from his diagnosis in September 2017 to his passing eight months later were wrought with uncertainty, desperation and hope. Before the bone marrow transplant, I remember my father pouring over binders of information about the side effects, complications and quality of life during and after the procedure. This treatment was undoubtedly risky, with only 60% success rate among all bone marrow transplant patients and several years before full recovery.

Three weeks into the bone marrow transplant, the doctors and nurses were all very positive. They said my father was fairing remarkably well and sent him home a week early. Ten days later, my father was intubated in the ICU with multiple organ failure. The team of doctors expressed to us the severity of the situation but they maintained optimism that they could still save my father's life. With each day that passed, their optimism faded and it became clear my father would not survive.

Since my father's passing, I have been trying to create sense from the chaos of cancer, starting from diagnosis, through treatment, and death. I have listened to my mother agonize over whether she missed something the doctors said or failed to tell the doctors enough. I looked through the spiral-bound notebooks my parents used during visits to the clinic: only short phrases and a few keywords were recorded. We had no records from my father's final weeks. I recognized that managing medical information during cancer treatment is an incredible challenge. And even more specifically, I recognized that keeping track of everything a doctor says is an enormously difficult task with the greatest burden placed on patients and their caregivers.

My father's experience with cancer and the healthcare system was not uncommon. Patients in medical environments face a host of challenges ranging from extended stays in inpatient care, long periods in the waiting rooms, confusion about treatment options and misunderstandings of insurance coverage, to name a few. Researchers and industry professionals across multiple disciplines have conducted ethnographic studies to formally identify unmet needs of patients in hospitals and medical clinics. Among their findings are the recurring challenges of passive exchange of information from doctors to patients (i.e. doctors speak and patients listen), patients' low information retention, unmanageable amounts of information per appointment, and exam rooms ill-equipped for patients to interact with their health data [1]–[3]. Traditionally these topics may not have been considered real problems because patients were expected to unequivocally trust their doctor's instructions. However, as medical information becomes increasingly available online and personal health monitoring technology becomes more accessible, people are taking a more active role in maintaining their own health [4].

Along with the general challenges of medical environments and managing health information, cancer patients have a uniquely difficult medical experience. Not only is diagnosis emotional for patients and their loved ones, treatments are often a harrowing test of mental and physical endurance. Oncologists may prescribe combinations of surgeries, chemotherapy, radiation, immunotherapy, or medication, to name a few [5]. The effects of these illnesses and their treatments cannot be adequately described with words. Patients endure a range of symptoms from the cancer itself as well as side effects to their prescribed therapies. Such experiences may include, but are not limited to, pain, nausea, fatigue, or diminished mental state [6]. In addition, patients may experience changes in their physical appearance, may be unable to work, and may be required to relocate or travel long distances to receive their treatment.

Cancer patients are also burdened with the responsibility of managing their personal health information. After diagnosis, patients must immediately begin coordinating treatment and appointments with multiple physicians. During clinic appointments, cancer patients receive verbal information about their diagnosis and instructions for treatments from doctors and nurses. They must process this information, record comprehensive notes, ask questions, and make life-altering decisions within these meetings. Outside of the clinic, patients must be fastidious about their medication regimen which can include upwards of ten types of medications with various timing instructions and dietary restrictions. They must review and organize their decontextualized notes and information pamphlets from their oncologists. Patients and their caregivers often find the sheer quantity of information from this process overwhelming.

Despite the tens of thousands of health apps available online, very few address the specific needs of cancer patients. Many apps are targeted for preventative care such as fitness trackers, diet logs and medication schedulers. Only recently have applications specific to patients with medical conditions begun to emerge [7]. Within the past decade, researchers have conducted ethnographic studies to identify the unique conditions and needs of cancer patients at all points of their treatment journey. Cancer and its associated medications introduce significant hardship to patients ability to capture and retrieve information during treatment [8], [9]. Such challenges include diminished attention due to stress or treatment side effects, inability to accurately capture different types of information, and uncomfortable physical accommodations in the clinic environment. Based on these studies, researchers and medical information enterprises have begun developing new systems for information management, social support, improved patient-oncologist communication, and data visualization.

In this thesis, I address the challenge medical patients face for managing their health information. In particular, I focus on the challenge of capturing, reviewing and extracting information from medical appointments for patients enduring serious and often emotionally demanding health conditions, such as cancer. First, I hypothesize that a multimodal interface presenting information from medical appointments through text, audio and labels identifying positive and negative information will help patients review and understand information from conversations with their doctors. Second, I hypothesize that important positive and negative information can be extracted from a conversation with machine learning algorithms using features from the textual content and prosody of speech. In this context, I define positive information as information that should cause a patient to feel optimistic about their treatment options, health outcome, diagnosis, or health resources. I define negative information as information that should cause a patient to feel pessimistic about their treatment options, health outcome, diagnosis, or health resources. And I define neutral information as information that should not cause a patient to feel either pessimistic or optimistic about their treatment options, diagnosis, health outcome, or health resources.

To evaluate my hypothesis, I developed (1) a web interface that represents a conversation through a text transcript and an audio recording with labeled positive and negative information, and (2) a machine learning algorithm using features from text and prosody to classify important positive and negative information from a medical conversation. I conducted 25 user studies to collect fictional medical conversation data to be used in the machine learning algorithm and evaluate the effectiveness of my web interface for facilitating information review and understanding. In these conversations, I assumed the role of the doctor and participants assumed the role of the patient. Results from the user studies show that this multimodal representation of information using audio and text facilitates review of medical conversations. More specifically, the positive and negative labels of the text influence users' perception and encourage reflection about the information. However, the effect of the web interface on participants' understanding cannot be determined from this study. Results from the machine learning algorithms show that, with a dataset containing speech from a single speaker, positive and negative information can be identified from text and prosody with an accuracy of 90.6% and an F1-score of 0.888.

While I focus on conversations between doctors and patients, I propose that findings from this thesis may be relevant to emotional dialogs in general. Instances of emotional dialogs could include political speeches, debates, sportscasting, psychology or theater. In these examples, a tool identifying information based on textual and prosodic analysis of speech could be useful for external observers to navigate and understand important information from these events. Alternatively, more personal instances of emotional conversations may include couples' counseling, important presentations, or art and design critiques where a person may be too overcome with emotion to hear and understand the other side of the conversation. In these scenarios, having a multimodal record that shows what and how information was communicated could help a person review the conversation from a new perspective and improve understanding.

# Chapter 2

# Related Work

## 2.1 Medical Information Management

Patient engagement with their health information has a positive influence on their experience in the clinical environment and overall health outcome [10], [11]. This finding has encouraged researchers and industry experts to rethink how patients and doctors could interact with health information. Advancements in telehealth such as remote-patient-monitoring and secure electronic data transfer have made healthcare services more available, particularly in rural and underserved regions [12]–[14]. Mobile applications and wearable sensors for monitoring health metrics are empowering individuals to become more engaged with their health, contributing to positive health status [15], [16]. Interactive visualizations of medical data within medical environments offer the potential for predictive models and clearer communication between medical professionals and patients [17]–[19].

Cancer patients face a particularly intense challenge for managing and engaging with health-related information. New all-in-one applications for organizing health information are emerging in response to these challenges. For example, Klasnja et al. developed a customizable personal health information management system for breast cancer patients to record and link health logs, calendar events, and external information [20]. Jacobs et al. deployed tablet devices to aid breast cancer patients with organizing and remembering their health information [21]. Other researchers have incorporated interfaces for community support and online forums to address the emotional burden of diagnosis and treatment for cancer patients [22], [23]. However, none of these studies address the problem of capturing, reviewing and understanding information shared in conversations between doctors and patients.

## 2.2 Information Capture and Retrieval

Methods for information capture and retrieval often involve note-taking and note-review. Traditional note-taking is a situational task demanding varying levels of accuracy, attention and technological intervention [24]. Regardless of the setting, in the best of circumstances, note-taking is difficult [25]. While taking notes in a lecture, an office meeting or a doctor's office, a person must take in a continuous stream of audio and visual information, understand which elements are most important, then record that information so they will remember what it means later. For a doctor, the demanding nature of manual note-taking means that doctors must spend a significant amount of cognitive effort and time writing down notes instead of interacting directly with their patients. For patients, note-taking requires them to be mentally, physically and emotionally prepared to discuss their health information. This cognitive burden for medical patients poses a risk for misremembering or omitting important information given by the doctor or other medical staff.

Hundreds of note-taking programs are available online, each one supporting different platforms, levels of complexity and input [26]. These applications work well for many general purposes such as taking notes in a meeting or writing down grocery lists. But these systems still do not reduce the cognitive load on users for determining what information is important, nor do they include significant features supporting note-review [27]. In an attempt to make information capture and retrieval more effective, several studies are exploring new methods for interacting with information across multiple media sources. Researchers have implemented visual and voice interfaces for interacting with audio and other non-text-based media [28], [29]. Other studies have investigated how combinations of text-based media and short video-clips can facilitate more efficient video review and content retrieval [30]–[32].

## 2.3 Language Processing

Imagine if you never had to take your own notes and instead all the notes were written down for you. Since the 1960s, the solution to this idea was a designated human transcriber or transcriptionist. This person would have the designated function of transcribing a complete textual record of activities like court sessions, cinematic productions, doctors' notes, or academic lectures. Recently, human transcription has been shown to be very successful within the medical field, particularly for aiding doctors with note-taking during appointments with their patients [33], [34].

Advances in machine learning have enabled significant development in automatic speech recognition (ASR), speaker recognition (SRE), and natural language processing (NLP). Commercial ASR products offer transcription services with accuracies near 90% [35], [36]. However, there are many remaining challenges associated with automatic speech recognition including context-specific vocabulary and linguistic ambiguities. Researchers are investigating multimodal strategies for combining textual speech data with video to help machines disambiguate the meaning of sentences within a given visual context [37]. Along with semantics, speaker identification is an ongoing challenge and current technology in ASR and SRE primarily address a single person dictating or a simple phone conversation. Researchers in ASR and SRE have created new tools to differentiate several speakers in group conversations using triangulation of sound and voice identification vectors [38], [39].

Automatic speech recognition shows promising applications in the medical field. Academic researchers and industry experts have begun developing new speech recognition models specifically for medical conversations between doctors and patients [40], [41]. While the examples here focus on the benefit that ASR and SRE may provide to medical professionals, such a system could also be beneficial to medical patients who must record and revisit information from their appointments. Patients may find that accurate records of these conversations serve as important resources when trying to understand their diagnosis and make treatment decisions.

## 2.4 Language Understanding

Transcribing a conversation is part of the solution, but the problem of extracting meaning from a conversation still remains. There are several ways to interpret the meaning of a conversation, and certainly human-human communication is interpreted through multimodal channels. In this section, I focus on the meaning of the textual content of the conversation. Work related to a computer's understanding of text is an important topic in machine learning and natural language processing. Researchers have constructed new systems that demonstrate understanding of language in text by generating relevant answers to factual queries [42] and relating textual statements to visual descriptors [37]. Still more are investigating applications of artificial intelligence for story interpretation and evaluation of author intention [43].

Language understanding and information extraction are becoming increasingly relevant for healthcare applications. For over 20 years, natural language processing and machine learning systems have been developed to extract important events from clinical notes and construct the timeline at which they occur in a patient's healthcare experience [44]–[46]. Recently applications of this work have extended to interpreting clinical notes to facilitate clinical decision-making for cancer care, palliative care and psychology [47]–[49]. While these studies explore the applications of language understanding for healthcare professionals, in this thesis I developed a tool for language understanding for patients.

## 2.5 Affect and Paralinguistics

Nonverbal communication plays a central role in how we understand and interact with other people. In fact, emotional expression is so important to human communication that we often impose affective characteristics onto non-emotive objects: consider times when you have gotten angry at your computer or when you have become emotionally attached to a toy. Studies in affective computing assert that integrating emotional intelligence into computers will create better interactions between people and machines [50]. Such interactions apply to applications including emotion coaches for those on the autism spectrum and adaptive educational environments [51], [52].

Paralanguage is the study of extralinguistic vocal cues that inform human communication. These cues include tone of voice, grunts, sighs, pauses or exclamations that inform human traits and states such as gender, age, mood or emotion [53], [54]. Prosody, a subset of paralanguage, studies extralinguistic qualities of speech with specific regard to tone, pitch, accent and rhythm. Because how we say things is often indicative of our emotional state, researchers have developed machine learning models based on the prosodic elements of speech for emotion recognition [55], [56]. Others have taken a multimodal approach to the challenge and developed models based on combinations of textual and prosodic analysis of speech [57]. However, paralanguage informs us about more than just emotional state. Medical professionals and researchers have manually and computationally developed methods for identifying mental disorders and physical illness based on qualities of speech [54], [58], [59]. And other researchers have utilized multimodal systems of face-tracking and acoustic analysis for determining participant interest and boredom in unstructured conversations [60].

# Chapter 3

# Representing a Conversation

## 3.1 Design of a Prototype Multimodal Interface

The first task in this thesis was to determine whether a multimodal interface containing text and audio records of a conversation, with annotations identifying positive and negative information, would help a patient review and understand the information discussed in their medical appointments. For this purpose, I constructed a simple prototype web interface with HTML, CSS, and JavaScript, hosted locally on my personal computer with node.js. The primary features of the interface include a text transcript of a recorded conversation, an audio playback interface for the recorded conversation, a chart to visualize the total number of positive and negative speech events the doctor contributed to the conversation, and filters to isolate positive and negative information in the transcript (Figure 3-1). This design enables users to navigate through the conversation through multiple modalities: they can choose to only read the text, only listen to the audio, or some combination of the two. Users can click on a speech event in the transcript and the audio cursor will update to the corresponding time in the recording. Conversely, users can click on a time in the audio playback interface, and the transcript will scroll to the corresponding position.

Obtaining transcripts of conversations was a multi-step process (Figure 3-2). An audio file of a conversation was converted to MP3 using Adobe Media Encoder CS6. The MP3 file was uploaded to Amazon Web Service (AWS) S3, the storage service for

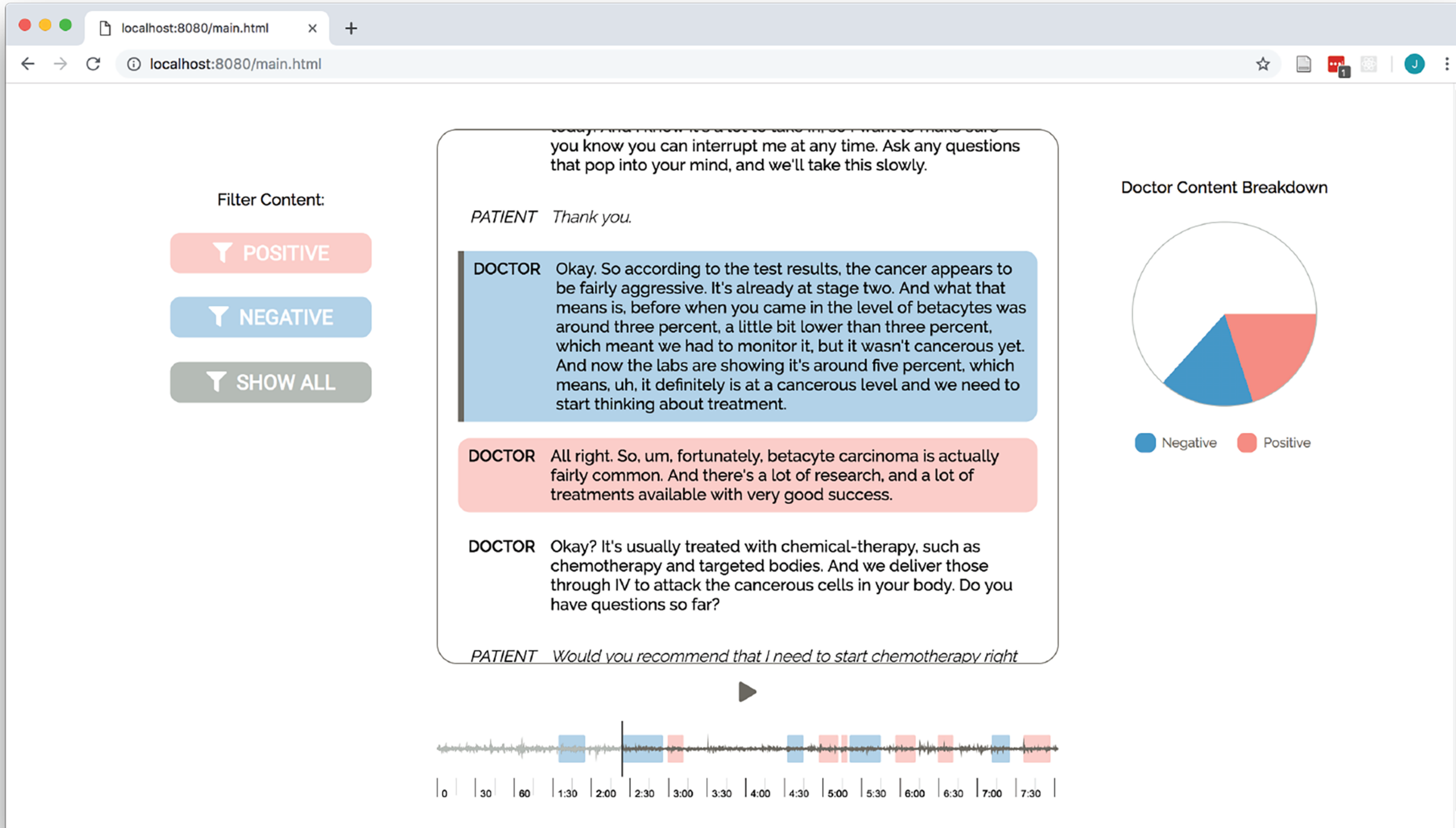


Figure 3-1: The prototype web interface.

