ABSTRACT

Title of thesis: AUGMENTATIVE AND ALTERNATIVE COMMUNICATION INTERFACE WITH A CONTEXT-BASED, PREDICTIVE LANGUAGE MODEL

Timothy Lewis, Master of Science 2018

Thesis directed by: Professor Cynthia Matuszek
Department of Computer Science
and Electrical Engineering

Communication is ubiquitous in the world. The need to describe one's needs and desires is a necessity in all aspects of life. When a person loses the ability to physically speak, and the motor skills to activate speech systems that require physical input, there are limited alternatives for communication. Eye-gaze tracking offers unique opportunities to support communication; however, many of the current eye-gaze tracking communication programs are unintuitive, inaccurate, and outdated. Many of the existing communication programs employ a visual keyboard that the user can use to visually type out the words and sentences they wish to speak. In this paper, I investigate an alternative to the standard virtual keyboard methods, developing a communicative flow that is intuitive and easy to use without speech. I create a simple application for participants to construct messages. I introduce a language model concept that gathers information about the environment and using language groundings, determines the salience of items within a scene. This, in turn,

allows the language model to make context sensitive text prediction to better assist a user in communicating quickly and effectively. Using a gaze tracking device as input to the interface, I then perform a user study to evaluate the efficacy and intuitiveness of the communication device. I found that when a user is presented with context based portions of speech, they can communicate easier and quicker than with a generic language model.

AUGMENTATIVE AND ALTERNATIVE COMMUNICATION INTERFACE WITH A CONTEXT-BASED, PREDICTIVE LANGUAGE MODEL

by

Timothy Lewis

Thesis submitted to the Faculty of the Graduate School of the University of Maryland, Baltimore County in partial fulfillment of the requirements for the degree of
Master of Science
2018

Advisory Committee:
Professor Cynthia Matuszek, Chair/Advisor
Professor Tim Finin
Professor Francis Ferraro

Acknowledgments

I would like to express my gratitude to my advisor, Dr. Cynthia Matuszek, for providing me the opportunity to perform work in the Interactive Work and Language Lab. She has always made herself available for guidance. It has been a pleasure working with and learning from her, and I look forward to extending this experience as I continue on with my academic career.

I would also like to thank the other members of my committee, Dr. Frank Ferraro and Dr. Tim Finin, for their engagement, questions, and insightful comments..

The experiments described in Chapter 4 were performed onsite at Washington State University by Alexandra Kerr, Marcus Blaisdell, and Kaniz Madhobi under the direction of Dr. Matthew E. Taylor. A paper describing this work and conclusions drawn from it, with these additional authors, will be submitted at a robotics and learning conference in Summer 2018. I would like to especially acknowledge the significant contribution to the project that these studies represent and thank the WSU team.

A special thanks to Dr. Amy Hurst at UMBC for providing great insight and direction in preliminary research.

Thank you to my friends and family who have provided encouragement and support throughout this whole process.

Last, but not least, I would like to thank my girlfriend, Rachel. She has always been proud of me for every accomplishment. She has put up with me spending late nights on the computer. She has always been there for me. Thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAC	Augmentative and alternative communication
ALS	Amyotrophic lateral sclerosis
ERICA	Eye-gaze Response Interface Computer Aid
LIS	Locked-in syndrome
MND	Motor neuron disease

Chapter 1: Introduction

1.1 Outline of Thesis

This thesis presents research that has been performed on developing a communication interface for users to control via gaze tracking as an input. My goal is to develop a communication interface that allows a user to communicate easier, quicker, and more effectively than existing solutions. The concept for the supporting language model is what sets this communication interface apart from many existing interfaces. By collecting information about a user's environment, I aim to develop language groundings to determine salience within a given scene. With this information, the language model can produce context driven text prediction that will be more applicable to a given situation.

For this particular research, the interface presents users with text prediction options using a simulated grounded language model as a corpus. There are four different methods used to populate the text suggestions presented in the interface, outlined in Chapter 3. By employing the use of a gaze tracking device, I am able to better simulate the user experience of someone who may not have the ability to interact with such an interface otherwise. An interactive user study was performed, including a usability survey, and results were collected and analyzed.

This chapter provides an overview of several of the technical concepts relied upon in this research. Chapter 2 discusses some existing technologies and related research to this topic of study. Chapter 3 explains what I expected to find from this research and outlines my hypotheses for this research. Chapter 4 details the actions I took in designing and developing the communication interface, my goals for the supporting language model, and the user study that was performed. Chapter 5 presents the results and observations from the user study. Chapter 6 discusses conclusions I drew from the research and user study. Chapter 6 also discusses planned and potential future work on related topics and other domains that this research may be applied to in the future.

1.2 Locked-in Syndrome

Locked-in Syndrome (LIS) is a condition resulting in the paralysis of nearly all voluntary muscles except for the muscles that can control the movement of the eyes [1, 2]. LIS also renders patients with the inability to speak, allowing them to only communicate via eye movements. Patients with LIS remain conscious and awake, and cognitive function is typically unaffected. There are several causes that can lead to LIS and other conditions that have similar symptoms.

Motor neuron disease (MND) is a group of diseases that affects the nerves of the brain and the spinal cord [3]. MND causes messages from the brain to the muscles to gradually decrease, leading to atrophy. The most common form of MND manifests itself as amyotrophic lateral sclerosis (ALS). ALS is a a progressive neurodegenerative disease [4]. Since the disease is progressive the symptoms get worse over time [5]. Early symptoms of ALS include muscle stiffness or weakness. As the disease progresses, voluntary muscle movement can deteriorate, leading to the loss in ability to move, speak, eat, and, ultimately, breathe. Many of these symptoms are similar to those of LIS. There is currently no cure or treatment that halts or reverses ALS. Life expectancy of a person diagnosed with ALS averages from about two to five years, however, this can be highly variable, and progression of the disease is not always predictable.

Although the sequence of symptoms and the rate of disease progression is variable, individuals with ALS will eventually lose the ability to stand or walk, and user their hands and arms. This typically includes the ability to write. The disease will also usually progress to the point where vocal speech will become difficult or impossible. However, muscles in the eyes may not be affected, or may be affected only very late in the disease progression. Individuals with ALS that suffer from the symptoms will still be required to communicate their needs and desires just like every other individual.

1.3 Augmentative and Alternative Communication

The American Speech-Language-Hearing Association defines augmentative and alternative communication (AAC) as, "all forms of communication (other than oral speech) that are used to express thoughts, needs, wants, and ideas" [6]. AAC systems can typically be classified into one of two categories: unaided systems and aided systems. Both types have the same purpose of providing tools and strategies to solve everyday communication challenges [8]. Unaided systems do not require anything external to an individual's body. This include gestures, body language, facial expressions, eye gaze, and sign language [7]. Aided systems utilize other materials, equipment, tools, or devices. Aided systems can also be broken down into two sub-categories: basic and high-tech. Basic systems include pen and paper and pointing to letters or picture boards. Introducing a computer to the system will transition it to a high-tech system. This can include touching letters or pictures on a screen and then converting it to speech.

The ALS Association differentiates augmentative communication and alternative communication [14]. Augmentative communication is used by people who still have some speech ability but may be degraded or have limited abilities. Augmentative modes of communication are used to support, or supplement what the person is able to say verbally in this case. Alternative communication is required when a person has no speech. In this case, patients must rely on an alternative communication method to make all of the needs and desires known.