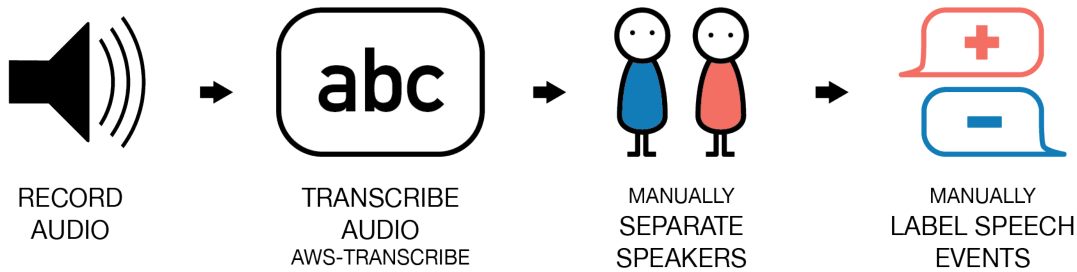


Figure 3-2: The pipeline for obtaining transcripts of conversations.

AWS machine learning tools, and then submitted to AWS Transcribe, a transcription service. AWS Transcribe returned a full transcript of the audio file, timestamps associated with each recognized word and estimated speaker segmentation. However, due to the inaccuracies of the transcription, I manually separated by speaker and corrected transcription errors for each conversation. In this first task, I also manually labeled the speech events as positive, negative or neutral.

## 3.2 User Study Procedure

With the working prototype, I designed a controlled study to evaluate how this representation of information may assist patients in reviewing and understanding information from their medical appointments. The study protocol was approved by the Committee on the Use of Humans as Experimental Subjects (COUHES) at Massachusetts Institute of Technology. The study consisted of two phases: (1) Appointment Phase, and (2) Review Phase. Each session of the study took place in a private room reserved in the Architecture and EECS CSAIL facilities on MIT campus.

Before beginning the study, participants were informed of the full procedure and provided their consent. They were informed that the study was investigating methods of information capture within the medical setting. The true intent of the study was withheld to prevent any bias for or against the proposed representation of information. Participants were also informed that the study required them to participate in a fictional cancer diagnosis. Due to the emotional nature of this type of conversation, I felt participants needed to be fully aware of their role in the study.

For the Appointment Phase of the study, participants and I (the study proctor) enacted a fictional appointment between a patient and a doctor in a private room. The participant assumed the role of patient and I assumed the role of doctor. Participants were provided with a single fact-sheet of backstory information for their role including the nature of their make-believe health-condition, the nature of the appointment and a pseudonym to use during the conversation (Appendix A). During the conversation they were allowed to use this sheet of paper to take notes. As a fictional doctor, I used a predefined script to deliver a fictional cancer diagnosis to the participant using the SPIKES protocol (Table 3.1). This protocol is a method used by medical professionals to deliver bad news in an empathic and humane manner [61]. In the script, I informed the acting patient that they had been diagnosed with a fictional cancer called, “betacyte carcinoma” (Appendix B). I included relatively positive information such as high 5-year survival rates with treatment, several available treatment options, and successful research supporting the disease. I also included negative information such as the cancer diagnosis, uncertain long-term prognosis and severe side effects from the treatment. The acting patient was allowed to ask questions at any time during the conversation. Conversations lasted anywhere from 5 minutes to 15 minutes, depending on the user's engagement and responses. I recorded each conversation using the Voice Memos app on an iPhone 5S.

After completing the conversation in the Appointment Phase, I collected participants' notes and participants departed for four to five hours. This break was included to simulate the time a real patient may experience between receiving a real diagnosis and returning home to discuss the information with their family. During the gap between the appointment phase and review phase of the study, I prepared the audio recording for the web interface. As described earlier, this process involved transcribing the audio to text and manually annotating speech events within the conversation as positive, negative or neutral.

After the designated break, participants reconvened with me in a private room for the Review Phase of the study. In this phase, participants completed a questionnaire about their experience in the earlier appointment first by referencing only

**SPIKES Protocol**

| | |
|---|---|
| **Setting** | Arrange to speak to the patient in a private room. Make eye contact and offer gestures of reassurance. |
| **Perception** | Ask the patient about what they are expecting from the appointment and what information they already know. |
| **Invitation** | Ask the patient how much information they would like to know about their diagnosis. |
| **Knowledge** | Deliver information about the diagnosis in small chunks using nontechnical language. |
| **Empathy** | Assess patient's emotional reaction. Offer comfort and/or ask patient how they are feeling. Let the patient know you are connected with how they feel. |
| **Strategy & Summary** | Ask the patient if they are ready to hear about treatment plans for the future. Explore the patient's knowledge and expectations of treatment. Create a dialog where patients can express their fears and concerns. |

Table 3.1: SPIKES protocol for delivering bad news.

their handwritten notes and then by using the web interface. Before using the web interface, participants were required to watch a 1-minute video introducing the basic functionality of the interface. The survey was designed to evaluate changes in participants' understanding and perception of information in the acted conversation as well as their experience using the web interface (Appendix C). The survey consisted of 31 questions: five binary questions (yes or no), one trinary question (positive, negative or neutral), 12 short answer questions and 13 rubric questions rated on a Likert scale of 0 (least) to 4 (most). The survey was divided into three sections. In the first section participants could only reference their manual notes to answer questions about the acted conversation. In the second section, participants could reference the web interface and their manual notes to answer questions about the conversation. In the third section, participants answered questions about their experience using the web interface.

After completing the questionnaire, I conducted short interviews with participants about their experience in the acted scenario and their reactions to the web interface. An outline of the interview questions can be found in Appendix D. I asked participants about their trust in the classification of information and their reactions when the system disagreed. I asked them to reflect on challenges they faced and their sense of control over their health information in their role as patient. I also asked participants when they thought audio would be useful for managing their health information. Finally, I asked participants to share how they would want a system like the one I designed to be integrated into their healthcare experience. At the end, participants were rewarded with a $20 Amazon gift card.

## 3.3 User Study Results

In total, 25 participants between ages 18 to 50 years old agreed to participate in the study. There were 17 female participants and 8 male participants, all from the MIT community. The average age was 25 years old, the youngest was 19 years old and the oldest was 44 years old. The group included 13 graduate students, 7 undergraduate students and 5 members of MIT staff. Participants came from several departments and programs including MIT Media Arts and Sciences, Electrical Engineering and Computer Science, Architecture, Aeronautics and Astronautics, Mechanical Engineering, Math, Physics and Urban Studies and Planning.

I was first interested to determine whether the representation of information in the web interface affected how participants felt they understood the conversation. I asked participants to respond to the question, "How well do you feel you understand the content of the conversation?" using a Likert scale of 0 (Not at All) to 4 (Very Well). Participants ranked their understanding first using only their manual notes and again using the web interface to review the conversation. Unexpectedly, the average reported level of understanding was identical for the initial condition using only manual notes and for the final condition using the web interface, with a score of 3.28 out of 4. However, there were changes in understanding at an individual level.

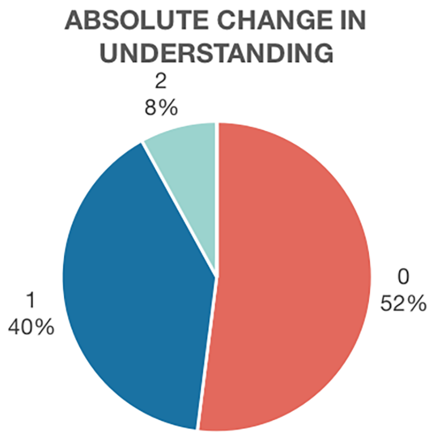


Figure 3-3: Magnitude of change in participants' understanding after using the web interface.

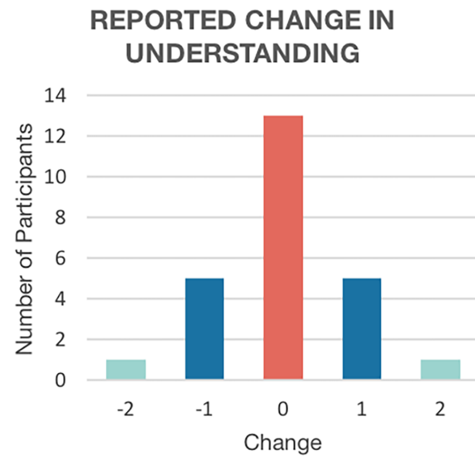


Figure 3-4: Change in participants' level of understanding of the conversation after using the web interface.

I computed the difference between participants' reported levels of understanding with their manual notes compared to their understanding with the web interface to identify changes per individual. From this comparison, 13 out of 25 participants reported no change in their understanding of the conversation. The remaining 12 out of 25 participants reported an absolute change of up to 2 on a scale of 0 (no change) to 4 (maximum change), as shown in Figure 3-3. Of the participants who indicated a change in understanding, 6 showed a positive change in understanding and 6 showed a negative change in understanding (Figure 3-4). The positive change in understanding indicates that participants gained additional comprehension of the conversation after viewing the web interface.

Among the reports of negative change in understanding, three may be interpreted to mean that the participants realized how little they originally understood the conversation after using the web interface. In an interview, one participant who reported a negative change in understanding expressed that she would have wanted to do additional research because she realized:

> *"It's common knowledge that there is chemo and these effects, but there is nothing new I left [the appointment] knowing, just the name of the cancer."—Guadalupe Babio, Graduate Student*

Indication that a change in understanding did actually occur for these three participants is further supported by reported changes in their perception, confidence and optimism about the conversation. However, the remaining three negative changes in understanding do not exhibit consistency with the other responses in the survey nor comments in the interviews.

Based on the positive responses regarding usability and comments in the interviews, the results describing changes in participant understanding were not due to confusion using the interface, but from ambiguity in the survey questions. In hindsight, these questions should have been phrased more clearly. The question "How well do you feel you understand the content of the conversation?" after viewing the web interface may have had mixed interpretations. Participants may have responded with their current level of understanding or they may have reassessed their original level of understanding. A clearer set of questions may have been, "Now that you have used the web interface, how well do you think you understood the conversation originally?" and "How well do you understand the conversation now?" This change would account for cases when a participant overestimates their understanding of a conversation initially, then after reviewing the web interface, discovers they actually did not understand the conversation very well from the beginning. However, based on the current responses, the effect of the web interface on participant understanding is inconclusive. Additional user studies and methods for evaluating comprehension would be required to determine a significant relationship between this representation of a conversation and changes in patient understanding of information.

Delving further into participants' understanding of the conversation, I was interested to see whether the interface influenced participants' perspective of the positive and negative information in the conversation. I calculated the difference in participants' reports of the valence of the conversation using only their manual notes compared with their reports using the web interface. The results showed that 16 out of 25 participants experienced an absolute change up to 2 on a scale of 0 (no change) to 4 (maximum change) and 9 participants experienced no change in their perception of the information (Figure 3-5). Of the participants who did report a change in their

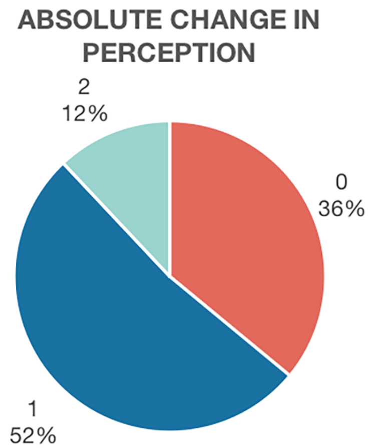


Figure 3-5: Magnitude of change in participants' perception of the conversation after using the web interface.

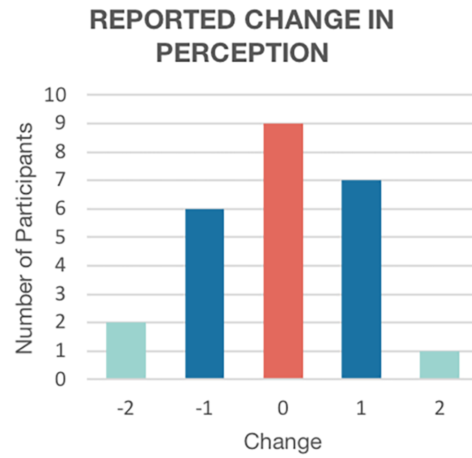


Figure 3-6: Change in participants' perception of the conversation after using the web interface.

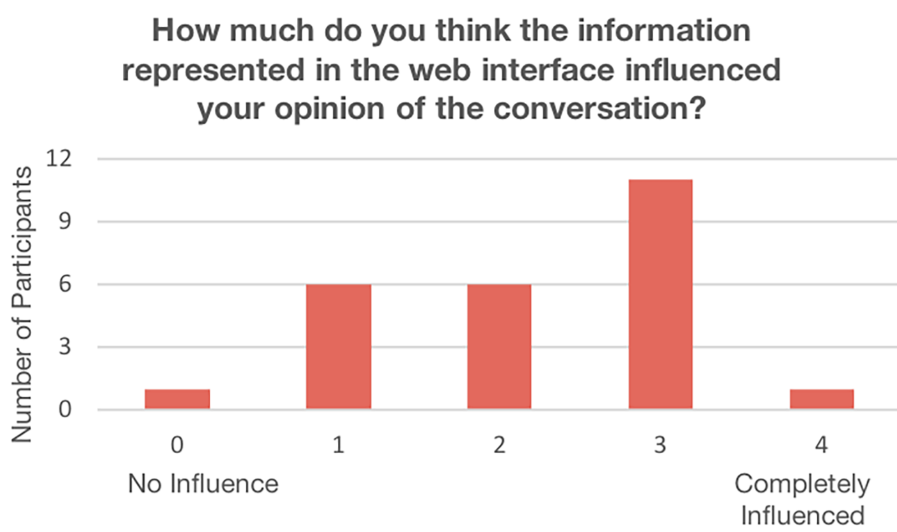


Figure 3-7: The web interface's influence on participants' opinion of the conversation.

perceived valence of the information, 8 experienced a positive change and 8 experienced a negative change (Figure 3-6). A change in the positive direction may have arisen when a participant exited the appointment feeling the overall diagnosis was very negative, but after viewing the web interface they realized there was positive information to consider. One participant who reported a positive change said:

> *"I was also surprised by how much positive was in [the appointment]. So, it kind of also made me think, oh maybe it wasn't as bad as I thought it had been."—Anastasia Ostrokowski, Design Researcher*

In the opposite direction, a participant may originally have left the conversation feeling that the conversation went very well and that the prognosis was promising, but upon seeing the interface they realized the conversation was not as positive as they originally perceived. A participant who experienced a negative change said:

> *"During the conversation, I thought we had more positive information. But when I saw the interface, I realized that there was so much less positive information, like there was only one part. So, I was kind of amazed." —Graduate Student*

In support of these findings, when asked how much the web interface influenced their opinion of the conversation, Figure 3-7 shows that 18 out of 25 participants also reported a score of 2 or higher on a scale of 0 (no influence) to 4 (completely influenced). In this respect, the web interface did influence participants' perception of the positive and negative information shared in the conversation.

The survey included questions regarding participants' feelings about their treatment options. Participants were asked to respond to the question, "Based on the conversation, how would you rate your optimism about your treatment options?" on a Likert scale from 0 (Not At All Optimistic) to 4 (Very Optimistic) using only their manual notes as reference. Later they responded to the question, "After viewing the web interface, how do you feel about the treatment options discussed in the conversation?" on a Likert scale from 0 (Very Pessimistic) to 4 (Very Optimistic). Results from the survey showed that 11 participants reported an absolute change of up to 2

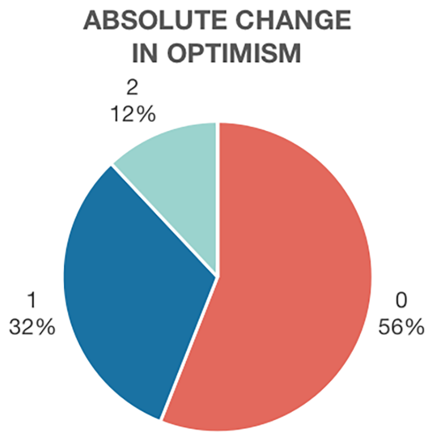


Figure 3-8: Magnitude of change in participants' optimism about treatment after using the web interface.

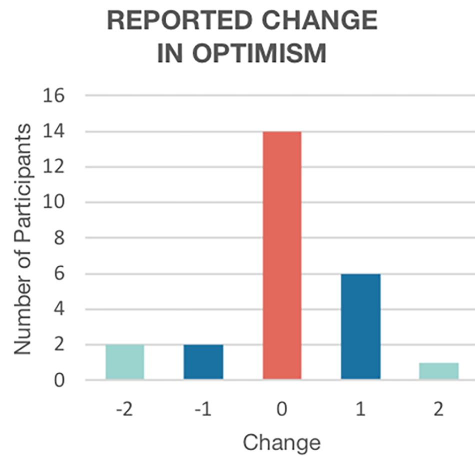


Figure 3-9: Change in participants' optimism about their treatment after using the web interface.

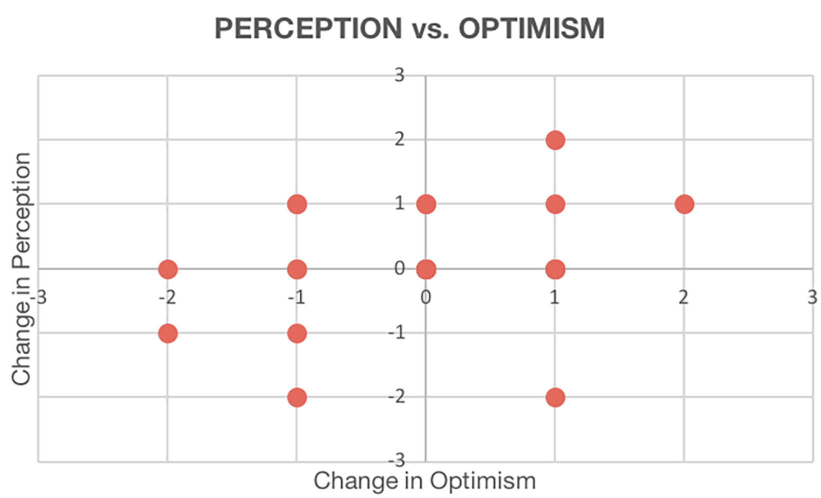


Figure 3-10: No clear correlation was found between participants' change in perception and their change in optimism.