The content and configuration of aided systems can vary greatly depending on the capabilities and comforts of the individuals using the systems. High-tech systems are constantly being developed and improved to better accommodate user's needs.

1.4 Eye and Gaze Tracking

Eye and gaze tracking devices can be used to facilitate high-tech aided AAC systems. The tracking devices can be employed as an intermediary interface between the individual user and a computer system.

Oculography is the method for recording eye positions and movements [9]. There are several different methods to perform oculography including electro-oculography, sceleral search coils, infrared oculography, and video oculography. Electro-oculography uses sensors attached to the skin around the eyes to measure electric fields as the eyes move. This method requires very close proximity to sensors; however, it does have the benefit of being able to detect eye movement when the eyes are closed (such as when sleeping). Sceleral search coils is implemented by embedding small wire coils into a modified contact lens. The contact lens is then placed in the eye. Using a mirror in the contact lens, reflected light can be measured, and by determining the orientation of the coil in a magnetic field, an individual's eye orientation can be determined at high accuracy. However, this method is highly invasive. Infrared oculography measures the intensity of the reflection infrared light illuminating the eye. This method can be easy to implement, using a system mounted to spherical glasses; however, the detection range can be limited. Video oculography is the most widely used and widely researched method for eye tracking. It uses one or more cameras to determine the movement of eyes based off of the images captured.

In order to perform gaze tracking, the location of the eye must be detected from a collected image. Information about the eye regions and head pose can be gathered, and a gaze direction can be estimated. Eye and gaze tracking technologies can be used as input devices for computer systems. If an individual is presented with a computer screen, a gaze tracker can detect where the user is looking and interpret this to control a computer cursor.

1.5 Grounded Language

The goal of language grounding is to interpret the physical world into a representation of natural language tied to it [11]. Humans have the ability to see objects, identify, and describe them in a way that other humans can typically easily understand. It becomes a much larger problem space when attempting to have a system, such as a computer or robot, that can interpret these physical objects and assign meaning to them. For instance, if we have a system that understands the meaning of stripes and the meaning of horse, we might be able to have the system learn the meaning of zebra as stripes plus horse [10]. But, in this case, how do we get the system to understand stripes and horse in the first place?

There are several other challenges that are introduced when studying language grounding [12]. In order to develop a robust, open-ended flexible intelligent system that can handle such tasks as language communication, we must consider the whole systems approach; that is, integrating multiple subfields of artificial intelligence, such as sensorimotor intelligence, high level knowledge representation, and reasoning and verbal interaction, into a single system. Another challenge lies in the leveraging of robot systems capable of using their computing power, sensor arrays, and motor controls to assist in developing language models. There must be a connection made between the intelligent system platforms and the natural language aspects of the problem. Lastly, human language is constantly evolving. New words are constantly being created and the meaning of existing words are changing. We need to develop a system that is flexible enough to adapt to the ever-changing cultural system of language.

Chapter 2: Related Work

2.1 Existing AAC Products

With modern technology, patients with ALS who are unable to speak use electronic devices to make themselves understood [14]. There is a portion of ALS patients who can speak but are unable to physically type to utilize certain electronic devices [15]. In a Windows environment, a keyboard and mouse input are essential. In many cases, there is an on-screen keyboard available for text input. There is also the option of an on-screen mouse to use as input in place of the physical hand-held mouse. These virtual counterparts allow users to have access to the same Windows features that any other user would have using the physical devices. There are also many different options for text to speech tools available that will speak whatever is typed into the computer. Some of the more popular, free of cost text to speech applications that may be more popular include E-triloquist, MyFTC, Natural Reader, and PVoice.

A mobile device, such as a cellular phone or tablet, can be used for communication. If a patient has difficulty with speech but has hand and finger function, they can use applications such as Speak It, Small for Aphasia, Locabulary, or Proloquo2Go as voice output [13]. In a similar function, a personal computer can also be used for communication via voice output software. If a patient no longer has use of their hands, they may use a headmouse as input to the computer. A user operates the computer through movements of the head that are being sensed by a receiver connected to the computer. Dedicated speech generating devices can also be used but tend to be very expensive, ranging from approximately $5000 to over $15000. These systems include VMax from Dynavox, ECO from Prentke Romich, and Lightwriter from Tobii/ATI.

The EyeLive system is a method to track the gaze of ALS patients. The EyeLive system can track eye movement to determine the direction a user is looking, and by using a nine-cell grid, creates a hierarchical selection algorithm to generate text input that can be used for communication. After initial experimentation it was found that the gaze tracking was inconsistent, and typing speed was below a desirable rate, requiring much training and practice to increase speeds [16]. Because the interface still employs a single character selection method to craft full messages, it requires three interface selections for each character, rendering it slower than a standard QWERTY keyboard. However, this is an interesting approach to implementing a communication interface due its simplistic design.

Two communication methods currently in use by ALS patients are the EyeGaze and GazeTalk applications [18]. GazeTalk uses a hierarchical menu structure with word prediction in order to facilitate communication. EyeGaze employs the more traditional virtual QWERTY keyboard for eye typing. Through very small sample size studies, it was demonstrated that the typing rate with the EyeGaze system was higher than the GazeTalk system. The EyeGaze also had a lower error rate than GazeTalk. There are beliefs that these results might have been observed due to user's familiarity with the simple, standard keyboard layout. There was also more error in the GazeTalk application due to the prediction method. Once a word had been predicted, a user would have to backspace through the word in order to correct it. With the simple keyboard of EyeGaze a user would type exactly what they wanted, resulting in the need for less error correction.

Another product in use by patients with ALS is the Eye-gaze Response Interface Computer Aid (ERICA) [19]. ERICA, when combined with other tools, can provide onscreen keyboards, full words, command keys, abbreviation expansion, Windows-based programs, and environmental controls. During a study of 15 individuals with ALS, it was found that 14 of the individuals were able to continue using the ERICA system successfully through their progression. It showed that a desire existed in the study participants to continue communication and use of other applications. It also led to the conclusion that many methods of AAC can require a lot of lead time to to obtain, set up, and learn, resulting in high over costs.

2.2 Related Research

2.2.1 Gaze Tracking and Interface

Liu et al. introduces an eye-gaze tracking system in an attempt to overcome more expensive and complicated models [24]. The system presented, called EyeLive, uses infrared sensors installed on a pair of glasses to detect the direction of the eye gaze as well as blinking. The EyeLive interacts with a computer interface using a grid divided into nine areas, mostly focusing on five of the areas, leaving out the extreme corners. Running through multiple trials of user testing demonstrated reasonably successful results, showing accuracy when targeting specific points in 5 regions. As technology advances, more and more affordable and reliable solutions for gaze tracking will become commercially available for use.

As stated by Majaranta and Räihä, "most eye typing systems are implemented by having a virtual keyboard on the screen." It is stated that while gaze tracking input can be fast, the resolution can be an issue. It can be difficult for gaze tracking systems to detect minute differences in angles of the gaze. It is also mentioned that, "when a traditional keyboard is used, the user can feel the keys and hear the clicks while typing. S/he can also observe the changes on the screen as the letters appear. Before the user of an eye typing system can select the focused item, the system should give clear feedback." Some type of on screen feedback is needed to avoid any confusion as to where a user is focused while using a gaze tracking device. This is a good thing to keep in mind when designing a user interface. There are many things to take into consideration when designing the interface including the size and layout of selectable entities. In regards to the interface specifically, Majaranta and Räihä came to the conclusion that typing issues had not really been studied in detail at the time. They state that there are some interactions that have been largely overlooked including methods for editing text, selecting a chunk of text, scrolling the text, and undo methods [17].

Some conclusions from Carrington, Hurst, and Kane in a study of "Wearables and Chairables" can also be applied to this problem. Their study showed that users had a desire for technologies that can fit to the form factor of the wheelchair they may be confined to. They also showed interest in wearable technology [21]. By following these guidelines, we can design an interface to work with technology that is streamlined to fit comfortable on a wheelchair (i.e. tablet computer display for an interface and smaller, less intrusive gaze tracking device).

2.2.2 ALS and Communication and Technology

In a study of "Duration of AAC technology use by persons with ALS," it was found that the mean duration of AAC use was 28.4 months among all participants. One of the observations observed from the study was that there may be an increase in the number of ALS patients opting for invasive mechanical ventilation. This results in the need for access to AAC with considerably less motor control as the lives of the individuals are prolonged [20]. Overall this results in a greater demand for more accurate, easy to use AAC gaze tracking system. As gaze tracking technology improves, so will the usefulness of the applications and user interfaces tied to them.

Beukelman, Fager, and Nordness outline the importance of a timely referral for communication support for people with ALS [23]. Because the rate of progression of the disease varies from patient to patient, it can be difficult to determine the exact time of when one should begin using an AAC system. It is suggested that an early adoption is preferred. They also state that general acceptance of AAC tools is generally high, and AAC tools are often used until near the end of life of the ALS patient. It is proposed that AAC systems that are adaptable to the level of motor function of a user is very beneficial. As a user progresses with ALS it may be necessary to transition from an interactive touchscreen to a gaze tracking interface. One of the future goals of the AAC system outlined in this paper is to have the language model learn and adapt while the user is still able to speak and as they start to progress with the symptoms of ALS.

Research performed by Richter, Ball, Beukelman, Lasker, and Ullman reported that AAC technology of varying levels is employed to fulfill different purposes [26]. It was reported that users will express quick needs using unaided or low technology methods of AAC. More in-depth information is often communicated using high-tech AAC systems. Fried-Oken et al performed a study, surveying caretakers of patients diagnosed with ALS to determine the main purposes of AAC device use [22]. Based on the purposes included in the survey, they found that AAC devices are most used for getting basic needs met and clarifying needs; however, it was noted that over one quarter of participants also responded that they did not use an AAC device for those purposes at all. Participants deemed that making sure basic needs are met and providing instructions or directions to others were among the most imported uses for AAC devices. A study performed by Joan Murphy, consisting of observing fifteen people with ALS in their homes [25]. While none of the subjects used any type of high-tech AAC device, it was found that communication was still a very important part of their lives. It was observed that it was most important to communicate to develop and maintain social closeness rather than to convey needs or wants or transfer information. The goal for the development of my AAC system is to be beneficial for users in all communication situations.

2.2.3 Grounded Language

An experiment conducted by Chen and Mooney leveraged a symbolic form of an event in order to produce commentary related to the actions taking place [29]. Their focus was on extracting existing commentary, parsing, and assigning it to pieces of corresponding action. By training a language model in this fashion, they were able to successfully simulate sporting events and generate accurate commentary to go along with it. This idea behind this language model parallels some of the aspects I will be aiming to reach in future work. I aim to have a system that will be able to generate language based on objects or actions around the central agent.

Dindo and Zambuto developed a probabilistic approach to learning a grounded language model [28]. They created an interactive training system to teach the salience of objects to a robot using visual cues and descriptive language. From the description, the language model can assign semantic categories and classes to objects which will eventually develop into lexical item, branching a word to a semantic category. They have also extended their work to applying syntactical constraints to the words during training of the language model. This allows for a firmer representation of word meaning in the language model. This work provides great foundation for developing a grounded language model.

Relying heavily on hidden Markov models, Yu and Siskind developed a grounded language model that learns word meanings from pairings of videos and sentences [27]. Unlike several other grounded language models, this work results in scenes labeled with complete sentences rather than just individual words. They developed a language model that is able to analyze an entire scene or series of scenes in video form and develop some type of meaning for it. The material covered in this research could contribute greatly when considering real life scene interpretation on-the fly in order to gather language groundings to support environmental context.

2.2.4 Context Understanding

Dey and Abowd performed a study to better understand context and context awareness in computer systems and interactive applications [30]. They defined context as, “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.” This definition was stated to make it easier to associate the types of context needed for various application scenarios. Dey and Abowd also listed location, identity, activity and time as being the most important types of context. These researchers also defined context-aware as, “A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the users task.” Context-aware applications have the ability to adapt to various situations based on context. All of this research provides informative and thought-provoking insight when developing an application with context in mind.

Chapter 3: Hypothesis

The scope of this study is to investigate the usage of a user interface with a simulated language model developed with the purpose of enabling a user to communicate via a gaze tracking device as input. My hypotheses are mostly targeted to the users' reactions towards their interaction with the interface, as well as the comparison of responses with context driven text prediction versus standard text prediction. In order to determine the reaction of users towards the interface, a user study is performed, and surveys are administered throughout the study and afterwards. In order to determine users' feelings towards the communication system over all, we will focus in on survey provided after all tasks of the study are completed. The survey included questions about the interface's usability, complexity, and efficacy, as well as the interaction with the interface using the gaze tracking device. I believe that users will be positive towards the interface and communication system as a whole.

During the study, participants are introduced to the interface in four different incarnations. These four methods are outlines in Table 3.1. The methods vary on the corpus of text relied upon for text prediction, as well as the amount of suggested text provided. The language model behind the interface and the user study procedures are elaborated upon in Chapter 4.

Table 3.1: The four text prediction methods tested, which vary in whether they generate words (W), multiple words (MW), and sentences (S), and whether they are based on an existing English corpus or a preliminary contextual corpus.

Method	Corpus	W	MW	S
StdEng	Standard English	✓	✓	
ContWord	Contextual	✓		
ContWords	Contextual	✓	✓	
ContSent	Contextual	✓	✓	✓

I believe that an interface that provides context-based text predictions with complete sentence suggestions will result in a more pleasing user experience. I believe that users will feel that they can communicate easier and quickly when using an interface with context driven text prediction with complete sentence suggestions when compared to an interface with other methods of text prediction. This hypothesis will be tested against results from a survey administered after each task during the user study. The survey asked for responses from participants based on each one of the methods outlines in Table 3.1.

I also believe that users will in fact require less time to complete tasks in which an interface is provided that uses context-based text prediction with complete sentence suggestions. This will be verified using data recorded during the user study. For each task a participant completes, the number of clicks and the time to complete each task is recorded. The averages between the different interface population methods will be compared against each other in order to determine if there is any significance in the difference in time to complete tasks or the required number of clicks to complete tasks.

Chapter 4: Approach

4.1 Initial Data Collection

To begin this study, it was decided that four scenarios would be created for the basis of the language models and testing.

The four scenarios were:

1. You are on the first floor of a multi-story building. You would like to get to the fifth floor. You move into the elevator, and you happen to end up being positioned towards the back of the elevator. However, there is someone standing next to the button panel. What do you say to them?

2. You have broken your leg and have it in a cast. You are supporting yourself and moving around with crutches. You are wearing a backpack with a water bottle in it, but you are unable to maneuver on the crutches to access it. You are getting thirsty and would like your water bottle. What do you say to a nearby friend?

3. You are roaming the halls of work and pass a co-worker. They stop, smile, and say, “Hey, how's it going?” Your day is going very well. With what do you reply?