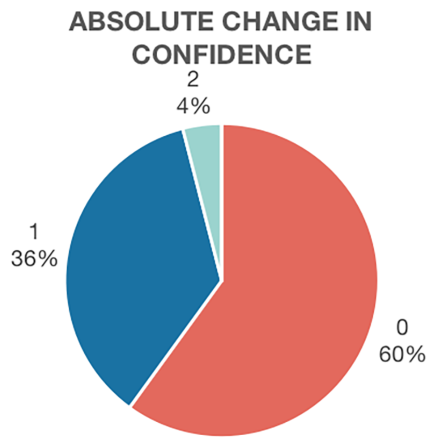


Figure 3-11: Magnitude of change in participants' confidence about making a decision for their treatment after using the web interface.

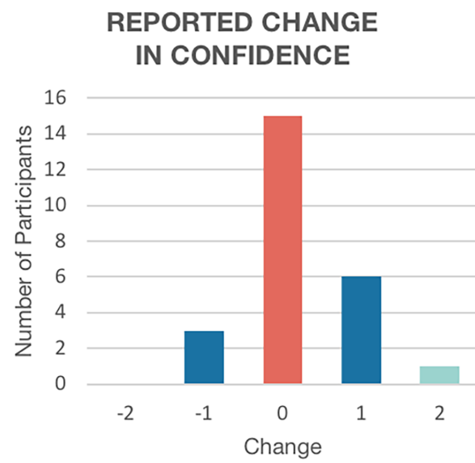


Figure 3-12: Change in participants' confidence about making a decision for their treatment after using the web interface.

(on a scale of 0 to 4) and 14 participants reported no change after viewing the web interface (Figure 3-8). Of the 11 participants who reported a change, seven participants became more optimistic and four became less optimistic after viewing the web interface (Figure 3-9). Surprisingly, based on the survey responses, there does not appear to be a strong correlation between participants' change in perception of the information and their reported change in optimism. Of the four participants who became less optimistic, three also reported a negative change in perception of information. However, of the seven who became more optimistic, two reported no change and two reported a negative change in perception of the information. It is possible that the imprecise wording on the Likert scales did not have equivalent meaning to all users and therefore resulted in ambiguous results.

Additionally, I was curious to find whether the web interface impacted participants' confidence about making a decision for their treatment. First, I asked participants to respond to the question, "Based on this conversation, how confident would you feel about making a decision about your treatment?" on a Likert scale of 0 (Not At All Confident) to 4 (Very Confident) using only their manual notes. Then I asked participants, "After viewing the web interface, how confident do you feel about making a decision about your treatment?" with the same Likert scale. Based on the difference between the responses, 10 participants indicated an absolute change of 1 or more (on a scale of 0 to 4) and 15 participants indicated no change in confidence after viewing the web interface (Figure 3-11). Of the 10 participants who indicated a change, 7 felt more confident about making a decision after viewing the web interface and 3 felt less confident (Figure 3-12).

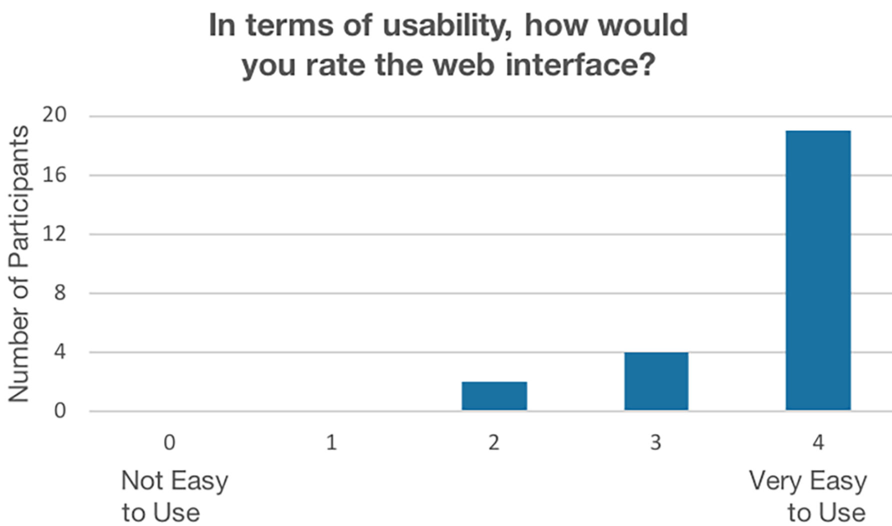


Figure 3-13: Usability of the web interface.

Results regarding the usability and experience with the web interface were very positive. All participants reported a score of 2 or higher on a scale of 0 (Not Easy to Use) to 4 (Very Easy to Use) when asked about the usability of the interface (Figure 3-13). More specifically, 19 out of 25 participants rated the usability as "Very Easy to Use." When asked whether the web interface was helpful for finding important information, 23 out of 25 participants indicated a score of 3 or higher on a scale of 0 (Not At All Helpful) to 4 (Very Helpful), as illustrated in Figure 3-14. Going further, 20 out of 25 participants also indicated that the positive and negative labels were helpful for understanding the content of the conversation (Figure 3-15).

Participants provided feedback about features that they found successful, features they found unsuccessful and improvements they would have liked to experience. The most successful features in the web interface were the highlighted positive and negative

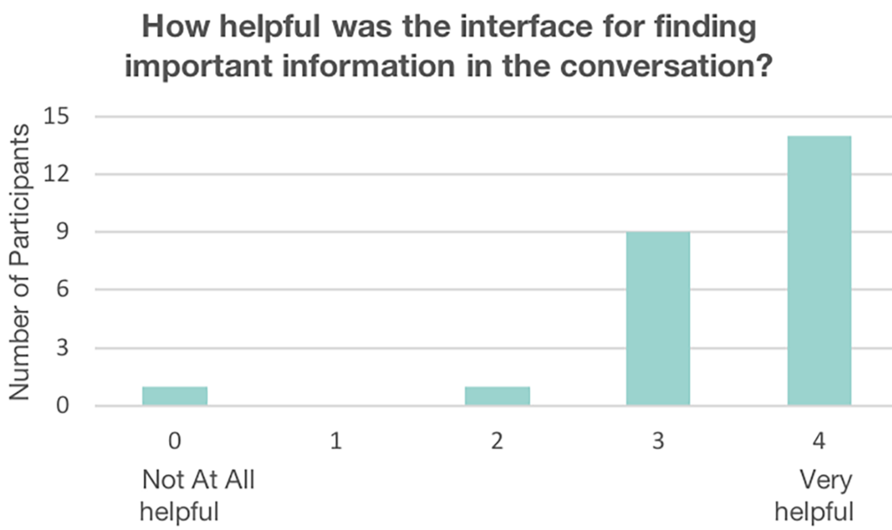


Figure 3-14: Helpfulness of the web interface for finding important information in the conversation.

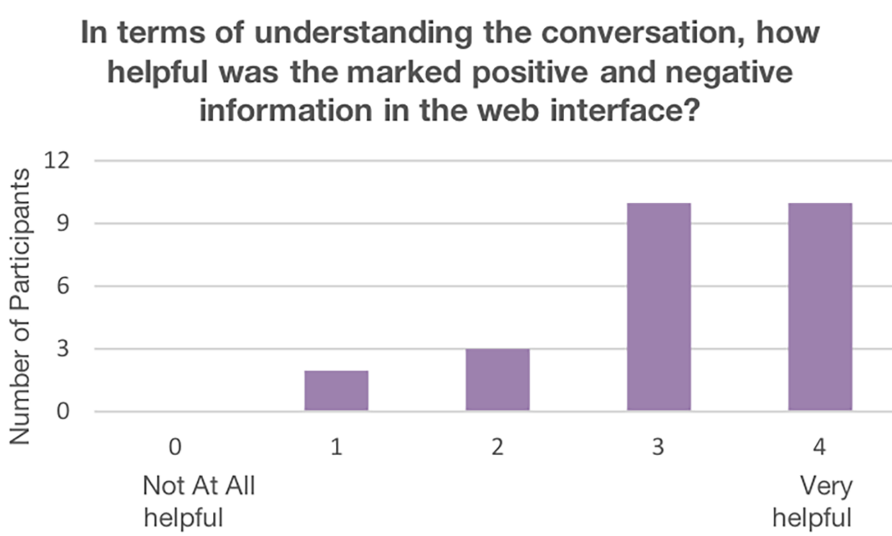


Figure 3-15: Helpfulness of the positive and negative labels in the web interface for understanding the conversation.

content, the transcript, the filters, the ability to navigate the transcript by clicking on the audio track, and the visualization of positive and negative information on the audio track. The least successful aspect of the web interface was that the positive and negative labels did not capture all of the important information in the conversation. As a result, participants suggested that more categories of information be labeled in the interface, particularly information about the treatment regimen.

In the final part of the user studies, I interviewed participants about their experience receiving the diagnosis. Obviously, a fictional scenario where the acting patient is already aware they will receive a cancer diagnosis is not equivalent to a real-life experience. However, most participants did make an effort to put themselves in the patient's shoes by imagining how they might react if the situation were real. Despite these limitations, several participants reported that receiving even a fictional diagnosis was somewhat emotional. One participant shared:

> *"Even though it's a fictional conversation, I think the first thing that came to mind, when you said, oh this is what you have, was like shock. You're totally bewildered."—Parul Koul, Undergraduate Student*

I consistently observed a delay between a participant hearing the cancer diagnosis and responding to it in a way that indicated they understood what was happening. After hearing the cancer diagnosis, participants often responded with the phrase, "Okay," and did not begin asking questions about their health options for several minutes, if at all. One participant described his experience acting:

> *"It took a while, there was a period where, the conversation progressed pretty far before I realized like, let's backup to square one to, what exactly is this disease? What is the prognosis? How is chemotherapy going to impact my life? Questions I didn't really think to ask when you first presented the news."—Justin Lueker, Graduate Student*

This lag in reaction time was also apparent in participants' manual notes. From the note records of 24 participants (one record was lost), 10 participants completely forgot to take notes during the conversation, five participants wrote less than six words and one participant recorded information about the fictional doctor's bedside manner instead of about the diagnosis. Of the remaining eight participants, only one took very thorough notes although, as she commented later:

> *"Afterwards they [the notes] can easily become confusing and the information is not clear. Also as I reached the bottom [of the page] then I had gone back and written up top and I realized I didn't know what order that had come in or what exactly it [my notes] were referring to. I had written down the 5-year and 20-year survival rates, but I wasn't sure if that was after one of the specific treatments or in general. I think it's easy for it to get out of order and scattered."—Graduate Student*

I asked participants about the specific challenges they experienced during the fictional diagnosis and whether the web interface helped them deal with any of those challenges. Participants commented that keeping track of all the medical terms and thinking of what questions to ask was very difficult. Some also found that even within the 4 to 5-hour break, they had forgotten some important information from the conversation. The web interface did not help alleviate those challenges during the conversation, but participants commented that it did serve as a useful tool for reviewing the conversation and could be helpful for further independent research:

> *"It gives a good overview of the whole situation and helps you view all the information in one context. And maybe it could be useful for figuring out what information you have and what information you still need to get. It would be useful to be able to look at it and then figure out, what are the next steps? Where do you go from here? What more do you need to know?"—Graduate Student*

Receiving a serious diagnosis, such as cancer, can be a remarkably disempowering experience. Regarding this aspect of the patient experience, I asked participants about their sense of control over their health information during the conversation and whether the interface changed their sense of control. Most participants commented they felt reasonably in control of their health information during the appointment, primarily because the fictional doctor answered all of their questions. Had this been a real diagnosis, I may have encountered different responses. With the web interface, 13 participants commented that they felt more in control of their health information because they had a tangible record of information for reference. Unexpectedly, three participants commented that the web interface actually made them feel less in control because they felt the positive and negative labels were telling them how to think.

Although the positive and negative classification of information was annotated manually for the user studies, in Chapter 4 of this thesis I develop an algorithm to automatically classify the information. Because this type of system would become an application of AI in healthcare for interpreting the valence of information, I was curious about participants' trust in the system. It is important to note that participants were not aware of how the information in the web interface had been labeled. All participants assumed that the conversations were labeled algorithmically. With this assumption 24 out of 25 participants expressed that they trusted the classification of information because it mostly agreed with their personal opinion. When the labels in the system did not match their personal opinion, 13 participants said it caused them to question the classification system overall. Some saw this as reason to pay more attention to the rest of the information in the transcript, just in case the classification did not accurately capture positive or negative important information. Others simply disregarded the system as wrong. Still others said the labels encouraged them to consider how the information could be interpreted differently. Two participants commented they were concerned that the labels might be misleading and cause users to ignore other important information that is not labeled as positive or negative, but is otherwise equally or more important.

I also discussed the multiple modalities of information review and representation in the web interface. Only eight participants said they used the audio as a tool to review the information. Most participants preferred to use the transcript because it was much faster to review and because they did not want to listen to their own voice. However, participants who did not use the audio commented that access to the audio record was valuable as a source of ground truth or for confirming the tone of the conversation. Nine participants said that the audio helped them trust the information in the interface and six said that the audio added a human aspect that was not captured in the transcript.

Finally, I asked participants how they would want this type of interface to be integrated into their healthcare experience. Most participants wanted this system to be available through an online health portal and wanted medical staff to be responsible for recording the conversations. When asked if they thought the web interface would be helpful for sharing information with or receiving health information from a loved one, 16 participants thought it would be valuable. Of those 16, 9 participants thought the audio feature could be helpful to hear exactly what the doctor said.

> *"My mom will tend to ask a lot of follow up questions when we have stuff. And like, I can imagine, particularly, if you get something that's a lot of bad news, you don't want to respond to a million follow up questions, particularly things that may not have been answered by the doctor, that it would be nice to give that [web interface] and they [parents] could see exactly what the doctor said."—Gabriel Terrasa, Undergraduate Student*

> *"Not everybody is capable of transmitting this information. So, for example, if my grandma goes to the doctor, probably she doesn't go alone, but if she goes alone, it's good that then she can show this [web interface] to my mother or other people."—Guadalupe Babio, Graduate Student*

One participant thought it could be a useful tool to monitor doctor performance and another thought it could be a useful tool to ensure patient compliance with physician instructions. Several participants commented they would have liked to see more features in the interface such as linking medical vocabulary to external resources and a summary page of the conversation.

## 3.4 Discussion of User Study Results

Based on the quantitative and qualitative results from the user study, I conclude that the prototype web interface is a helpful tool for reviewing medical conversations. In particular, a transcript with indications marking the positive and negative content is helpful for revisiting a conversation and finding important information. And although the audio playback feature was not useful for reviewing the information, the feature offered an element of truth and helped participants trust the content of the interface. However, further investigation is necessary to determine if the web interface positively affects patient understanding.

Additional improvements of this system might include labeling major topics in a conversation such as diagnosis or treatment details, a summary page describing the main points of the conversation, or linking external resources to keywords within the transcript. Of course, more user-testing with real medical patients is also necessary to determine what tools and features would be most helpful for managing information from medical conversations.

Reviewers from previous presentations of this thesis have expressed concerns that the visualization of the positive and negative content in the web interface may negatively affect a patient's psychological state, and thus result in poor health outcomes. This is an important topic that should be considered in future studies. However, this thesis is not about the psychological effect of information on patient outcomes; it is about helping patients gain better understanding and control over their health information. Studies in the United States and Europe found that patients want honesty and transparency regarding their health information, particularly when a diagnosis is grave [61]. Based on my personal experience with my father's disease, having the ability to reflect on a conversation, considering all the positive and all the negative information, is valuable. When a prognosis is extremely poor, patients and their caregivers need to know so they can make the necessary preparations with their family and choose how to live their remaining days. Alternatively, when a prognosis is uncertain, patients should have access to information that clearly identifies the risks and benefits of their treatment options. While my tool is a far cry from this vision of information management for patients, it is a step towards helping patients engage with their health information.

# Chapter 4

# Extracting Information From a Conversation

The second task of this thesis was to develop a machine learning algorithm to identify important positive and negative information from medical conversations using text and prosody. The dataset for this system consisted of the 25 recorded fictional medical conversations from the user studies in Chapter 3 and two additional recorded conversations from pilot studies. In total, these conversations amount to 3 hours, 25 minutes and 56 seconds of audio data, of which 2 hours, 37 minutes and 39 seconds the doctor character (me) is speaking. In this study, I analyzed only the doctor's speech to identify important positive and negative content in the conversations.

As mentioned in section 3.1, the collected audio recordings were transcribed to text, manually separated by speaker, and manually labeled as positive (1), negative (-1), mixed (2) and neutral (0). The definitions of these classes are described in Table 4.1. I manually separated speech events by listening for pauses and topic changes within the dialog. The length of speech events ranged from 0.68 to 56.47 seconds, averaging approximately 9.9 seconds overall (Figure 4-1). The labels for each speech event were determined by a majority vote from my opinion and the opinions of three additional people. Initial tests with all labels showed that the mixed category did not improve the classifier, therefore mixed speech events were omitted from the set.