4. You have been having stomach pains after eating each day for the past week. You are visiting your doctor. They ask you how you are doing. What is your response?

These scenarios were selected for being common interactions with other people that could lead to natural responses. After selecting these scenarios, I solicited anonymous responses to a preliminary survey via a Google Form. Survey participants were presented with various scenarios and asked to type responses. The scenarios required spoken messages. These responses were then used in support of the language model behind the interface.

4.2 The Interface

The interface was designed and developed in an attempt to create a simple and intuitive input system. The main page of the interface, as seen in Fig. 4.1, displays the outgoing message, buttons with suggested text, and several utility buttons. The field at the top of the screen provides a clear display of the currently crafted outgoing message. The buttons directly below the message field will vary depending on the scenario. Fig. 4.1 shows the interface as it would look when providing single word text predictions. Fig. 4.2 shows the interface as it would look when providing single word and multiple word predictions. Fig. 4.3 shows the interface as it would look when providing complete sentence or phrase predictions. When a button is selected, the text within is then added to the outgoing message, and the suggested text contained in the buttons is updated as well.

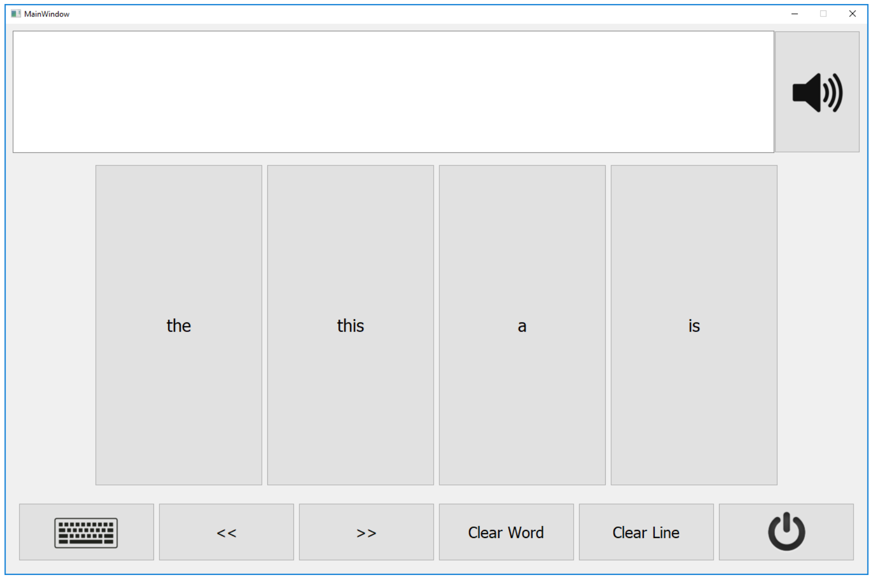

Figure 4.1: The main screen of the developed interface. It is currently showing single word predictions

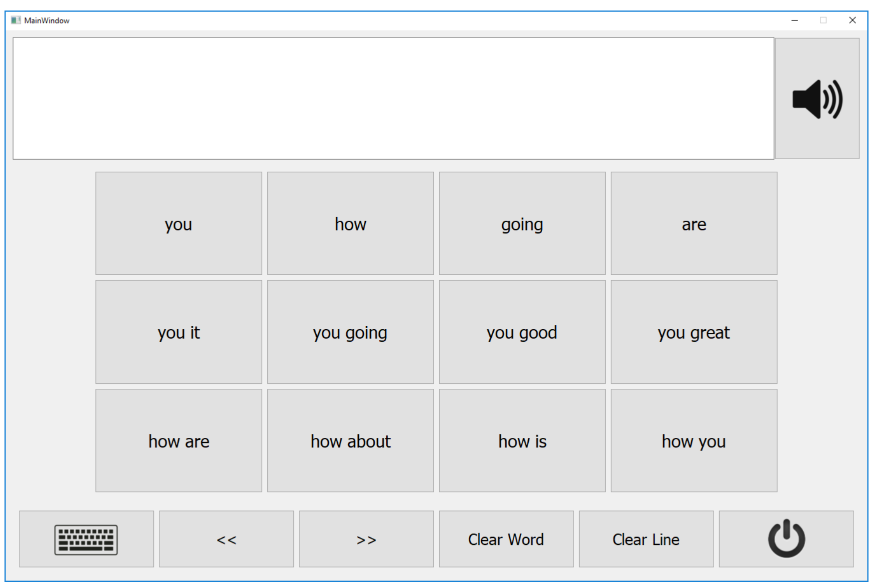

Figure 4.2: The main screen of the developed interface. It is currently show single word and multiple word predictions.

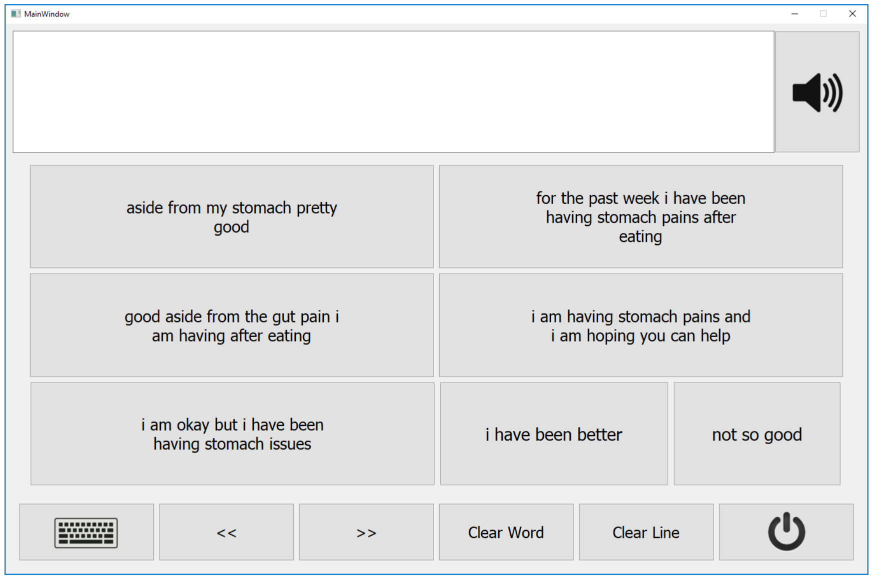

Figure 4.3: The main screen of the developed interface. It is currently showing full sentence predictions.

The text is drawn from the language model based on the current outgoing message. At the bottom of the interface screen are utility buttons. From left to right in Fig. 4.1, the utility buttons are as follows:

1. The button that contains the image of a keyboard will take the user to a page with an on-screen QWERTY style keyboard, as seen in Fig. 4.4.
2. The [<<] and [>>] buttons will cycle through pages of suggested text.
3. The [Clear Word] button will remove the last word from the message field.
4. The [Clear Line] button will clear the entire message field.
5. The button with the power symbol on it will close the application.

There is also a button located directly to the right of the message field contain-

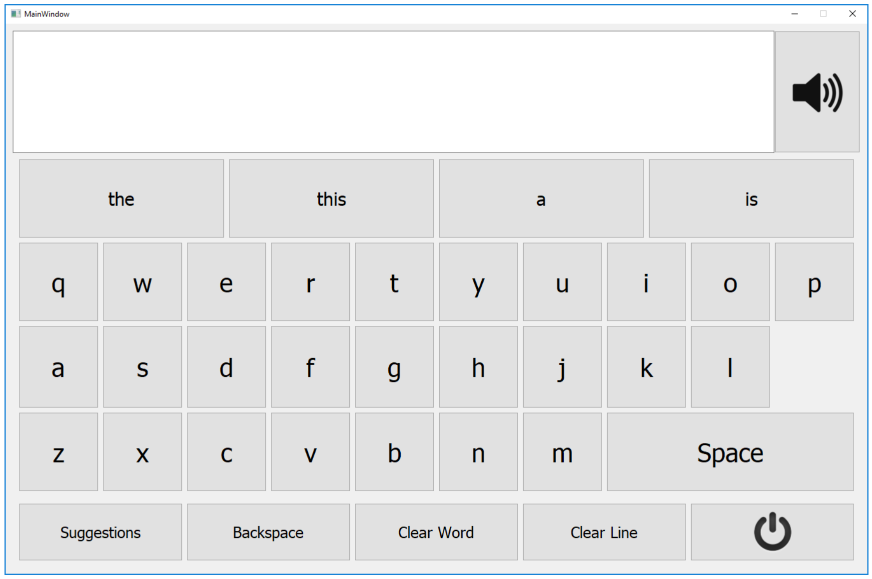

Figure 4.4: The QWERTY keyboard page of the interface.

ing an image of a speaker. This button will read the message aloud and complete the current task for the user study.

The utility buttons for the QWERTY style keyboard screen of the interface in Fig. 4.4 are very similar to those seen on the regular text prediction screen. There are several exceptions. The [Suggestions] button take the user back to the screen referenced previously with only predictive text selections. The [Backspace] button will delete the last character in the message field. The [SPACE] button inserts a space character into the outgoing message field. The rest of the character keys input the character one would expect into the message field. The QWERTY keyboard screen also has single word auto-complete and predictions directly below the message field. The QWERTY page allows for a more customized input. If the user does not like any of the predicted text on the main interface screen, they may choose to input text via the QWERTY keyboard.

The interface was designed to be straightforward to use with very little learning curve. The interface should be easy to read by users. The size of the interface was implemented to approximately fit on a standard tablet computer for future development and testing uses.

4.3 The Language Model

The long-term goal for this AAC application is to have a language model that has been taught by the user over time in order to learn proper phrasing. The overall system will also involve sensor readings from the environment to develop context. The system will be able to intelligently select portions from the language pool depending on the context for suggestions through the interface.

In order to simulate a fully functioning language model, I first needed to have the system "learn" from the user. To accomplish this, I used the responses to the preliminary survey. The responses to the survey were then collected and compiled into an n-gram language database for each scenario leveraging the Presage predictive text entry platform. It added the collected language to an already compiled English language database. The standard data set is built into the Presage platform and is trained from the novel Don Qijote. This n-gram system allowed for predictive text to be implemented based on the text that had already been selected for output. In order to simulate the language model being context-based, the scenario specific language model can be implemented on the back end. In order to populate full

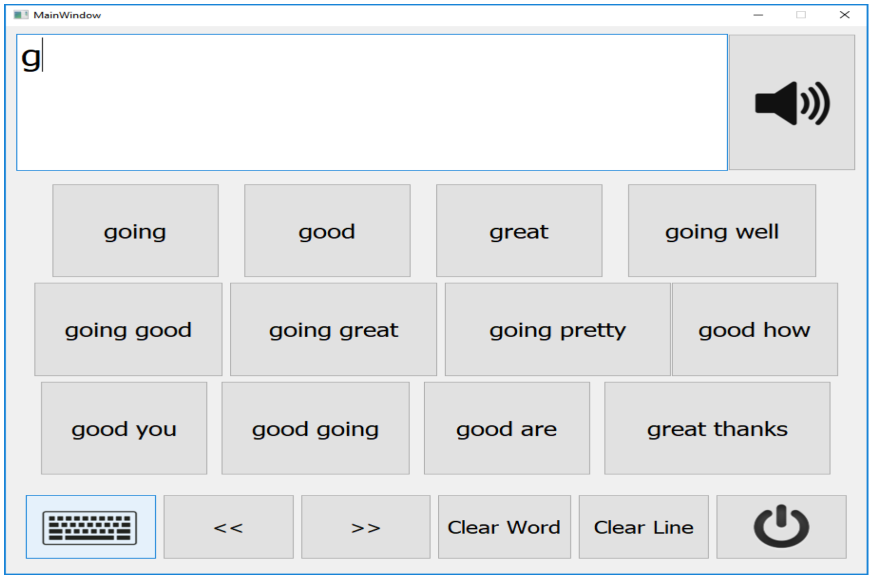

Figure 4.5: The main screen of the developed interface. It is currently showing the interface after a user has just started a message.

sentence suggestions, sentences are taken directly from the list of responses to the preliminary survey.

The interface as shown in Fig. 4.5 shows the text prediction and auto-complete in action. The scenario being presented is: *You are roaming the halls of work and pass a co-worker. They stop, smile, and say, "Hey, how's it going?" Your day is going very well. With what do you reply?* The text prediction method supplying suggestions to the interface is context based with one and two word suggestions.

Fig. 4.6 shows the interface after a user has begun responding to the scenario: *You have broken your leg and have it in a cast. You are supporting yourself and moving around with crutches. You are wearing a backpack with a water bottle in it, but you are unable to maneuver on the crutches to access it. You are getting thirsty and would like your water bottle. What do you say to a nearby friend?* The main

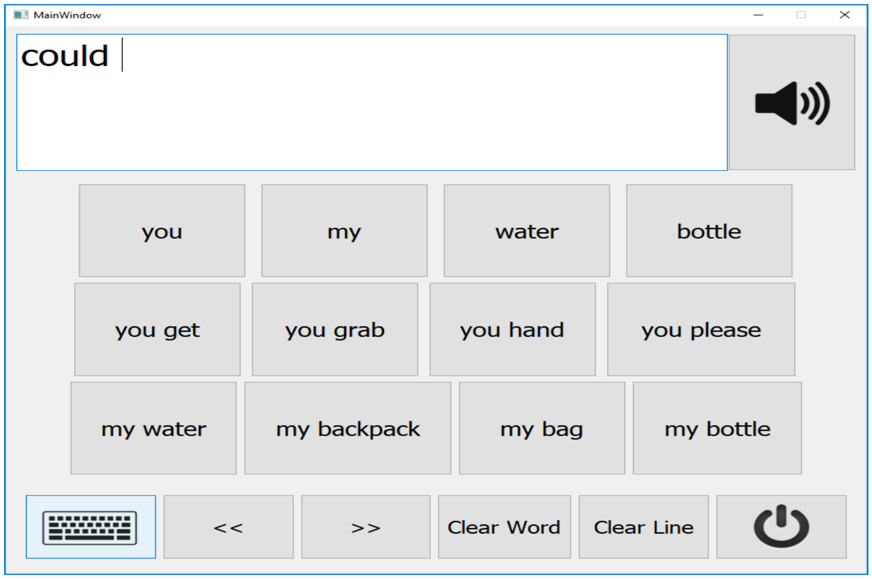

Figure 4.6: The main screen of the developed interface. It is currently showing the interface after a user has selected a suggested word.

interface is shown presenting one and two word suggestions that are context based.

Fig. 4.7 shows the interface in the same state as previously described, however, it is displaying the QWERTY screen. The single word suggestions based off the current input can be seen directly below the message field. These suggestions are context based similar to the main screen suggestions.