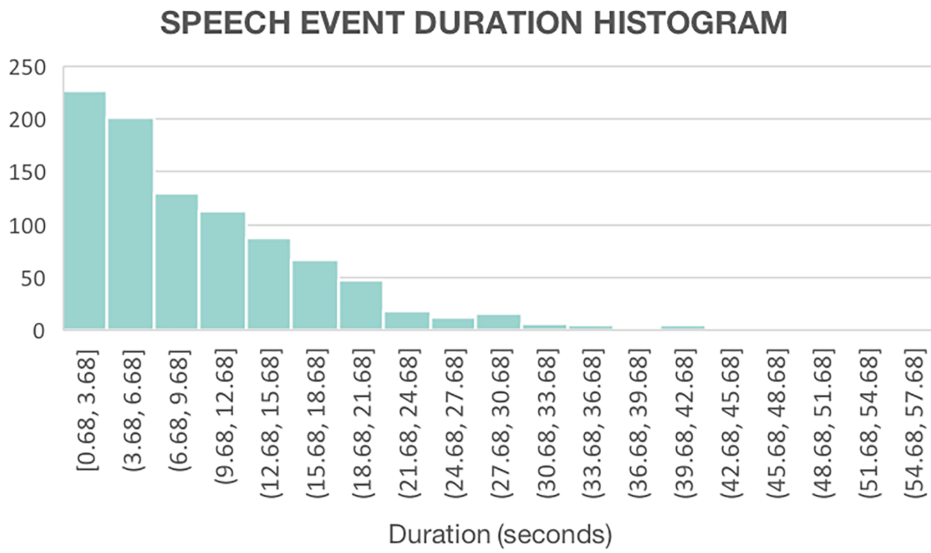


Figure 4-1: Histogram of speech event durations.

**Category Definitions**

| | | |
|---|---|---|
| **Mostly Positive** | 1 | Information that should cause a patient to feel optimistic about their treatment options, health outcome, diagnosis, or health resources. |
| **Mostly Negative** | -1 | Information that should cause a patient to feel pessimistic about their treatment options, health outcome, diagnosis or health resources. |
| **Neutral** | 0 | Information that should not cause a patient to feel either pessimistic or optimistic about their treatment options, diagnosis, health outcome, or health resources. |
| **Mixed** | 2 | Information that contains a roughly even distribution of positive and negative information about treatment options, diagnosis, health outcome, or health resources. |

Table 4.1: Definitions of positive, negative, neutral and mixed categories.

## 4.1 Dataset and Feature Extraction Methodology

For this task I used a multimodal approach to language understanding by analyzing a conversation by the content of the text along with the prosody in speech. In this study, I analyze only the doctor's speech to identify important positive and negative medical information for the patient. However, future studies may also incorporate the patient's response into the analysis.

### 4.1.1 Text Features

I used IBM Natural Language Understanding (IBM NLU) to extract sentiment and emotion measurements from the text [62]. IBM NLU required text passages to be six words or longer to perform the sentiment and emotion analysis. Speech events that were shorter than six words were omitted from the dataset. In total, I extracted four features related to the textual content of the conversations from IBM NLU: sentiment, joy, fear and sadness.

The sentiment analysis from IBM NLU returned confidence scores as decimal values on a range from -1 to 1. The more negative or positive a value, the more confident the model was that the text was negative or positive, respectively. Confidence scores close to zero meant the model considered the text content to be neutral. I empirically selected threshold values as follows: scores less than -0.5 were assigned -1 (negative), scores greater than 0.5 were assigned 1 (positive), and scores between -0.5 and 0.5 inclusive, were assigned 0 (neutral).

Emotion scores were returned on a different scale. IBM NLU returned probabilities for five emotions: joy, sadness, fear, anger, and disgust. Probability scores of 0.5 meant the model was uncertain if that emotion was present in the text. Scores higher than 0.5 meant the model was more certain that the emotion was present and scores lower than 0.5 meant the model was more certain that the emotion was not in the text. From these metrics, I thresholded the emotion scores as follows: scores higher than 0.5 were assigned 1 (emotion present) and scores lower than 0.5 inclusive were assigned 0 (emotion not present). I did not expect anger and disgust to be present in the doctor's speech events, so I only considered joy, sadness and fear measurements from this analysis.

I thresholded the IBM NLU emotion and sentiment scores to ensure that my machine learning algorithm was correctly interpreting their values. From initial tests with decision tree classifiers, I found unexpected and illogical operations at decision nodes such as low joy scores or high fear scores corresponding to positive information. Thresholding the emotion and sentiment scores reduced these types of artifacts in the algorithm.

In addition to language understanding, I also used Python's Natural Language Toolkit (NLTK) to identify features from the words in the text [63]. I tokenized each speech event and considered contractions such as “can't” or “won't” to be single words. I then computed lexical diversity and average word length. With initial tests, lexical diversity and word length did not show significant variance between positive, neutral, and negative speech events in this dataset. These features may become more relevant with a larger dataset containing more diverse conversations such as conversations from annual checkups, routine appointments during treatment, or diagnosis of serious illnesses.

### 4.1.2 Prosodic Features

I used Affectiva Automotive AI via Affectiva Emotion as a Service UI to analyze the audio recordings for prosodic features [64]. This service returned a time-based sequence of values corresponding to detected anger, laughter, and levels of vocal arousal in the audio recording. Vocal arousal is a measurement that accounts for loudness, pitch and phonemic duration. Because I was considering only the doctor's speech in my analysis, I did not expect anger and laughter to be relevant vocal features, and therefore only used the vocal arousal data for prosodic feature extraction. Future studies may also consider the patient's contribution to the conversation, in which case anger may become a relevant vocal feature.

The raw vocal arousal data consisted of 1.2 second speech events sampled every 0.3 seconds. This meant that a 1.5 second audio sample would have five sequential, but overlapping arousal measurements. As a starting point for analyzing this data, I applied a box filter with a window size of 5 to approximate the visualization tool on the Affectiva Emotion as a Service UI. Next, I applied a Guassian convolution filter of window size 11 and standard deviation of 1.5 to smooth the signal. I then applied a linear transform to scale the data to a range between 0 and 1. Finally, the smoothed signal was separated into segments defined by the start and end time of each speech event in the conversation.

Using the SciPy signal library [65], I extracted peaks from the normalized data whose maximum value was larger than 0.15. The threshold for the peak values was determined empirically. From the extracted peaks, I calculated the average peak height, the minimum peak height, the maximum peak height, the mode peak height (rounded to the nearest 0.1), the number of peaks per speech event and the average peak frequency per speech event. In addition to the peak properties in the data, I also computed the average vocal arousal and the average curvature of the vocal arousal data per speech event. I computed the curvature by taking the second derivative of the vocal arousal data using a centered window of size 11.

Along with vocal arousal features, I also considered the average rate of speech in each speech event. I computed this value as the number of words per speech event divided by total time per speech event. Further investigation of the rate of speech may also consider more localized speech rate or hesitations to characterize different types of speech events.

## 4.2 Training Classifiers

### 4.2.1 Evaluation

The goal of this task was to identify important information within a conversation using positive, negative and neutral labels. I acknowledge that neutral speech events may also contain important factual information from a conversation, however for the scope of this thesis I focused on identifying important information using positive and

| | **Positive** | **Negative** | **Neutral** |
|---|---|---|---|
| **Positive** | True Positive (TP) | False Negative (FN) | False Neutral (FU) |
| **Negative** | False Positive (FP) | True Negative (TN) | False Neutral (FU) |
| **Neutral** | False Positive (FP) | False Negative (FN) | True Neutral (TU) |

Table 4.2: Classification prediction confusion matrix definitions.

| | **Positive** | **Negative** | **Neutral** | **MACRO** |
|---|---|---|---|---|
| **Precision (Prec)** | $\frac{TP}{TP + FP_{neg} + FP_{neu}}$ | $\frac{TN}{TN + FN_{pos} + FN_{neu}}$ | $\frac{TU}{TU + FU_{pos} + FU_{neg}}$ | $\frac{\Sigma\,(Precision)}{3}$ |
| **Recall (Rec)** | $\frac{TP}{TP + FN_{pos} + FU_{pos}}$ | $\frac{TN}{TN + FP_{neg} + FU_{neg}}$ | $\frac{TU}{TU + FP_{neu} + FN_{neu}}$ | $\frac{\Sigma\,(Recall)}{3}$ |
| **F1-score** | $\frac{2 * Prec_{pos} * Rec_{pos}}{(Prec_{pos} + Rec_{pos})}$ | $\frac{2 * Prec_{neg} * Rec_{neg}}{(Prec_{neg} + Rec_{neg})}$ | $\frac{2 * Prec_{neu} * Rec_{neu}}{(Prec_{neu} + Rec_{neu})}$ | $\frac{\Sigma\,(F1\text{-}score)}{3}$ |
| **Accuracy** | $\frac{TP + TN + TU}{ALL}$ | | | |

Table 4.3: Accuracy calculations.

negative valence. As described in Table 4.2, with three classification categories, there were six possible outcomes: true positive (TP), false positive (FP), true negative (TN), false negative (FN), true neutral (TU) and false neutral (FU). To compute the accuracy of the classifier I considered precision, recall and accuracy (Table 4.3).

### 4.2.2 Training and Validation

Initial tests with this dataset showed that it contained a total of 842 speech events: 141 positive, 138 negative and 563 neutral. To correct this imbalance, I duplicated sufficient positive and negative speech events to equal the number of neutral speech events within the set. The resulting dataset included a total of 1689 speech events: 563 positive, negative and neutral labeled speech events. With this balanced dataset the chance probability of choosing any particular label is exactly 1/3 (0.33).

I used the balanced dataset to train decision tree models and support vector machine models using the Python machine learning library SciKit Learn [66]. Each model was trained with a random 70% train and 30% test split. Cross validation scores were also computed for each model using and 80% train and 20% test split. I trained separate models for text features and prosodic features, then combined text and prosodic features into a multimodal classifier for comparison. Features in each of these were selected empirically and the final models were selected according to the highest average cross validation score.

### 4.2.3 Results

The final multimodal decision tree classifier achieved the highest accuracy of 90.6%, which is 2.71 times better than chance, and an F1-score of 0.888. Unexpectedly, the accuracy after combining text features and prosodic features did not improve the final model. In fact, the average accuracy of the Prosody model is slightly higher than that of the Text-Prosody model (Fig. 4-2). It is possible that by analyzing prosody alone, the decision tree model found distinct vocal patterns in the way I delivered positive and negative information, thus resulting in a very high accuracy for the Prosody model. However, with the Text model only achieving an accuracy of 0.541, combining the text analysis with the prosodic analysis may have introduced more disorder into the dataset. The final tree structure contained 216 decision nodes and 217 leaf nodes with a minimum of 1 speech event and a maximum of 58 speech events at a single leaf node. The average number of speech events per leaf node was 5.4 with a 7.76 standard of deviation.

The confusion matrix in Table 4.5 gives a more detailed look at the classification results. The model achieves recall values near 100% for positive and negative speech events, but only 65.2% for neutral speech events (Table 4.6). The model also exhibits precision of 82.2% for positive labels, 91.5% for negative labels, and 96.2% for neutral labels. These results suggest that the model is well-equipped to separate positive information from negative information, but less equipped to separated neutral information from positive or negative information. Based on the precision measurements,

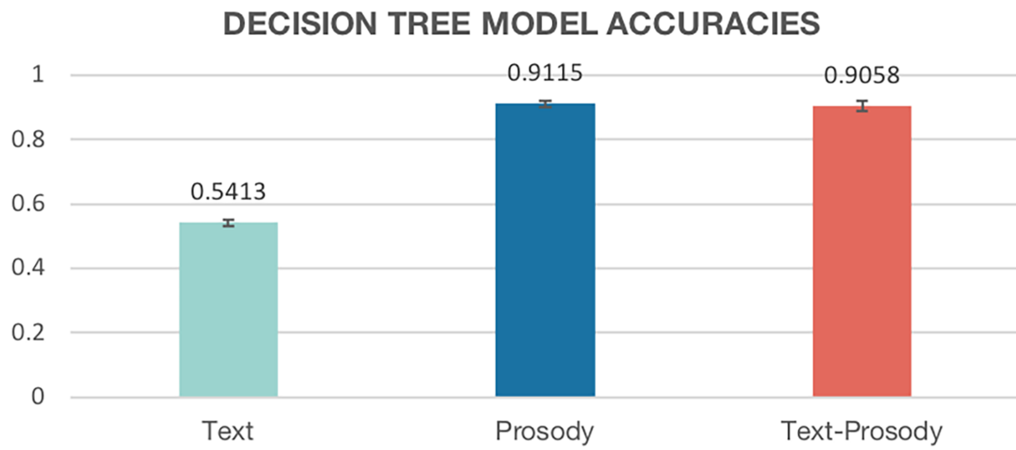


Figure 4-2: Accuracy of decision tree models.

| **Text-Prosody Model Cross Validation Scores** | | | | |
|---|---|---|---|---|
| 0.914 | 0.882 | 0.917 | 0.923 | 0.893 |

Table 4.4: Cross validation scores of decision tree model.

| | **Positive** | **Negative** | **Neutral** |
|---|---|---|---|
| **Positive** | 176 | 0 | 0 |
| **Negative** | 0 | 172 | 4 |
| **Neutral** | 38 | 16 | 101 |

Table 4.5: Confusion matrix of classification results from decision tree model.

| | **Positive** | **Negative** | **Neutral** | **Macro Scores** |
|---|---|---|---|---|
| **Recall** | 1.000 | 0.977 | 0.652 | 0.876 |
| **Precision** | 0.822 | 0.915 | 0.962 | 0.900 |
| **F1-score** | 0.903 | 0.945 | 0.777 | 0.888 |

Table 4.6: Accuracy results from decision tree model

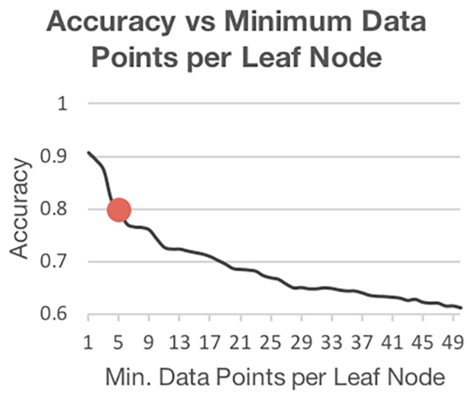


Figure 4-3: Accuracy as a result of the minimum data points per leaf node.

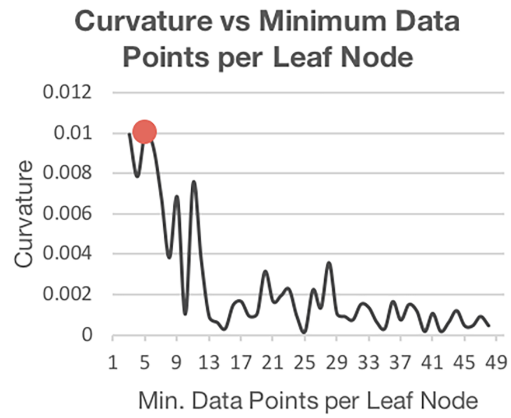


Figure 4-4: Curvature versus the minimum data points per leaf node.

the model appears slightly better at separating negative information from neutral information than positive information from neutral information. Although the cross-validation scores are reasonably consistent (Table 4.4), the nearly perfect accuracy and recall suggest that the model is overfitting to the dataset.

In an attempt to minimize overfitting to the training data, I modified the previous decision trees to require a minimum of five data points per leaf node. I selected this number by plotting the accuracy of the model against the minimum number of data points per leaf node (Figure 4-3) to identify the point with the maximum curvature (Fig. 4-4). This change reduced the overall accuracy of each decision tree model. The Text-Prosody decision tree achieved an average accuracy of 79.8% and F1-score of 0.726. Again, the Text-Prosody model faired worse than the Prosody model (Figure 4-5).

A closer inspection of Tables 4.8 and 4.9 show that all accuracy measurements are significantly reduced with this model. Recall for neutral and positive speech events is reduced by approximately 20%. Recall for negative speech events performs slightly better and is only reduced by about 13%. Precision is reduced by approximately 17% for positive speech events, 15% for negative speech events, and 25% for neutral speech events. While the decline in accuracy in all categories is sizeable, the reduction in precision for neutral speech events suggests that many of the deeper decision nodes in the original decision tree were identifying true neutrals. These scores also suggest that,

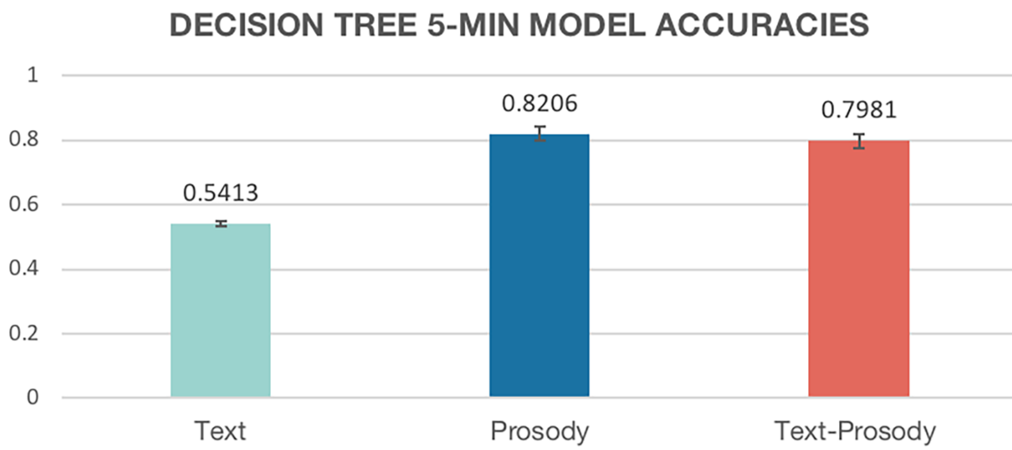


Figure 4-5: Accuracy of decision trees with a minimum of 5 points per leaf node.

| Text-Prosody Model Cross Validation Scores | | | | |
|---|---|---|---|---|
| 0.799 | 0.817 | 0.817 | 0.810 | 0.807 |

Table 4.7: Cross validation scores of Decision Tree 5-Min.

| | **Positive** | **Negative** | **Neutral** |
|---|---|---|---|
| **Positive** | 140 | 18 | 18 |
| **Negative** | 14 | 149 | 13 |
| **Neutral** | 50 | 24 | 81 |

Table 4.8: Confusion matrix of classification results from Decision Tree 5-Min.

| | **Positive** | **Negative** | **Neutral** | **Macro Scores** |
|---|---|---|---|---|
| **Recall** | 0.796 | 0.847 | 0.523 | 0.722 |
| **Precision** | 0.686 | 0.780 | 0.723 | 0.730 |
| **F1-score** | 0.737 | 0.812 | 0.607 | 0.726 |

Table 4.9: Accuracy results from Decision Tree 5-Min.