The interface as shown in Fig. 4.8 is what the main interface screen would look like if the user described previously had entered another character into the message field. Updated context based, one and two word suggestions can be seen on the interface.

Fig. 4.9 shows the QWERTY screen of the interface after a partial response has been entered into the text field. The auto-complete suggestions in this case are predicted using the standard English library.

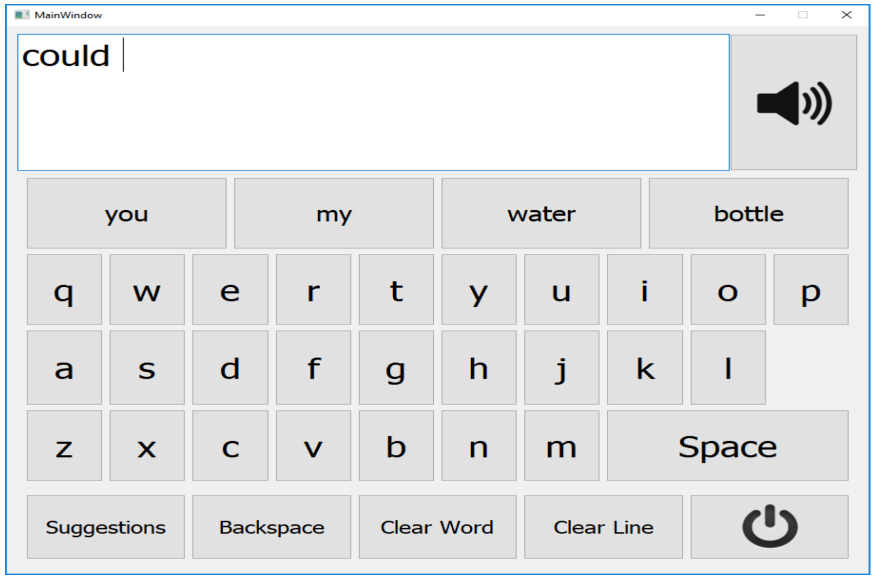

Figure 4.7: The QWERTY screen of the developed interface. It is currently showing the interface after a user has selected a suggested word.

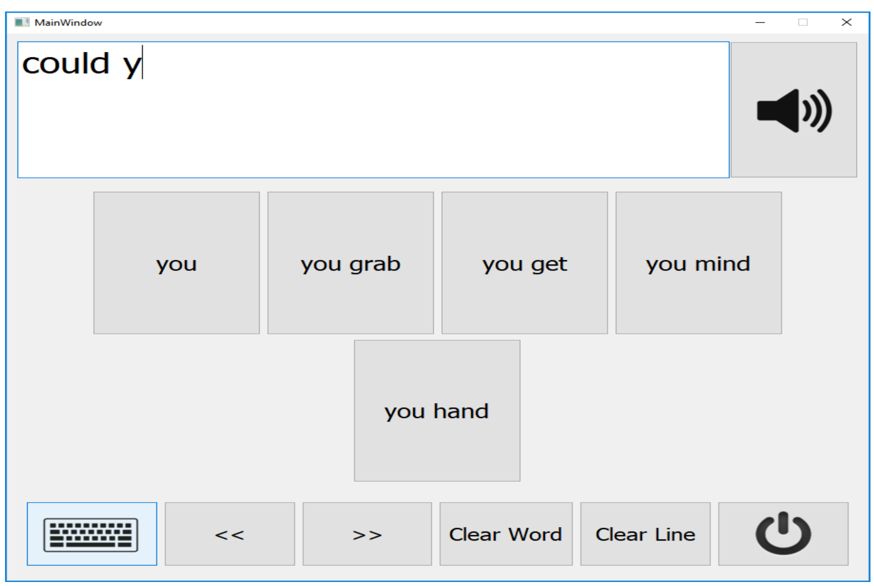

Figure 4.8: The main screen of the developed interface. It is currently showing the interface after a user has entered in a partial word.

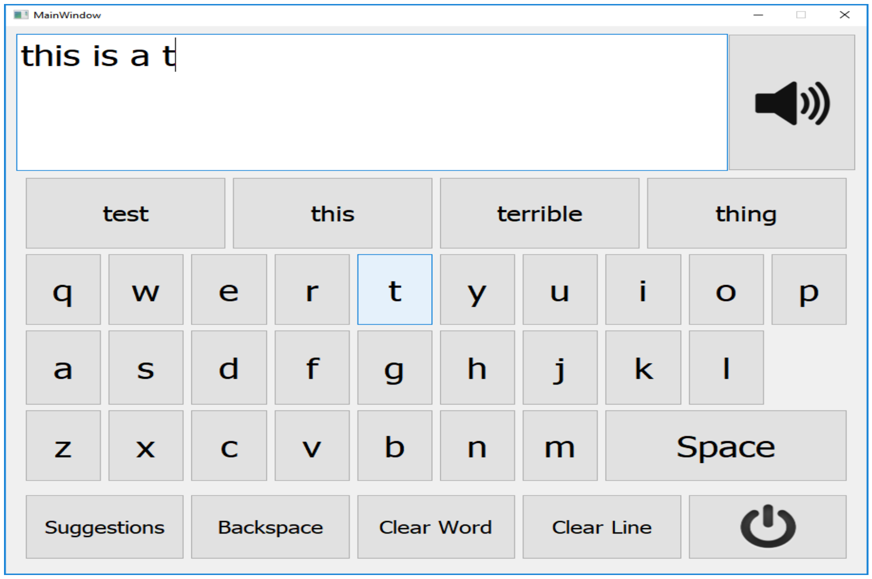

Figure 4.9: The QWERTY screen of the developed interface. It is currently showing the interface after a user has entered in a partial word.

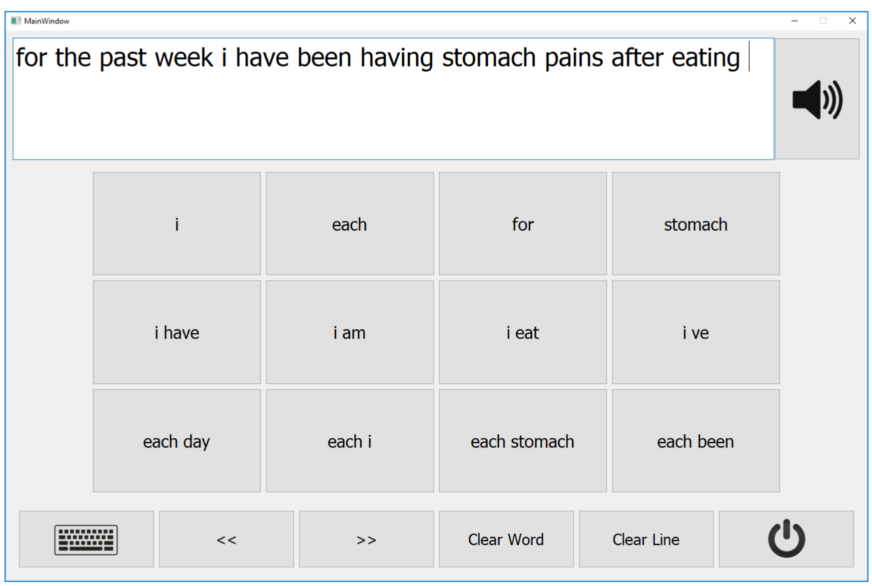

Figure 4.10: The main screen of the developed interface. It is currently showing the interface after a user has selected a complete sentence to output.