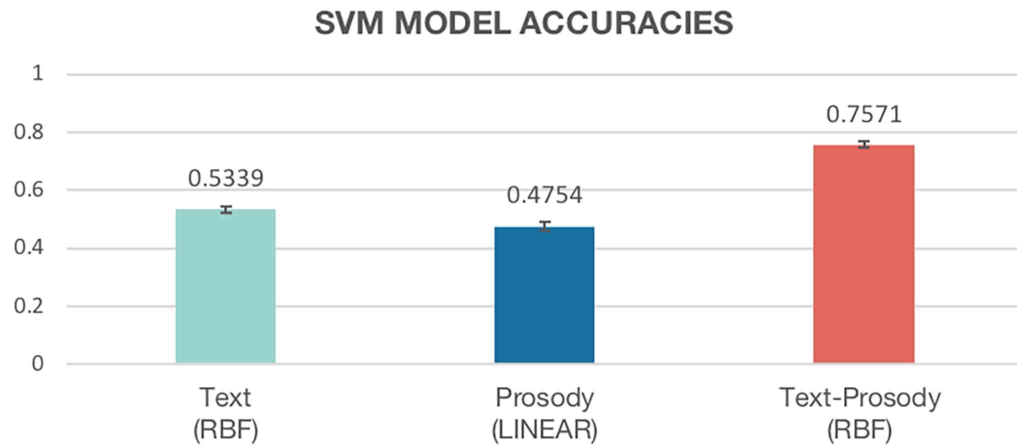


Figure 4-6: Accuracy of SVM models.

| **Text-Prosody Model Cross Validation Scores** | | | | |
|---|---|---|---|---|
| 0.773 | 0.743 | 0.761 | 0.756 | 0.753 |

Table 4.10: Cross validation scores of SVM model.

even with more constraints on the structure of the decision tree, negative information is most successfully identified out of the three categories.

The accuracy of the SVM model was lower than those from the decision tree models (Figure 4-6). The Text SVM achieved an average accuracy of 53.4% and the Prosody SVM achieved an average accuracy of 47.5%. Unlike the decision tree models, the combined Text-Prosody SVM model resulted in significantly better accuracy than the Text or Prosody models. This model achieved an accuracy of 75.7%, which is 2.27 times better than chance, and an F1-score of 0.721.

Looking closer at the classification results in Tables 4.11 and 4.12, the Text-Prosody SVM model achieved 73.6% recall for all positive speech events, 77.3% recall for negative speech events and 63.8% for all neutral speech events. The model also achieved 68.8% precision for positive labels, 71.4% for negative labels and 77.6% for neutral labels. From these results, this model appears to be slightly better at correctly classifying negative information than positive information.

To evaluate how well my models perform with other speakers, I obtained permission to use five additional conversations of doctors and nurses speaking with patients.

| | **Positive** | **Negative** | **Neutral** |
|---|---|---|---|
| **Positive** | 128 | 24 | 22 |
| **Negative** | 35 | 140 | 6 |
| **Neutral** | 23 | 32 | 97 |

Table 4.11: Confusion matrix of classification results from SVM model.

| | **Positive** | **Negative** | **Neutral** | **Macro Scores** |
|---|---|---|---|---|
| **Recall** | 0.736 | 0.773 | 0.638 | 0.716 |
| **Precision** | 0.688 | 0.714 | 0.776 | 0.726 |
| **F1-score** | 0.711 | 0.743 | 0.700 | 0.721 |

Table 4.12: Accuracy Results from SVM Model.

One conversation was a training video by Canadian Culture and Communication for Nurses (CCCN) providing a good example of how to deliver bad news to patients [67]. The four additional conversations were videos of real patients talking to real doctors from Brown Alpert Medical School [68]–[71]. These videos included a variety of appointments such as routine checkups, disclosing medical errors, and cancer diagnosis. There were four different medical professionals in these videos: three were female and one was male. Again, I analyzed only the doctors' (or nurses') speech and established ground truth by the selecting the majority vote from four peoples' opinions.

The resulting accuracy of my three models on these conversations can be seen in Table 4.13. The original decision tree still achieved the highest accuracy of all the models at 69.8%, but fared considerably worse than the 90.6% accuracy achieved with the data containing only my voice. The second-best model was the SVM with an accuracy of 63.2%, which is nearly 10% lower than the model's accuracy with the data containing only my voice.

Taking a closer look at the confusion matrices for each of these models in Tables 4.14, 4.15 and 4.16, the accuracy is primarily due to the models' success at identifying neutral speech events. Each model exhibits particular weakness at identifying positive information.

| Model | Accuracy | F1-score |
|---|---|---|
| Decision Tree | 0.698 | 0.589 |
| Decision Tree 5-Min | 0.575 | 0.490 |
| SVM | 0.632 | 0.533 |

Table 4.13: Accuracy of models on five additional conversations of patients speaking with doctors.

| | **Positive** | **Negative** | **Neutral** | **Recall** |
|---|---|---|---|---|
| **Positive** | 6 | 0 | 7 | 0.462 |
| **Negative** | 3 | 9 | 7 | 0.474 |
| **Neutral** | 11 | 4 | 59 | 0.797 |
| **Precision** | 0.300 | 0.692 | 0.808 | |

Table 4.14: Confusion matrix of classification results from decision tree model on five additional conversations.

| | **Positive** | **Negative** | **Neutral** | **Recall** |
|---|---|---|---|---|
| **Positive** | 5 | 1 | 7 | 0.385 |
| **Negative** | 5 | 8 | 6 | 0.421 |
| **Neutral** | 20 | 6 | 48 | 0.649 |
| **Precision** | 0.167 | 0.533 | 0.787 | |

Table 4.15: Confusion matrix of classification results from Decision Tree 5-Min model on five additional conversations.

| | **Positive** | **Negative** | **Neutral** | **Recall** |
|---|---|---|---|---|
| **Positive** | 5 | 3 | 5 | 0.385 |
| **Negative** | 5 | 11 | 3 | 0.579 |
| **Neutral** | 14 | 9 | 51 | 0.689 |
| **Precision** | 0.208 | 0.478 | 0.864 | |

Table 4.16: Confusion matrix of classification results from SVM model on five additional conversations.

It may be worth noting that the speakers in each of these conversations had very different paralinguistic characteristics. One female speaker had a fairly deep voice and spoke with a very somber and slow cadence. The second female speaker had a higher voice and could be described as shy. The third female speaker had a fairly high voice and spoke quickly. The male speaker also had a fairly high-pitched voice and spoke in a friendly, but very fast and clipped manner. I describe these characteristics because the qualities of my voice that define good or bad news may not be generalizable to all personalities, cultures or regional speech patterns. Larger and more diverse datasets may show that a generalizable model can be trained by culture or personality type. Alternatively, future studies may find that hyper-personalized models trained per individual care provider are scalable and offer the highest accuracies.

# Chapter 5

# Conclusions and Future Work

In the first task of this thesis, I built a novel multimodal prototype web interface that represents information from medical conversations to facilitate patient review and understanding of their health information. The interface included a transcript as well as an audio playback system to revisit the content of a conversation. Information in the transcript and corresponding time segments on the audio timeline were highlighted to inform the user of important positive and negative information. Labels in this first task were determined manually. Additional features included a chart visualizing the total number of positive and negative speech events in the conversation and filters to isolate positive or negative information.

I evaluated the prototype web interface in a controlled study with 25 participants. Results from the study indicate that, at least in a fictional setting, the web interface helps patients review the content of conversations from medical appointments. More specifically, features such as the text transcript and the positive and negative labels on information helped participants navigate through the conversation and find important information. The web interface's effect on participants' understanding of information and optimism about their health options cannot be determined from these results. Further studies and additional methods for evaluation will be necessary to explore these topics.

In the second task of this thesis, I developed machine learning algorithms to extract the important positive and negative information from medical conversations. The algorithms were trained on a dataset of 27 fictional conversations where I assumed the role of doctor. For the purpose of this study, I only used speech events containing my voice. I considered features extracted from the text using IBM NLU and prosodic features from vocal arousal using the analysis from Affectiva Automotive AI. The most successful algorithm was a decision tree, achieving an accuracy of 90.6% and an F1-score of 0.888. Unexpectedly, the decision tree's accuracy did not improve when combining textual and prosodic features into the learning model. A decision tree using only prosodic features achieved an accuracy of 91.1%.

The machine learning algorithms were trained only on my voice, and were therefore very likely to overfit to the way I speak. To test the generalizability of my models, I collected five additional conversations from CCCN and Brown Alpert Medical School. As expected, the accuracy results from these conversations were significantly lower than on the conversations containing only my voice. The decision tree model still achieved the highest performance with an accuracy of 69.8% and F1-score of 0.589.

Future work for extracting positive and negative information from conversations based on prosody and text will require a significantly larger dataset. Particularly within the medical context, it will be important to collect a dataset of real doctors speaking with real patients with speakers from different geographic regions, different personalities and of different genders. However, further studies may show that a truly generalizable model is not realistic and that individual healthcare providers will need personalized models to interpret their manner of speaking and conveying information. Future algorithms may also consider the patient's reaction in addition to the doctor's information.

A significant comment from the study participants was that the positive and negative labels did not highlight enough of the important information in the conversation. Users indicated they would have liked additional labels to mark information such as treatment options, medications, and side-effects. Others also indicated they would have liked a summary page in addition to the transcript to help them quickly assess the main points of the conversation. In the future, additional machine learning and natural language processing algorithms may be used to develop these features in the web interface and further aid patients with review and retrieval of their medical information.

While the focus of this study was the impact of information representation and information extraction in the medical context, these findings may be relevant to other fields as well. As discussed in Chapter 2, the way we speak has an enormous influence on how information is received and understood. A machine learning system that uses prosody and text to identify important elements in a vocal exchange could become relevant in sportscasting, public presentations, design critiques and more. Current tools for aiding human-human communication offer limited modalities such as audio inputs and visual outputs. However, these systems place the burden of interpreting, understanding and synthesizing information completely on the end-user. In emotional communication settings, interpreting and managing information can be overwhelming, thus rendering existing methods of information management too limiting. With machine learning, emotionally intelligent multimodal systems may offer new methods for interacting with information within emotional contexts. Such systems could encourage end-users to reflect on and consider alternative perspectives, which may improve understanding and engagement with information.

# Appendix A

# User Study: Fact Sheet

Your name:

- Mr./Ms. Doe
  *(If you have a preferred prefix, please inform the study proctor now)*

Scenario:

- Betacytes are a cell mutation detected in the blood.

- Last year, your doctor identified a high level of betacytes in your body. At that time, the level did not pose any serious health risks, but was higher than normal levels. As a result, your doctor has you come into the clinic every 6 months to monitor the level of betacytes in your body.

- High levels of betacytes are generally associated with fatigue, loss of appetite and weight loss. You may express to the doctor that you've experienced any of these symptoms.

- You are aware that betacytes in the body can sometimes lead to cancer. But so far, the doctors say the level of betacytes in your body is not cancerous.

- You had recent labs taken a few days ago to check the current status of betacytes in your body. Today you will get the results from your most recent medical tests.

- You are your current age (in real life).

- You came to this appointment by yourself.

- You recently got a new pet. You can share this information with the doctor if appropriate.

DISCLAIMER:
All medical conditions mentioned in this scenario and the following conversation are fictional. Any information you may encounter or think you know about these medical conditions external to this conversation are not related to this study.

# Appendix B

# User Study: Script

Level of Betacytes

| Healthy | Not-Cancerous | Stage I | Stage II | Stage III |
|---|---|---|---|---|
| 1% | <3% | <4% | <5% | <6% |

Treatments

| | | |
|---|---|---|
| Chemotherapy | Pegaspar<br>Methotrexamine | Travels through the bloodstream to kill cancer cells. |
| Targeted Antibodies | Arrantraxibar<br>Cytocaptopurine | Latches on to cells with specific cancerous markers to kill them. |
| Radiation | | Radioactive particles are delivered either externally or internally through IV to damage the DNA in the cancerous cells, causing them to die. |

Statistics

| | |
|---|---|
| 5-year survival | More than 90% of people live past 5 years. |
| 20-year survival | 65% stay in remission<br>35% may develop cancer again |
| No treatment | Very aggressive – 1 year.<br>Less aggressive – 5 years. |

Side Effects

Hair loss, nausea, fatigue, loss of appetite, weaker immune system.

Other Answers:

- Betacyte Carcinoma is not genetic.
- It can spread to solid tissue.
- I don’t have that information with me today, but I will check with the care team and get back to you.
- I don’t have the specifics about that information, but the care team will provide you with additional materials with more details about that.

Hi Mr./Mrs. Doe. It's nice to see you again, thanks for coming in today.

[…]

How are you today?

[…]

Wonderful. Alright, is there anyone here with you today that you would like to be here while we discuss your information?

[…]

Well, I'd just like to start by asking how have you been feeling since we last met?

[…]

Ok, so last year we identified an abnormal level of **betacytes** in your system which is why we have been having regular checkups every six months. Do you remember that?

[…]

I have the results from your most recent labs right here. How much detail would you like today?

[…]

Mr./Mrs. Doe, unfortunately it does look like the number of **betacytes** has recently increased to a critical level. I'm sorry to say that this level of **betacytes** in your body indicates that you now have a cancer called **betacyte carcinoma**.

[…]

The cancer appears to be fairly aggressive. The results from the lab show it has already advanced to stage II.

[…]

Mr./Mrs. Doe, I realize this is a lot to take in and I just want to let you know that I and the care team are all here to support you.

[…]

It is important we start thinking about treatment as soon as possible. Can I tell you about some of the treatment options now?

[…]

**Betacyte carcinoma** has been widely studied for many years, which means there is a lot of research and treatments available. With treatment, survival rates are very high.

[…]

**Betacyte carcinoma** is most commonly treated with chemical therapies such as chemotherapy or targeted therapies. And how these treatments work is that we deliver them through an IV to attack the cancerous cells to prevent them from multiplying and growing.

[…]

A combination of chemotherapy and radiation is also an effective treatment strategy, particularly for aggressive cancers.

[…]

In your case, we will start with a chemical therapy and see how you respond. Chemical therapies are often a cure for **betacyte carcinoma**, however, cancer progresses differently for everyone and we'll be paying close attention to make sure we apply the most effective treatment.

[…]

The care team is going to provide you with more detailed information about your treatment options. Take some time to look over these options and discuss with your loved ones.

[…]

Before you leave today, you'll need to schedule your next appointment at the front desk so we determine the next steps.

# Appendix C

# User Study: Survey

# Review the Appointment

The terms "positive," "negative," and "neutral" are used to describe information from the conversation. The definitions of these terms are outlined below:

• POSITIVE information implies a good outcome or makes you feel optimistic.
• NEGATIVE information implies a bad outcome or makes you feel pessimistic.
• NEUTRAL information makes you feel NEITHER optimistic NOR pessimistic.

1. **Email address** *

## Manual Notes

Please refer to your handwritten notes to respond to the following questions based on the simulated doctor's appointment earlier today.

As mentioned previously, the definitions of positive, negative and neutral are as follows:

• POSITIVE information implies a good outcome or makes you feel optimistic.
• NEGATIVE information implies a bad outcome or makes you feel pessimistic.
• NEUTRAL information makes you feel NEITHER optimistic NOR pessimistic.

2. **In your opinion, was there any positive information from the conversation?**
   *Mark only one oval.*

   YES
   NO

3. **If yes, please briefly describe each piece of positive information. (Bullet points are fine)**

4. **In your opinion, was there any negative information from the conversation?**
   *Mark only one oval.*

   YES
   NO

5. **If yes, please briefly describe each piece of negative information. (Bullet points are fine)**

6. **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?**
   *Mark only one oval.*

   - YES
   - NO

7. **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)**

8. **How well do you feel you understand the content of the conversation?**
   *Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| NOT AT ALL | | | | | | VERY WELL |

9. **In your opinion, the overall the content of the conversation was:**
   *Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| Very Negative | | | | | | Very Positive |

10. **Based on the conversation, how would you rate your optimism about your treatment options?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| NOT AT ALL optimistic | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | VERY optimistic |

11. **Based on this conversation, how confident would you feel about making a decision about your treatment?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| NOT AT ALL confident | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | VERY confident |

## Web Interface

Please respond to following questions using the web interface provided to you by the study proctor. You may also reference your manual notes.

As mentioned previously, the definitions of positive, negative and neutral are as follows:

• POSITIVE information implies a good outcome or makes you feel optimistic.
• NEGATIVE information implies a bad outcome or makes you feel pessimistic.
• NEUTRAL information makes you feel NEITHER optimistic NOR pessimistic.

## The web interface highlights what the doctor considers positive information or negative information. The next three questions are designed to help you become familiar with the web interface.

12. **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?**

________________________

13. **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.**

14. **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:**

    *Mark only one oval.*

    - Positive
    - Negative
    - Neutral

## Using the web interface as an additional source of information, please respond to the following questions based on your understanding of the conversation.