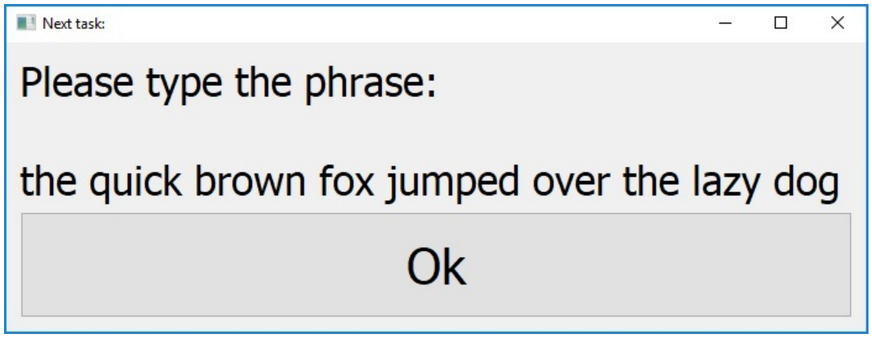

Figure 4.11: One of the familiarization tasks asked of users for the interface.

An example of the interface demonstrating the selection of a suggested complete sentence with next word(s) suggested can be seen in Fig. 4.10. The outgoing message that has been entered can be seen in the text field. The interface has suggested what it determines is the next best word or series of words based on the supporting language model.

4.4 The User Study

Participants of the user study were asked to interact with the developed AAC interface using a Tobii Dynavox PCEye Mini eye tracker for input. There were specific pieces of the interface developed to help facilitate the study. Before each task, the user was presented with a screen providing instructions as to what to do next as seen in Fig. 4.11 and Fig. 4.12. Once a user reads the text on the prompt, they must click the [Ok] button to continue on to the task itself. Once a user is completed with a task (finished entering text into the outgoing message field), they click the text-to-speech button (the button with the image of the speaker on it).

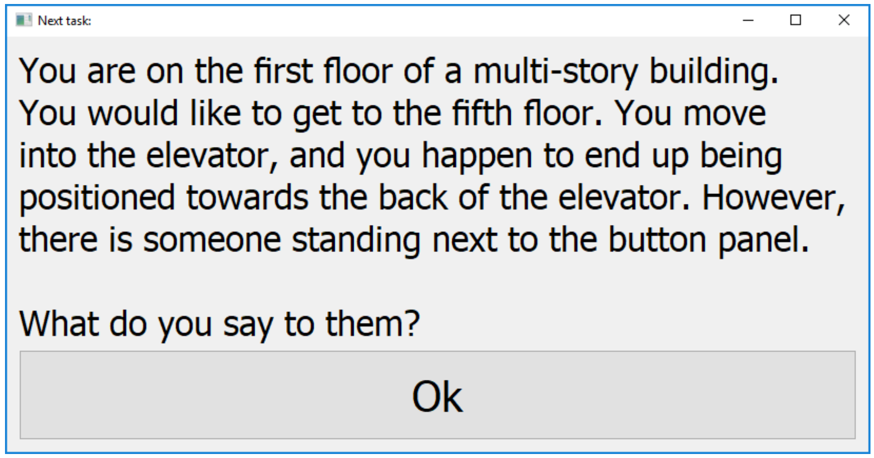

Figure 4.12: One of the scenario prompts displayed as part of the interface.

Each user was provided with familiarization tasks for the interface. This allowed the user to become familiar with the capabilities of the interface but did not reveal the different experimental methods or the context-based, predictive text language models. Each user would receive the same set of tasks, also referred to as scenarios. The first familiarization task specified a sentence to type into the outgoing message field and the single word predictive text options available to them provided all of the words necessary to complete the sentence. It was designed to gain familiarity with the main screen of the interface providing suggested text. The second familiarization task specified another sentence to type into the outgoing message field. For this task, the single word predictive text was not sufficient to complete the sentence, forcing the user to utilize the QWERTY keyboard. The goal of this task was to gain familiarization with the QWERTY keyboard input screen of the interface.

After the familiarization task, the user is presented with a series of four additional tasks, the scenarios enumerated in the previous section, in a random order. The user would be presented with single word, standard text prediction on the main screen of the interface for the first scenario, regardless of the scenario. The user would be presented with context trained, single word predictive text for the second scenario they were tasked with completing. For the third scenario, the user would be presented with context-sensitive, single and multiple word text prediction. On the final scenario, the users were presented with context-sensitive, single word, multiple word, and complete sentence predictions to select from.

For each task, the user was able to input text into the outgoing message field by selecting from the suggested text on the main screen or switching to the QWERTY keyboard to enter text. Users were able to switch between screens at will. After each scenario is completed, the user was asked to complete a short survey regarding the task. After all scenarios are completed, the user was asked to complete a survey in regard to the overall experience with the interface and study.

4.5 Surveys

The questions asked in each survey can be seen in Appendix A. All questions asked users to respond using a five step Likert scale. Although there were too few responses in this user study for statistically significant qualitative interpretation, I am able to provide anecdotal analysis.

The survey given after all tasks were completed gave insight to how users felt about the overall communication system. There were several questions regarding the complexity, consistency, and ease of use for the interface. Several of the questions were related to using the gaze tracking device to control the mouse cursor and select buttons. This survey will provide information as to how the overall communication system was received and can provide valuable input into changes and improvements to be made to the system in the future.

The surveys given after each task were aimed at gathering information about how the user felt using the communication system for the various methods outlined in Table 3.1. The questions in those surveys are related to how easy the participants found the tasks, how frustrated they were using the interface to complete tasks, and how effectively they thought they could communicate using the interface. Since these results were recorded on a task by task basis, it allowed for the comparison between interface text prediction methods.

The interface program monitored the number of clicks a user makes to interact with the interface and records the number of clicks required to complete each task. The amount of time it takes a user to complete each task was also recorded by the program. These numerical values allow for the comparison of time required and total interaction required to complete the tasks based on the methods of text prediction provided by the interface. By analyzing average values from task to task, I can see which methods of text prediction appeared to better support quick and effective communication.

Chapter 5: Results

5.1 Overall Interface Reactions

As stated previously, participants of the user study were administered a brief survey after interacting with the provided interface for each scenario task. There were four of these short surveys in total. After all tasks were completed, a comprehensive survey was administered regarding the overall study. The surveys allowed responses based on a five step Likert scale. The exact surveys can be referenced in Appendix A. There was a total of twenty participants in the user study.

By observing the results of questions 1 through 10 of the final, comprehensive survey, we can see participants reception of the communication system and some of its individual components. The number of responses for each question can be see in Fig. 5.1 and Fig. 5.2. We can analyze the number of different responses for each question to observe certain trends and to draw conclusions.