15. **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)**

16. **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)**

17. **Based on your understanding, the overall content of the conversation was:**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| Very Negative | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Very Positive |

18. **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| Very Pessimistic | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Very Optimistic |

19. **After viewing the web interface, how confident do you feel about making a decision about your treatment?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| NOT AT ALL confident | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | VERY confident |

20. **How well do you feel you understand the content of the conversation?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| NOT AT ALL | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | VERY WELL |

21. **How much do you think the information represented in the web interface influenced your opinion of the conversation?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| No Influence On My Opinion | ◯ | ◯ | ◯ | ◯ | ◯ | Completely Influenced My Opinion |

22. **Was there information that the web interface highlighted as positive that you thought was NOT positive?**
*Mark only one oval.*

◯ YES
◯ NO

23. **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)**

24. **Was there information that the web interface marked as negative that you thought was NOT negative?**
*Mark only one oval.*

◯ YES
◯ NO

25. **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)**

## Your Experience

As mentioned previously, the definitions of positive, negative and neutral are as follows:

- POSITIVE information implies a good outcome or makes you feel optimistic.
- NEGATIVE information implies a bad outcome or makes you feel pessimistic.
- NEUTRAL information makes you feel NEITHER optimistic NOR pessimistic.

26. **In terms of usability, how would you rate the web interface?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| Not easy to use | ◯ | ◯ | ◯ | ◯ | ◯ | Very easy to use |

27. **How helpful was the interface for finding important information in the conversation?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| NOT AT ALL helpful | ◯ | ◯ | ◯ | ◯ | ◯ | VERY helpful |

28. **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| NOT AT ALL helpful | ◯ | ◯ | ◯ | ◯ | ◯ | VERY helpful |

29. **How would you rate your overall experience with the web interface?**
*Mark only one oval.*

| | 0 | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|
| TERRIBLE | ◯ | ◯ | ◯ | ◯ | ◯ | VERY PLEASANT |

30. **What do you think worked well in the interface?**

31. **What do you think did NOT work well in the interface?**

32. **Do you have comments or suggestions to improve the web interface?**

☐ Send me a copy of my responses.

Powered by Google Forms

# Appendix D

# User Study: Interview Questions

Outline of Interview Questions:

1. Did you trust the positive/negative classification of information in the transcript? Why?
2. Can you share some of the biggest challenges during the conversation?
3. How did the introduction of the interface alleviate those challenges, if at all?
4. How much control of your health information did you feel during the conversation and part 1 of the questionnaire?
5. Did interface change your perception of control of your health information? Why?
6. If you did not use the audio features of the interface, why not?
7. When do you think you would use the audio features, if ever?
8. How would you like a system like this to be integrated into your healthcare experience?

# Appendix E

# User Study: Survey Responses

| | Participant 1 | Participant 2 | Participant 3 |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | The explanation about the percentage of chance to get cured from the cancer and detailed information about the options and those cure chances | 90% Survival rate for 5 years with treatment of ~3 months. 20 year survival rate was ~60% but that seems pretty normal. | - The conditions was treatable<br>- There were multiple treatment options<br>- The success rate was very high |
| **In your opinion, was there any negative information from the conversation?** | NO | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | | I have cancer. | - That I have cancer<br>- That the side effects are really harmful |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | YES | YES | NO |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | The cancer stage and its meaning<br>The side effects of therapy sessions | There's no information on what treatment option is known to be the best. | |
| **How well do you feel you understand the content of the conversation?** | 4 | 3 | 3 |
| **In your opinion, the overall the content of the conversation was:** | 3 | 3 | 1 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 3 | 4 | 3 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 3 | 3 | 1 |

| | **Participant 4** | **Participant 5** | **Participant 6** |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | Very informative and friendly doctor | There is options, now cancer is curable | At stage 2, 90% survival then after a few years it's 65% survival; multiple treatment options and many are covered by insurance; in around a month the treatment should show signs of helping |
| **In your opinion, was there any negative information from the conversation?** | NO | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | | either quimiotherpy or tarteget have the same impact, why decide one or other<br>the doctor don't empathise | Will likely have to take time off from work each week because of side-effects after treatment session; even with monitoring of betacyte levels twice a year the cancer was caught at stage 2 and not 1; dramatic spike in betacyte number in 6 months |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | YES | YES | NO |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | That I work in neuroscience / medical research field | the names of the treatments | |
| **How well do you feel you understand the content of the conversation?** | 4 | 3 | 4 |
| **In your opinion, the overall the content of the conversation was:** | 3 | 1 | 4 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 3 | 4 | 3 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 3 | 1 | 3 |

| | **Participant 7** | **Participant 8** | **Participant 9** |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | That I have a decently high chance of survival<br>That it is not hereditary and won't impact me if I want to have children in the future<br>That treatment can be done in a few months<br>That the doctors will be very supportive<br>That I am only stage 2 | the success rates of treatment are relatively high, and that this is a common disease/problem that there are multiple options for treating. | Good survival rates (90+% for 5 year, 60+ for 20 years).<br>Stage 2 Cancer can be treated and people can continue to work through treatment. |
| **In your opinion, was there any negative information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | That I have cancer in the first place<br>That the treatment will have bad side effects and will leave me unable to work for at least 1 day a week<br>That I can't do much to prevent it outside of getting treatments for it<br>That it might come back in the future | Sometimes other cancers arise unrelated, and the 20 year survival rate did not seem that good. Also, of course the diagnosis itself is negative, and the treatment side effects are negative (hair loss, etc.) | The phrase "the cancer is aggressive"<br>Side effects of treatment<br>The possibility of recurrence or developing another form of cancer after treatment<br>People working through treatment have lower energy |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | YES | YES | YES |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | That my insurance will be able to cover the treatments, though I was not told how much I'd have to pay so this may be positive or negative depending on how much | I need to have someone pick me up from treatment, treatment is once a week for four hours at a time for three months. There are two main options for treatment, with personal variation in reactions. | We can wait a week before making any decision.<br>Targeted antibodies has the same side effects as chemo but with no hair loss.<br>Other blood reports came out normal. |
| **How well do you feel you understand the content of the conversation?** | 2 | 2 | 3 |
| **In your opinion, the overall the content of the conversation was:** | 2 | 2 | 2 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 4 | 3 | 3 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 1 | 2 | 3 |

| | Participant 10 | Participant 11 | Participant 12 |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | 5-year with 90% survival rate, are various different treatment options | 90% survival rate for 5 years<br>two different effective treatments | Very little felt positive but some of the information around being "stage2" as opposed to other stages helped |
| **In your opinion, was there any negative information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | Chemotherapy methods are more successful but have more side effects vs. targeting antibodies; 20-year with only 65% survival rate; that my betacyte levels have increased significantly | you have cancer<br>less survival rate after 20 years | Looking at survival statistics was alarming, thinking about it not being curable |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | YES | NO | YES |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | That I'm at Stage 2 already - I didn't ask further to know how positive or negative this is | | Some of the conversation around the various treatments felt more neutral |
| **How well do you feel you understand the content of the conversation?** | 3 | 4 | 3 |
| **In your opinion, the overall the content of the conversation was:** | 1 | 2 | 2 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 2 | 4 | 2 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 3 | 3 | 3 |

| | Participant 13 | Participant 14 | Participant 15 |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | There was a high likelihood of surviving the cancer in the short-term<br>There was a choice in treatment<br>The cancer was caught early<br>The cancer most likely hadn't spread to other parts of the body yet<br>The cancer was being carefully monitored<br>There were good treatment options in place | 1) There is a treatment for the affliction.<br>2) The success rate of the treatment seems to be high. | -There's a 90% survival rate for 5 years after treatment<br>-Treatment typically eradicates all cancerous cells if I choose chemotherapy<br>-There are a lot of doctors and hospital staff willing and ready to support me with my decision<br>-I have some time to think about which treatment I want to go with |
| **In your opinion, was there any negative information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | The long term surviving the cancer (20 years) was not that strong<br>Diagnosed with cancer<br>No clear reason why the cancer became more aggressive going from stage 0 to 2<br>Effects of cancer treatment on lifestyle and health | 1) I am worried about the side effects of chemotherapy, and how this might affect my family.<br>2) I am worried about the cost of the treatment. | -My cancer cells have progressed and are now Stage 2<br>-The survival rate over a longer period than 5 years drops down pretty significantly even with treatment<br>-If I don't get treatment I have between a couple months and 5 years to live.<br>-Chemotherapy would cause my hair to fall out<br>-The less invasive treatment could potentially let other unidentified cancer cells continue living |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | YES | NO | YES |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | No clear benefits between one cancer versus the other | | -I was referred to my insurance for price information so since I wasn't given any it was pretty neutral |
| **How well do you feel you understand the content of the conversation?** | 4 | 4 | 3 |
| **In your opinion, the overall the content of the conversation was:** | 3 | 3 | 1 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 2 | 3 | 2 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 2 | 2 | 3 |

| | Participant 16 | Participant 17 | Participant 18 |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | NO |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | It was a highly popular and therefore well researched cancer, so there are some good treatments. 5-year survival rate is fairly high. | there is a 90% survival rate in 5 years and 65% in 20 years. there is treatment available because its a common cancer. there are medications | |
| **In your opinion, was there any negative information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | I had been diagnosed with the cancer. 20-year survival rate is not that high (60%). The most effective way to treat it would be chemo which can have strong negative side effects. | I have been diagnosed with cancer. I will be enduring chemo which affects hair growth etc. I have to start treatment soon. | Cancer diagnosis, discussion of side effects, suggestion that I should take a semester off |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | NO | YES | YES |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | | the info about the medicine and the appointments for treatment. I have had high betacytes for a while. | Description of treatment options, description of disease |
| **How well do you feel you understand the content of the conversation?** | 4 | 3 | 3 |
| **In your opinion, the overall the content of the conversation was:** | 1 | 1 | 1 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 2 | 3 | 2 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 4 | 3 | 0 |

| | **Participant 19** | **Participant 20** | **Participant 21** |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | Percentage of successful treatment | likelihood of survival rate, and the effectiveness of the treatment options | Good prognosis from chemo treatment; Assurance that hospital staff was "on my side" and would move quickly to carry out treatment if I choose to go through with it; Suggestion that I might still be able to work and carry out some normal activities during treatment regimen |
| **In your opinion, was there any negative information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | I could die from this disease | neither treatment options seem very pleasant, the 20yr survival rate is way lower than the 5yr | Learning about fatality of the disease; mentioning some percentage of deaths that occur after 5 years (even as low as 10% was still concerning to hear) |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | NO | NO | NO |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | | | |
| **How well do you feel you understand the content of the conversation?** | 4 | 3 | 2 |
| **In your opinion, the overall the content of the conversation was:** | 2 | 3 | 1 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 4 | 3 | 2 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 4 | 0 | 4 |

| | **Participant 22** | **Participant 23** | **Participant 24** |
|---|---|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | The treatment options were immediately and clearly explained to me. The treatment outcomes were also explained to me clearly without prompting, and my questions were also answered clearly. Overall, I appreciated that all the necessary information about treatment options, outcomes and side effects was provided to me even though I didn't feel fully ready/prepared to ask questions due to the shock/bewilderness I felt at receiving this diagnosis, though fictional. | 90% Survival after 5 years | Approaches to try to cure available; info of successful rates provided; some side effects advised, and the process is introduced (frequency and length each treatment, etc) |
| **In your opinion, was there any negative information from the conversation?** | YES | YES | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | Mentioning the point about insurance felt a little unpleasant, and I didn't really want to think about that dimension of the issue at the time. However, I can see how this might be necessary. | Cancer diagnosis, length of treatment. | death rates (1 minus each success rate); insurance may not cover; will die anyway no matter whether could live for a few years. |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | NO | YES | YES |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | | Types of care offered. | a lot people got cancel. |
| **How well do you feel you understand the content of the conversation?** | 3 | 4 | 4 |
| **In your opinion, the overall the content of the conversation was:** | 3 | 2 | 2 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 3 | 2 | 3 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 2 | 3 | 4 |

| | **Participant 25** |
|---|---|
| **In your opinion, was there any positive information from the conversation?** | YES |
| **If yes, please briefly describe each piece of positive information. (Bullet points are fine)** | The overall tone was polite and intoned some concern like when inquiring after my pet or if someone had come with me to the appointment. |
| **In your opinion, was there any negative information from the conversation?** | YES |
| **If yes, please briefly describe each piece of negative information. (Bullet points are fine)** | The negative information were instances when I asked a question and the doctor said they weren't sure on topics such as different types of medication. Another example of negative was the frequent mention of payment, insurance and other paperwork jargin. It is an important discussion to have but finding out you have cancer is not the time. I felt like I was treated as a patient with a number, courteous but not treated as a specific individual. |
| **In your opinion, was there other important but neutral information that was shared in the conversation earlier today?** | YES |
| **If yes, please briefly describe each piece of important but neutral information. (Bullet points are fine)** | Talking about options and next steps was comforting but once again it was done with emotional distance that left it feeling like we were talking about basic errands or a mundane topic like grocery shopping or a hair appointment. |
| **How well do you feel you understand the content of the conversation?** | 3 |
| **In your opinion, the overall the content of the conversation was:** | 2 |
| **Based on the conversation, how would you rate your optimism about your treatment options?** | 3 |
| **Based on this conversation, how confident would you feel about making a decision about your treatment?** | 2 |

| | **Participant 1** | **Participant 2** | **Participant 3** |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 15 percent | 10% | 50% |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | survival rate with the treatment is over 50% for people who survived past five years, and 65% for 20 year survival | I have cancer called betacyte | The doctor informed me that I will not have to live with this illness forever. |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Positive | Neutral | Neutral |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | treatment survival rate, the disease has been studied for many years, | this type of caner is widely studied, has a high successful treatment rate | - The illness is widely studied and has treatments<br>- The illness is treatable and has a high success rate |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | - I have a cancer<br>- the cancer is fairly aggressive, stage 2<br>- stage 2 means 5%<br>- side effects | I have cancer that if left untreated, is fatal. Aggressive stage II. With possible reoccurrence even after treatment. | - I have cancer<br>- The form of cancer I have is fairly aggressive |
| **Based on your understanding, the overall content of the conversation was:** | 1 | 2 | 2 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 2 | 2 | 3 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 3 | 3 | 1 |

| | **Participant 4** | **Participant 5** | **Participant 6** |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 8% | 30 | 10% |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | Treatments are good! | t's good that we have noticed | 90% survival rate at 5 years; 65% survival rate at 20 years; 35% at 20 years have recurring cancer of some kind |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Negative | Neutral | Positive |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | Treatments are looking to be effective for my condition. | -treatable<br>-stage 2<br>-we caught it early<br>-a lot of studies show a lot of succesfull | several treatments; high survival rate; most insurance covers treatment; chemotherapy and targeted antibodies are equally effective |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | I have 10% chance that it may not be treated. | -A lot of people lived beyond 5 years<br>-metastasize<br>-it can come back<br>-critical level<br>-fairly aggressive | betacytes at critical level now; stage 2 cancer; drastic rise from last lab results |
| **Based on your understanding, the overall content of the conversation was:** | 3 | 2 | 3 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 3 | 2 | 4 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 3 | 0 | 3 |

| | Participant 7 | Participant 8 | Participant 9 |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 10% | 12% | 90 |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | Betacyte carcinoma is widely studied | The two treatments are very successful, the five year survival rate with treatment is 90%. | Betacyte carcinoma is widely studied and there are a lot of options for treatment. |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Neutral | Neutral | Neutral |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | The cancer is early<br>The cancer is widely studied, good treatment success rate<br>Other patients are able to work despite having treatments<br>Insurance companies will cover it<br>This cancer is not genetic and the patient can still have children after treatment | My particular disease is widely studied, and the 5 year survival rate is quite good. | Stage 2 cancer has high survival rates<br>There are treatment options available |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | The patient has cancer<br>The cancer treatment will make the patient feel unwell | the betocyte cells are at a critical level, and chemotherapy presents bad side effects like hair loss, and nausea, and fatigue. | High levels of betacytes in system indicating stage 2 cancer (aggresive)<br>Side effects of treatment can bring down quality of life |
| **Based on your understanding, the overall content of the conversation was:** | 3 | 2 | 1 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 4 | 4 | 3 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 1 | 3 | 3 |

| | Participant 10 | Participant 11 | Participant 12 |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 8% | between 10-20% | less than 1/8th |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | That they have four different drugs available | Talking about my treatment options and how chemotherapy is the most successful. | the survival rates (with treatment is 90%) |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Neutral | Neutral | Positive |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | Widely studied with lots of information and treatment options; four different drugs available; 5-year survival rate is 90% | - talking about my cancer, and how it fairly common, and widely studied.<br>- Chemotherapy<br>- Targeted antibodies<br>- Success rates | treatment options, fact that it's not genetic, caught at stage 2, treatment stats |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | Betacytes have reached critical level and I have betacyte carcinoma; that it's aggressive and I'm at stage 2; that the 20-year survival rates are 65% with treatment; side effects from more successful treatment (chemotherapy) are more severe | - blood panels coming back higher than normal meaning I have cancer<br>-that my cancer is very aggressive | the cancer, it's aggressiveness, how it might spread to other parts of the body |
| **Based on your understanding, the overall content of the conversation was:** | 1 | 3 | 1 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 2 | 4 | 3 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 2 | 4 | 3 |

| | **Participant 13** | **Participant 14** | **Participant 15** |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 1/12 | 10% | 12.50% |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | The cancer is treatable with potential for a good outlook. | The doctors were able to detect my betacyte carcinoma fairly early, i.e. in Stage 2. This is quite good because I have a higher probability of treating it successfully. | The 5 year survival rate is 90%, and the 20 year survival rate is 65% |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Positive | Negative | Positive |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | The cancer was caught early<br>The cancer is treatable<br>The cancer is well studied, common, there are treatment options, and reported good outcomes<br>life should return to normal after treatment<br>cancer is treatable at stage 2 with good outcomes | 1) This is a fairly common type of cancer.<br>2) They were able to detect it at an early stage, which implies a higher probability of successful treatment.<br>3) This is not genetic, so I don't need to worry about passing it on to my kids.<br>4) There is a high likelihood of no relapse for many years after treatment.<br>5) There are multiple treatment options. | -The cancer I have is widely studies so there are many well-known treatment options<br>-The survival rate for 5 years is 90%<br>-Many people live cancer free for many years after treatment |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | betacytes have increased -> cancer<br>cancer is agressive<br>cancer jumped from stage 0 to stage 2, skipping a step | 1) I have betacyte carcinoma.<br>2) The cancer is at Stage 2, and I need to start thinking about immediate treatment.<br>3) I still have a chance of developing other cancers in my lifetime. | -The number of betacytes has increased and I now have cancer<br>-The cancer is Stage 2<br>-The level of betacytes in my blood is up to 5%<br>-If the cancer isn't treated it's fatal |
| **Based on your understanding, the overall content of the conversation was:** | 3 | 3 | 1 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 3 | 3 | 2 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 1 | 2 | 3 |

| | Participant 16 | Participant 17 | Participant 18 |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 11% | about 6 percent | 10 |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | High 5-year survival rate. | the survival rate is 60% for 5 years | The drugs at the hospital I am using are as good as anywhere else, but I am welcome to get a second opinon. |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Positive | Positive | Neutral |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | - it is a common cancer, so it well-researched and there are a lot of successful treatments available<br>- it has a 90% 5-year survival rate<br>- it is not genetic and cannot be passed on | there is a lot of treatment because its a common cancer. the five year survival rate is 90%. chemo is an effective treatment. | The cancer I have is common and has a relatively high survival rate.<br>The drugs at my hospital are just as good as drugs anywhere else. |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | - I'd been diagnosed with stage 2 betacyte carcinoma<br>- There is a lower 20-year survival rate for this type of cancer | I have cancer that is in an advanced stage | I have cancer.<br>Cancer is aggressive (stage 2) |
| **Based on your understanding, the overall content of the conversation was:** | 3 | 1 | 2 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 3 | 3 | 2 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 4 | 4 | 2 |

| | Participant 19 | Participant 20 | Participant 21 |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 20 | 18% | 12 |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | survival rate is high | The treatment will work and I'll have a normal life again | Good prognosis- statistics were given for five and twenty year survival rates |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Positive | Neutral | Neutral |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | This is a common disease and is widely studied, survival rate is high (over 90%), this cancer is non genetic so it won't be inherited to my kids | High survival rates the treatments will work | This type of cancer is quite common and there has been a lot of research on it. The chemotherapy for this type of cancer has generally good prognosis. |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | Betacyte count increased, I have stage 2 cancer, without treatment I will die, the treatment is painful, without treatment I will die within a year | quick progression of the cancer, negative side effects of treatment | The cell mutation has progressed to a point where cancer has been diagnosed; treatment is required to reduce the risk of fatality; treatment will have many unpleasant side effects |
| **Based on your understanding, the overall content of the conversation was:** | 1 | 1 | 2 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 3 | 3 | 3 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 4 | 1 | 4 |

| | Participant 22 | Participant 23 | Participant 24 |
|---|---|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | A little over 1/8th or 12-15% of the conversation is the reported as negative information. | 1/8 | 50% |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | The second piece of positive information is about the five year survival rate for betacyte carcinoma, which is over 90% with treatment. | The 90% survival rate after 5 years | The successful rates are fairly good with the treatments |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Neutral | Negative | Neutral |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | - Betacyte carcinoma is widely studied, common and multiple treatment options of good success are available<br>- The five year survival rate is over 90%, which is good<br>- Having betacyte carcinoma does not increase your chances of getting another cancer. It is also not genetic. | Betacyte Carcinoma is fairly common with lots of studies, high survival rate with treatment at stage 2, in the range where treatment will be affective, it isn't from genetics, expectation to return to a normal life | 1. that type cancel is widely studied; 2. the treatments' success rates are pretty good; 3. that type cancel is not genetic. |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | - Patients ("my") lab results show increased levels of betacytes, indicating that the patient has betacyte carcinoma. The cancer is aggressive and in stage 2<br>- Treatment options will affect the immune system to some extent and the patient will have to take more precautions<br>- 20 year survival rate is only 65% due to recurring betacyte carcinoma or occurrence of other types of cancers | cancer, increased number of betacytes, stage 2 cancer acting aggressively, lower survival rate after 20 years, side effects of chemo, | 1. those betacytes increased to a critically level; 2. at stage 2 now; 3. the cancer appears to be aggressive; 4. the 20-year survival rate is somewhat low. |
| **Based on your understanding, the overall content of the conversation was:** | 2 | 3 | 2 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 3 | 4 | 3 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 3 | 3 | 4 |

| | **Participant 25** |
|---|---|
| **The interface visualizes the total percentage of positive and negative information in the conversation. Approximately what percentage of negative information does the interface report?** | 25% |
| **The transcript of the conversation can be filtered by positive or negative information. Using the filters, please briefly summarize the second piece of positive information in the conversation.** | The second positive interaction is talking about the overall positive statistics around survival rates. |
| **You can listen to the audio recording of the conversation through the interface. The audio signal is highlighted to indicate times when positive information (blue) and negative information (red) occur. All other information is neutral. At time 2:45 on the timeline, the information is reported as:** | Negative |
| **Based on your understanding, what was the positive information conveyed in the conversation? (Bullet points are fine)** | The positive facts in the conversation revolved around positive outcomes and approaches. It happens when discussing treatment options and survival rates and the different things to expect. |
| **Based on your understanding, what was the negative information conveyed in the conversation? (Bullet points are fine)** | The negative conversation revolved around deviations of the positive statistics and often was mentioning the side effects of the treatment and the outcomes if I chose not to seek treatment. |
| **Based on your understanding, the overall content of the conversation was:** | 2 |
| **After viewing the web interface, how do you feel about the treatment options discussed in the conversation?** | 3 |
| **After viewing the web interface, how confident do you feel about making a decision about your treatment?** | 3 |

| | Participant 1 | Participant 2 | Participant 3 |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 3 | 3 | 3 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 1 | 3 | 3 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | YES | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | The other 35% may experience a reoccurring cancer of another cancer- it seems like still I have a fairly great chance to have the same cancer of have other | | |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | YES | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | - explanation of side effects- because those side effects are kind of predictable effects well-known for cancer and was not that surprising, and not sound like very bad compared to the fact that I got a cancer<br>- everyone responds differently to chemotherapy- I felt like I might have a better chance to work well with the chemotherapy than other people. Maybe because I am more optimistic | | |
| **In terms of usability, how would you rate the web interface?** | 4 | 4 | 4 |
| **How helpful was the interface for finding important information in the conversation?** | 4 | 3 | 4 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 3 | 3 | 4 |
| **How would you rate your overall experience with the web interface?** | 3 | 3 | 4 |

| | Participant 4 | Participant 5 | Participant 6 |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 4 | 1 | 4 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 1 | 4 | 2 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO | YES | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | - the five year survival rate, so that means how many people live beyond five years, is ninety percent | |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | NO | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | |
| **In terms of usability, how would you rate the web interface?** | 4 | 4 | 2 |
| **How helpful was the interface for finding important information in the conversation?** | 4 | 0 | 3 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 3 | 1 | 3 |
| **How would you rate your overall experience with the web interface?** | 3 | 3 | 3 |

| | Participant 7 | Participant 8 | Participant 9 |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 4 | 3 | 4 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 3 | 2 | 3 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO | YES | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | I thought the statement that treatment does not cause further cancer down the line is maybe more neutral than positive. | |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | NO | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | |
| **In terms of usability, how would you rate the web interface?** | 4 | 4 | 4 |
| **How helpful was the interface for finding important information in the conversation?** | 4 | 3 | 4 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 4 | 4 | 3 |
| **How would you rate your overall experience with the web interface?** | 4 | 4 | 3 |

| | Participant 10 | Participant 11 | Participant 12 |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 3 | 4 | 2 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 1 | 3 | 2 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO | NO | YES |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | mostly the information around survival statistics, the language i interpreted as negative bc it meant there was some likelihood of death/no cure |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | YES | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | I hadn't been sure if Stage 2 was a negative point or not, but the transcript/web interface make it more obvious that it is | | |
| **In terms of usability, how would you rate the web interface?** | 4 | 4 | 4 |
| **How helpful was the interface for finding important information in the conversation?** | 3 | 4 | 3 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 3 | 4 | 2 |
| **How would you rate your overall experience with the web interface?** | 4 | 4 | 3 |

| | **Participant 13** | **Participant 14** | **Participant 15** |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 4 | 4 | 2 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 3 | 0 | 1 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO | NO | YES |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | -I don't think the 20 year survival rate being 65% is very positive considering I'm only 20 |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | NO | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | |
| **In terms of usability, how would you rate the web interface?** | 4 | 4 | 3 |
| **How helpful was the interface for finding important information in the conversation?** | 4 | 4 | 3 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 4 | 4 | 1 |
| **How would you rate your overall experience with the web interface?** | 4 | 4 | 3 |

| | Participant 16 | Participant 17 | Participant 18 |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 4 | 2 | 3 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 2 | 3 | 1 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO | | YES |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | The betacyte count being at 5 percent instead of 3 percent now. I don't completely understand the medical details, and it also sounds like it should be bad that my count went up because thats why I have cancer. | A 65% 25-year survival rate is concerning |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | NO | YES | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | the 20-year survival rate being at 65%. in my opinion, hearing a cancer diagnosis is really scary and makes you feel like you're going to die, so even 65% survival sounded good to me. | |
| **In terms of usability, how would you rate the web interface?** | 3 | 4 | 4 |
| **How helpful was the interface for finding important information in the conversation?** | 4 | 4 | 4 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 4 | 3 | 2 |
| **How would you rate your overall experience with the web interface?** | 4 | 3 | 4 |

| | Participant 19 | Participant 20 | Participant 21 |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 4 | 4 | 3 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 3 | 3 | 3 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO | NO | YES |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | Hearing that there is only a 65% 20-year survival rate was frightening. I internalized that as "there is an almost 50/50 chance you only have 20 years to live", which is especially scary to hear as someone in his 20s |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | NO | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | |
| **In terms of usability, how would you rate the web interface?** | 4 | 4 | 3 |
| **How helpful was the interface for finding important information in the conversation?** | 4 | 3 | 4 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 4 | 2 | 3 |
| **How would you rate your overall experience with the web interface?** | 4 | 3 | 3 |

| | Participant 22 | Participant 23 | Participant 24 |
|---|---|---|---|
| **How well do you feel you understand the content of the conversation?** | 4 | 4 | 4 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 2 | 3 | 1 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO | YES | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | Where treatment becomes less effective, but the sentence after was positive, maybe the entire paragraph should not be rated as positive or not as a whole, but sentence by sentence. | |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | NO | NO | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | | | |
| **In terms of usability, how would you rate the web interface?** | 4 | 3 | 4 |
| **How helpful was the interface for finding important information in the conversation?** | 3 | 3 | 4 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 3 | 4 | 4 |
| **How would you rate your overall experience with the web interface?** | 3 | 3 | 4 |

| | **Participant 25** |
| --- | --- |
| **How well do you feel you understand the content of the conversation?** | 2 |
| **How much do you think the information represented in the web interface influenced your opinion of the conversation?** | 2 |
| **Was there information that the web interface highlighted as positive that you thought was NOT positive?** | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | |
| **Was there information that the web interface marked as negative that you thought was NOT negative?** | NO |
| **If you answered YES to the previous question, what was that information and why did you disagree? (Bullet points are fine)** | |
| **In terms of usability, how would you rate the web interface?** | 2 |
| **How helpful was the interface for finding important information in the conversation?** | 2 |
| **In terms of understanding the conversation, how helpful were the marked positive and negative information in the web interface?** | 3 |
| **How would you rate your overall experience with the web interface?** | 3 |

| | Participant 1 | Participant 2 | Participant 3 |
|---|---|---|---|
| **What do you think worked well in the interface?** | To see the contents in summary in terms of negative or positive, and search for that part by click the time line of the conversation recording | The content breakdown was nice to see an overall picture. I liked the audio track as it can tell how the conversation progressed (e.g. ended on mostly positive notes). | The simplicity and how intuitive the design was. You can easily figure out what the buttons do and how to operate the tools without someone needing to explain it to you. |
| **What do you think did NOT work well in the interface?** | I was wondering how it defines the positive and negative contents because some part I felt it has different view of mine, this wonder made me to also look up the neutral conversation part just to check if there's any missing contents that I felt positive/negative. | I think it would have been more empowering to be able to label the positive and negative sentences rather than being told that it was a positive or negative sentence. Most likely I would have labeled it as it was labelled. However, the act of labelling would have helped me understand the sense of the entire conversation. | It's hard to get exact timestamps because nothing appears when you hover over the timeline. |
| **Do you have comments or suggestions to improve the web interface?** | At the timeline of the conversation, the duration of the negative and positive part (color bar length) made me confused when I had to figure out the percentage of negative contents. I almost found that conversation length can represent the percentage of negative/positive contents but was not. | | N/A |

| | Participant 4 | Participant 5 | Participant 6 |
|---|---|---|---|
| **What do you think worked well in the interface?** | Overall UI | It is easy to navigate through the conversation | Identification of positive vs negative info in the transcript was good. Highlighting one over the other was helpful in focusing on one aspect at a time. |
| **What do you think did NOT work well in the interface?** | Nothing in particular. | The most important part of the conversation is what is neutral, that show the treatments etc The positive and bad things about the conversation is what the patient would get from the conversation not the important things that are the neutral ones. I mean the patient would leave knowing that his cancer is in a realluy good or bad stage and is curable or not not how | There were some details about what happens during/after the therapy that were neutral so not highlighted but would be important to review when making a decision, like the radiation therapy coming in as an additional step sometimes or the length of treatments. Did not get a good sense of overall conversation being positive or negative with the use of a pie chart perhaps because so much is neutral that it washed out the difference in the blue and orange. |
| **Do you have comments or suggestions to improve the web interface?** | Are there only two categories for determining the context of the conversation? positive vs negative? | highlight procedures and link it with webpage that the doctors would reccomend to look at. Otherwise the patient would look at doctor google. Kind of conceting with bibliography | It would be nice to have a scrollbar for the transcript to make skimming easier (cannot drag the time bar along the bottom for fast, smooth skimming). For those that are color blind, you should change your color scheme. The transcript did not update location when I was just listening; I had to click on the timeline to get the text to shift. |

| | **Participant 7** | **Participant 8** | **Participant 9** |
|---|---|---|---|
| **What do you think worked well in the interface?** | The conversation highlighting, assuming that it is able to properly capture positive and negative comments for all general conversations. It helped me remember info, such as this cancer being highly studied, that I had forgotten about. | I think the highlighted colors worked well for pointing out important parts of the conversation | The interface mostly made it quite easy to scan through positive and negative information separately |
| **What do you think did NOT work well in the interface?** | I am just slightly worried that some content might be missed if it is not given a specific label positive or negative label. | skipping through and selecting a specific spot in the convo didn't immediately show me where we were in the convo. | Some statements had both positive and negative information and had been marked as neutral by the interface |
| **Do you have comments or suggestions to improve the web interface?** | Maybe try to highlight not only positive and negative comments, but also information that the doctor views as important such as the treatment types | Maybe the segment of the convo could be highlighted in yellow if it is currently on the dial or something like that? | The interface could try and look for negative and positive sub-sections within a given block of information |

| | Participant 10 | Participant 11 | Participant 12 |
|---|---|---|---|
| **What do you think worked well in the interface?** | All of the different ways of jumping around the conversation (positive vs. negative, by time, by transcript) | it is easy to visualize everything - the coloring was helpful (blue = positive, red = negative) | Helped highlight parts of the conversation |
| **What do you think did NOT work well in the interface?** | Not being able to see the specific percentages of positive vs. negative and having to estimate the ratio | I couldn't tell what % was positive or negative. I can see it, but it does not give exact #'s. | not sure what the purpose of the interface is exactly.. is it for patients or doctors? for doctors, the interface helps them understand their patient care and delivery of information; for patients, it helps you understand what you saw as negative as a possible positive |
| **Do you have comments or suggestions to improve the web interface?** | | | easy to use |

| | Participant 13 | Participant 14 | Participant 15 |
|---|---|---|---|
| **What do you think worked well in the interface?** | I really liked the filtering ability of the interface, the highlights (also shades of colors were very well picked), and the time bar on the bottom | 1) The ability to jump to any point in the conversation by clicking on the timestamp.<br>2) The temporal variation of the sentiments shown at the bottom.<br>3) The ability to filter out the positive and negative comments separately. | The whole left side of the screen is pretty good. The sliding text and bottom scroller work really well. |
| **What do you think did NOT work well in the interface?** | I wish I could tell the time point exactly in the transcript versus having to estimate the time based on the time points below the bar | | |
| **Do you have comments or suggestions to improve the web interface?** | Add a tracing point for the time<br>I would also provide additional information resources somewhere so patients could also do their own research if they want more information | 1) It would be nice to see the percentage shares pop up when the cursor hovers over the pie chart.<br>2) If I had a really long conversation, I'd have to keep scrolling up and down to review all the information. It would be nice to open a new window with only the negative or positive comments when I click on the respective button (instead of the current highlighting implementation). | -The + and - buttons are confusing because they usually indicate increasing or decreasing a sliding scale of something, not choosing one or the other<br>-When I clicked the - sign for example, I expected pressing play to start playing at the first negative block, not just continue with where I was. |

| | Participant 16 | Participant 17 | Participant 18 |
|---|---|---|---|
| **What do you think worked well in the interface?** | It really did a good job labeling the positive and negative information. Also, having the important information marked as positive or negative made it easier to find the highlights of the conversation. I also think it had the right number of features. The pie chart, transcript, filters, and audio position all contributed to a positive experience and were easy to use. | Having both the text and timeline highlighted blue or red is nice. Also it has very clean designs and fonts. | Design of interface<br>Multiple ways to traverse conversation |
| **What do you think did NOT work well in the interface?** | No, just a few comments for improvement below. | I don't think that the percentage of positive and negative information in the pie chart is extremely useful. Not to mention, most of the conversation was neutral anyway. | The majority of the conversation (including important information about treatment options, description of the disease, etc) were left out<br>The purpose of using it isn't very clear (i.e. what am I gaining from looking at this?) |
| **Do you have comments or suggestions to improve the web interface?** | -I would have liked percentages on the pie chart.<br>-I wish you could drag the time marker instead of just clicking it along with an indicator of what time you are currently on. It'd be easier to remember at exactly what time important information was given that way.<br>-The time labels at the bottom are a little cluttered and took me a minute to figure out what was going on. Perhaps labels just every minute? I think having an indicator for what the position of the cursor (as mentioned above) would probably alleviate this as well.<br>- Being able to toggle on and off positive and negative information independently would have been more intuitive than the three buttons provided. Ideally both could be turned on at once to see all information that was not neutral at once. | the "+" and "-" buttons are misleading. because it is convention to use plus/minus buttons on apps/websites when you're adding something to a list or increasing the quantity buying a product online, it is confusing that the same buttons are used in a completely separate way. I also think that this website emphasizes the good/bad info, but there is a lot of important logistical information in the neutral section too that a user may overlook, so there could be some way to extract that info (medicine, etc). | Clicking on the red, blue, and white parts of the graph should have the same functionality as the filter content buttons |

| | Participant 19 | Participant 20 | Participant 21 |
|---|---|---|---|
| **What do you think worked well in the interface?** | color coding in the transcript matches the audio | Dividing the "big" positive and negative comments well. | The coloring scheme seemed to focus on the most important information, which correlated to the positive and negative information |
| **What do you think did NOT work well in the interface?** | | There's information that the positive/negative division didn't catch. For example - a negative for me was the re-occurrence of the cancer in 20yr survival rate, but the interface didn't classify that. | |
| **Do you have comments or suggestions to improve the web interface?** | If there is a borderline negative or a borderline positive comment, would that be marked as neutral information? However, I understand that the black or white nature of the positive/negative content is useful to maintain usability and clarity, and there might be a problem in differentiating these borderline negative/positive as gray zones. | I'm not sure if your interface is using an algorithm to classify positive or negatives. But if you're just looking for "key phrases" I don't think that is capturing the variety of nuance within a conversation. | |

| | Participant 22 | Participant 23 | Participant 24 |
|---|---|---|---|
| **What do you think worked well in the interface?** | I though the categorization of positive and negatives was remarkably accurate, and distilled the core of the conversation pretty well. The pie chart and the labeling on the audio tracks were helpful visualizations, and noticing that the negative parts of the conversation were only a little bit more than the positive parts helped me feel better about the conversation and diagnosis as whole. | The positive and negative highlights - the interface was also easy to use and understand. | the breakdown pie and the filter feature |
| **What do you think did NOT work well in the interface?** | Large parts of the conversation were labelled as useful, as they were mostly conveying information/answers to questions. If I had scrolled to a neutral portion of the transcript and used one of the filters, nothing changed in my view of the transcript and that was not very useful. The white color/label for "neutral" parts of the conversation was also a little confusing, because it made me feel like none of that information was important, when it actually is. | Locating the specific time was not so clear. | |
| **Do you have comments or suggestions to improve the web interface?** | I think it could be helpful if when using the filters, the transcript jumped to the first portion of the conversation in the category I want to see. Also, the labeling of the audio track is useful and I did listen to the audio fully when at first, but I didn't really use it much when answering the questions. The transcript was more useful to me for thinking about the conversation. | Perhaps time pointers can be placed in the text of the conversation as well. | |

| | **Participant 25** |
| --- | --- |
| **What do you think worked well in the interface?** | I think the interface was an easy breakdown of our conversation and helpfully broke it up into more digestible and easily processed segments. |
| **What do you think did NOT work well in the interface?** | The interface was a little difficult to navigate if you were looking for specific information such as a time in the conversation or when the doctor mentions insurance or treatment options. |
| **Do you have comments or suggestions to improve the web interface?** | Not at this time. |

# Bibliography

[1] M. J. Gonzales and L. D. Riek, “Co-designing Patient-centered Health Communication Tools for Cancer Care,” in Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare, ICST, Brussels, Belgium, 2013, pp. 208-215.

[2] S. R. Mishra et al., “Not Just a Receiver: Understanding Patient Behavior in the Hospital Environment,” in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2016, pp. 3103-3114.

[3] K. T. Unruh, M. Skeels, A. Civan-Hartzler, and W. Pratt, “Transforming clinic environments into information workspaces for patients,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2010, pp. 183-192.

[4] B. Shneiderman, C. Plaisant, and B. W. Hesse, “Improving Healthcare with Interactive Visualization,” Computer, vol. 46, no. 5, pp. 58-66, May 2013.

[5] “Types of Cancer Treatment,” National Cancer Institute. [Online]. Available: https://www.cancer.gov/about-cancer/treatment/types. [Accessed: 30-Jul-2018].

[6] “Side Effects,” National Cancer Institute. [Online]. Available: https://www.cancer.gov/about-cancer/treatment/side-effects. [Accessed: 30-Jul-2018].

[7] J. Freyne, “Mobile health: beyond consumer apps,” MobileHCI 2014, pp. 565-566.

[8] A. Das, A. Faxvaag, and D. Svan, “Interaction design for cancer patients: do we need to take into account the effects of illness and medication?,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2011 pp. 21-24.

[9] P. Klasnja, A. Civan Hartzler, K. T. Unruh, and W. Pratt, “Blowing in the Wind: Unanchored Patient Information Work During Cancer Care,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2010, pp. 193-202.

[10] M. Bonner and E. Mynatt, "Gauging the Patient-Centered Potential of Online Health Seeking," in Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, ICST, Brussels, Belgium, 2014, pp. 33-40.

[11] Christian Dameff, MD, Brian Clay, MD, and Christopher A. Longhurst, MD, "Personal Health Records: More Promising in the Smartphone Era?," J. Am. Med. Assoc., vol. 321, no. 4, 2019.

[12] C. Cannon, "Telehealth, Mobile Applications, and Wearable Devices are Expanding Cancer Care Beyond Walls," Semin. Oncol. Nurs., vol. 34, no. 2, pp. 118-125, May 2018.

[13] J. Shen and A. Naeim, "Telehealth in older adults with cancer in the United States: The emerging use of wearable sensors," J. Geriatr. Oncol., vol. 8, no. 6, pp. 437-442, Nov. 2017.

[14] "Center for Connected Health Policy." [Online]. Available: https://www.cchpca.org/about/about-telehealth. [Accessed: 11-Nov-2018].

[15] I. R. Yurkiewicz, P. Simon, M. Liedtke, G. Dahl, and T. Dunn, "Effect of Fitbit and iPad Wearable Technology in Health-Related Quality of Life in Adolescent and Young Adult Cancer Patients," J. Adolesc. Young Adult Oncol., Jun. 2018.

[16] R. M. Aileni, S. Pasca, C. A. Valderrama, and R. Strungaru, "Wearable health care: technology evolution, algorithms and needs," in Enhanced Living Environments: From models to technologies, R. I. Goleva, I. Ganchev, C. Dobre, N. Garcia, and C. Valderrama, Eds. Institution of Engineering and Technology, 2017, pp. 315-343.

[17] M. Craft et al., "An assessment of visualization tools for patient monitoring and medical decision making," in 2015 Systems and Information Engineering Design Symposium, 2015, pp. 212-217.

[18] D. Gotz and D. Borland, "Data-Driven Healthcare: Challenges and Opportunities for Interactive Visualization," IEEE Comput. Graph. Appl., vol. 36, no. 3, pp. 90-96, May 2016.

[19] J. H. Bettencourt-Silva, G. S. Mannu, and B. de la Iglesia, "Visualisation of Integrated Patient-Centric Data as Pathways: Enhancing Electronic Medical Records in Clinical Practice," in Machine Learning for Health Informatics, vol. 9605, A. Holzinger, Ed. Cham: Springer International Publishing, 2016, pp. 99-124.

[20] P. Klasnja, A. Hartzler, C. Powell, G. Phan, and W. Pratt, "HealthWeaver Mobile: Designing a Mobile Tool for Managing Personal Health Information during Cancer Care," AMIA Annu Symp Proc, vol. 2010, pp. 392-396, 2010.

[21] M. Jacobs, J. Clawson, and E. D. Mynatt, “Cancer navigation: opportunities and challenges for facilitating the breast cancer journey,” in Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing, 2014, pp. 1467-1478.

[22] J. Frost, N. Beekers, B. Hengst, and R. Vendeloo, “Meeting cancer patient needs: designing a patient platform,” in CHI ’12 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, 2012, pp. 2381-2386.

[23] S. Farnham et al., “HutchWorld: Clinical Study of Computer-mediated Social Support for Cancer Patients and Their Caregivers,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2002, pp. 375-382.

[24] P. A. Mueller and D. M. Oppenheimer, “Technology and note-taking in the classroom, boardroom, hospital room, and courtroom,” Trends Neurosci. Educ., vol. 5, no. 3, pp. 139-145, Sep. 2016.

[25] S. Chen, “A Pen-Eye-Voice Approach Towards The Process of Note-Taking and Consecutive Interpreting: An Experimental Design,” Int. J. Comp. Lit. Transl. Stud., vol. 6, no. 2, pp. 1-8, Apr. 2018.

[26] “The 10 Best Note Taking Apps in 2018: Evernote, OneNote, and Beyond.” [Online]. Available: https://zapier.com/blog/best-note-taking-apps/. [Accessed: 19-Apr-2019].

[27] K. Kim, S. A. Turner, and M. A. Prez-Quiones, “Requirements for electronic note taking systems: A field study of note taking in university classrooms,” Educ. Inf. Technol., vol. 14, no. 3, pp. 255-283, Sep. 2009.

[28] C. Goudeseune, “Effective browsing of long audio recordings,” in Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices - IMMPD 12, Nara, Japan, 2012, pp. 35-41.

[29] H.-Y. Lee, P.-H. Chung, Y.-C. Wu, T.-H. Lin, and T.-H. Wen, “Interactive Spoken Content Retrieval by Deep Reinforcement Learning,” IEEEACM Trans. Audio Speech Lang. Process., vol. 26, no. 12, pp. 2447-2459, Dec. 2018.

[30] A. Pavel, D. B. Goldman, B. Hartmann, and M. Agrawala, “SceneSkim: Searching and Browsing Movies Using Synchronized Captions, Scripts and Plot Summaries,” in Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST 15, Daegu, Kyungpook, Republic of Korea, 2015, pp. 181-190.

[31] A. Pavel, C. Reed, B. Hartmann, and M. Agrawala, “Video digests: a browsable, skimmable format for informational lecture videos,” in Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST 14, Honolulu, Hawaii, USA, 2014, pp. 573-582.

[32] M. Fleischman and D. Roy, "Grounded Language Modeling for Automatic Speech Recognition of Sports Video," in Proceedings of ACL-08: HLT, Columbus, Ohio, 2008, pp. 121-129.

[33] R. Gidwani et al., "Impact of Scribes on Physician Satisfaction, Patient Satisfaction, and Charting Efficiency: A Randomized Controlled Trial," Ann. Fam. Med., vol. 15, no. 5, pp. 427-433, Sep. 2017.

[34] K. Hafner, "A Busy Doctors Right Hand, Ever Ready to Type," The New York Times, 19-Jan-2018.

[35] J. Kincaid, "Which Automatic Transcription Service is the Most Accurate? 2018," Medium, 05-Sep-2018.

[36] A. Mason, "Comparing the Accuracy of Automatic Transcription Services," Medium, 12-Dec-2017.

[37] Y. Berzak, A. Barbu, D. Harari, B. Katz, and S. Ullman, "Do You See What I Mean? Visual Resolution of Linguistic Ambiguities," ArXiv160308079 Cs, Mar. 2016.

[38] A. de L. Fernandes, "CALTRANSCENSE: A REAL-TIME SPEAKER IDENTIFICATION SYSTEM," M.E. thesis, Univ. of California, Berkeley, USA, 2015.

[39] H. Su, "Combining Speech and Speaker Recognition - A Joint Modeling Approach," Ph.D. dissertation, Univ. of California, Berkeley, USA, 2018.

[40] B. Gur, "Improving Speech Recognition Accuracy for Clinical Conversations," M.E. thesis, Massachusetts Inst. of Technology, Cambridge, MA, USA, 2012.

[41] C.-C. Chiu et al., "Speech recognition for medical conversations," ArXiv171107274 Cs Eess Stat, Nov. 2017.

[42] A. Morales, V. Premtoon, C. Avery, S. Felshin, and B. Katz, "Learning to Answer Questions from Wikipedia Infoboxes," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, 2016, pp. 1930-1935.

[43] S. C. Bandler, "Interpreting Author Intentions by Analyzing Story Modulation," M.E. thesis, Massachusetts Inst. of Technology, Cambridge, MA, USA, 2019.

[44] B. Tang, Y. Wu, M. Jiang, Y. Chen, J. C. Denny, and H. Xu, "A hybrid system for temporal information extraction from clinical text," J. Am. Med. Inform. Assoc. JAMIA, vol. 20, no. 5, pp. 828-835, Sep. 2013.

[45] H. Harkema, J. N. Dowling, T. Thornblade, and W. W. Chapman, "Context: An Algorithm for Determining Negation, Experiencer, and Temporal Status from Clinical Reports," J. Biomed. Inform., vol. 42, no. 5, pp. 839-851, Oct. 2009.

[46] "MedLee — MedLingMap." [Online]. Available: http://www.medlingmap.org/node/69. [Accessed: 20-Apr-2019].

[47] K. Dinakar, J. Chen, H. Lieberman, R. Picard, and R. Filbin, "Mixed-Initiative Real-Time Topic Modeling & Visualization for Crisis Counseling," in Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI 15, Atlanta, Georgia, USA, 2015, pp. 417-426.

[48] A. W. Forsyth, "Improving Clinical Decision Making With Natural Language Processing And Machine Learning," M.E. thesis, Massachusetts Inst. of Technology, Cambridge, MA, USA, 2017.

[49] I. Chien, "Natural Language Processing for Precision Clinical Diagnostics and Treatment," M.E. thesis, Massachusetts Inst. of Technology, Cambridge, MA, USA, 2018.

[50] R. Picard, Affective Computing. The MIT Press, 2000.

[51] Y. Utsumi, J. Lee, J. Hernandez, E. C. Ferrer, B. Schuller, and R. W. Picard, "CultureNet: A Deep Learning Approach for Engagement Intensity Estimation from Face Images of Children with Autism," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 339-346. IEEE, 2018.

[52] J. M. Garcia-Garcia, V. M. R. Penichet, M. D. Lozano, J. E. Garrido, and E. L.-C. Law, "Multimodal Affective Computing to Enhance the User Experience of Educational Software Applications," Mob. Inf. Syst., vol. 2018, pp. 1-10, Sep. 2018.

[53] B. Schuller and A. Batliner, Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. United Kingdom: Wiley & Sons, Ltd., 2014.

[54] Mark Knapp, Judith Hall, and Terrence Horgan, "The Effects of Vocal Cues That Accompany Spoken Words," in Nonverbal Communication in Human Interaction, 8th ed., Monica Eckman, 2014, pp. 323-356.

[55] Taniya Mishra and Dimitrios Dimitriadis, "Incremental Emotion Recognition," INTERSPEECH, 2013.

[56] B. W. Schuller, "Speech Emotion Recognition: Two Decades in a Nutshell, Benchmarks, and Ongoing Trends," Commun. ACM, vol. 61, no. 5, pp. 90-99, May 2018.

[57] S. Yoon, S. Byun, and K. Jung, "Multimodal Speech Emotion Recognition Using Audio and Text," ArXiv181004635 Cs, Oct. 2018.

[58] Diagnostic and statistical manual of mental disorders, 5th ed. Washington, D.C.: American Psychiatric Association, 2013.

[59] A. Benba, A. Jilbab, and A. Hammouch, "Detecting multiple system atrophy, Parkinson and other neurological disorders using voice analysis," Int. J. Speech Technol., vol. 20, no. 2, pp. 281-288, Jun. 2017.

[60] B. Schuller et al., "Being bored? Recognising natural interest by extensive audiovisual integration for real-life application," Image Vis. Comput., vol. 27, no. 12, pp. 1760-1774, Nov. 2009.

[61] W. F. Baile, "SPIKES–A Six-Step Protocol for Delivering Bad News: Application to the Patient with Cancer," The Oncologist, vol. 5, no. 4, pp. 302-311, Aug. 2000.

[62] "Watson Natural Language Understanding," 28-Nov-2016. [Online]. Available: https://www.ibm.com/watson/services/natural-language-understanding/. [Accessed: 07-May-2019].

[63] "Natural Language Toolkit  NLTK 3.4.1 documentation." [Online]. Available: https://www.nltk.org/. [Accessed: 07-May-2019].

[64] "Affectiva Automotive AI," Affectiva. [Online]. Available: https://www.affectiva.com/product/affectiva-automotive-ai/. [Accessed: 07-May-2019].

[65] "SciPy  SciPy v1.2.1 Reference Guide." [Online]. Available: https://docs.scipy.org/doc/scipy/reference/index.html. [Accessed: 07-May-2019].

[66] "scikit-learn: machine learning in Python  scikit-learn 0.20.3 documentation." [Online]. Available: https://scikit-learn.org/stable/. [Accessed: 07-May-2019].

[67] CCCN, "Giving Bad News pt1.mov," Youtube. Jul 8, 2018. Available: https://www.youtube.com/watch?v=oMaTcGjOPsU.

[68] Jared Dichiara, "OB/GYN," Youtube. Apr 21, 2014. Available: https://www.youtube.com/watch?v=LK2SpVkWhiw.

[69] Jared Dichiara, "Delivering Bad News," Youtube. Jul 7, 2017. Available: https://www.youtube.com/watch?v=HEMc259fF$_0$.

[70] Jared Dichiara, "FM," Youtube. Apr 21, 2014. Available: https://www.youtube.com/watch?v=A2Bt0T9TXyk.

[71] Jared Dichiara, "Disclosing Medical Errors," Youtube. Jul 7, 2017. Available: https://www.youtube.com/watch?v=51ebFM7CPWQ.