For the first question, study participants were asked to respond to the following statement: *I think that I would like to use this interface frequently (given the situation where you must communicate via text).* A majority of users responded that they somewhat agreed with this statement. The second question asked users to respond to the following statement: *I found the interface unnecessarily complex.*

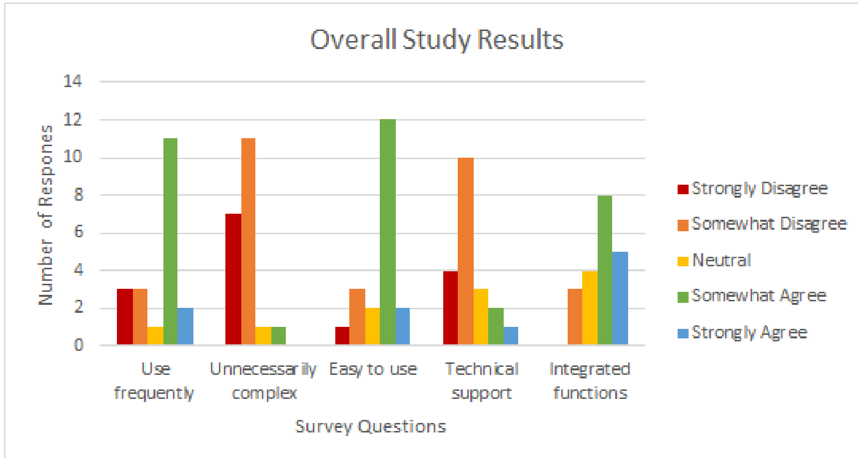

Figure 5.1: These are the responses to the survey administered after all tasks were completed. These questions correspond to those numbered 1 through 5 in Appendix A.

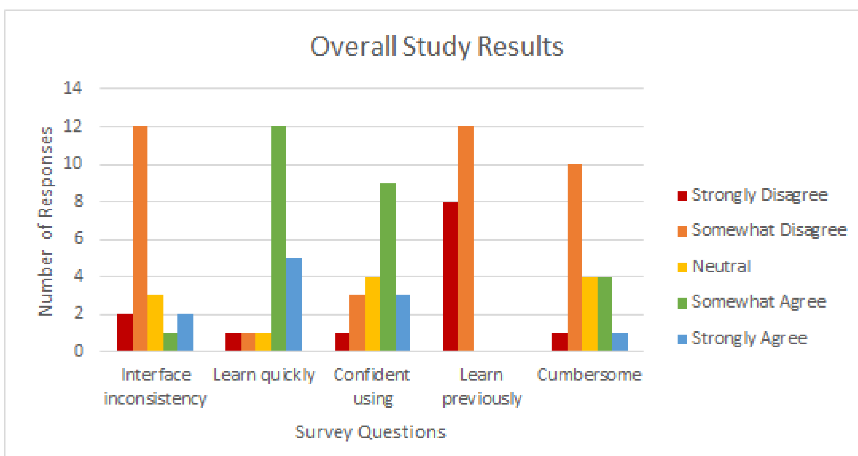

Figure 5.2: These are the responses to the survey administered after all tasks were completed. These questions correspond to those numbered 6 through 10 in Appendix A.

A majority of users responded that they somewhat disagree with this statement. The third question presented users with the following statement: *I thought the interface was easy to use*. A majority of users responded that they somewhat agree with this statement. The fourth question asked users to respond to the following statement: *I think that I would need the support of a technical person to be able to use this interface*. Half of the users responded that they somewhat disagreed with this statement. Only three users responded that they either somewhat agreed or strongly agreed with the statement. The fifth question of the survey prompted users to respond to the following statement: *I found the various functions in this interface were well integrated*. A majority of users responded that they either strongly agreed or somewhat agreed with this statement. The sixth question asked users to respond to the following statement: *I thought there was too much inconsistency in this interface*. A majority of users responded that they somewhat disagree with the statement. From this series of questions, it can seen that users believe that the interface was easy to use, well integrated, not unnecessarily complex, and non inconsistent. Most users thought that they would not require the support of a technical person to be able to use the interface and said that they would use the interface frequently if they were in a situation where they required such a thing.

The seventh question asked users to respond to the following statement: *I would imagine that most people would learn to use this interface very quickly*. A majority of users responded that they somewhat agree with the statement. The eight question prompted users to respond to the following statement: *I felt very confident using the interface*. A majority of the users responded that they either strongly agreed or somewhat agreed with the statement. The ninth question asked users to respond to the following statement: *I needed to learn a lot of things before I could start using this interface.* All users responded that they either somewhat disagreed or strongly disagreed with the statement. The tenth question of the survey asked users to respond to the following statement: *I found the interface very cumbersome to use.* Eight users responded that they somewhat disagreed with the statement, one user responded that the strongly disagreed with the statement, four users responded that they were neutral about the statement, four users responded that they somewhat agree with the statement, and one user said that they strongly agreed with the statement. From these four questions, it can be seen that users believed the interface can be picked up my most other people quickly, they did not need to learn many things prior to using the interface, and the interface was not cumbersome to use. Overall, users were fairly confident when interacting with the communication system.

5.2 Gaze Tracking Device as Input

Questions 11 through 14 of the final survey were targeted more towards users' reactions of using the gaze tracking device as input to the communication interface.

The eleventh question prompted participants to respond to the following statement: *I found it hard to look at the buttons long enough for them to be clicked.* Ten users responded that they somewhat disagreed with the statement, one user responded that they strongly disagreed with the statement, two users responded that

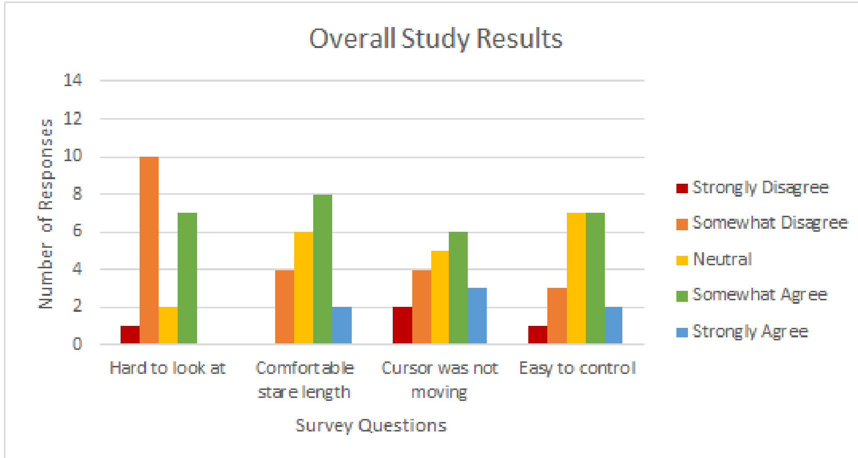

Figure 5.3: These are the responses to the survey administered after all tasks were completed. These questions correspond to those numbered 11 through 14 in Appendix A.

they were neutral towards the statement, and seven users responded that they somewhat agreed with the statement. The twelfth question asked users to respond to the following statement: *I didn't mind the staring at the buttons for the length of time needed to cause a click to occur.* Eight users responded that the somewhat agreed with the statement, two users responded that the strongly agreed with the statement, six users responded that they were neutral towards the statement, and four users responded that they somewhat disagreed with the statement. The thirteenth question asked users to respond to the following statement: *I often felt the mouse cursor wasn't moving to the button I wanted to click.* Six users responded that they somewhat agreed with the statement, three users responded that they strongly agreed with the statement, five users responded that they were neutral towards the statement, four users responded that they somewhat disagreed with the statement, and two users responded that they strongly disagreed with the statement. For the final question of the user study survey, participants were prompted to respond to this statement: *I could easily control the mouse cursor with my eyes.* Seven users responded that somewhat agreed with the statement, two users responded that they strongly agreed with the statement, seven users responded that they were neutral towards the statement, three users responded that they somewhat disagreed with the statement, and one user responded that they strongly disagreed with the statement. The responses to these four questions show that users may have had some difficulties when adapting to the use of the gaze tracking device. Participants seem less confident in saying that it was easy to look at buttons long enough to activate a click action, that they didn't mind staring at the buttons for the length of time required to activate the click, that the cursor was easily controlled, and that the button was often moving to the button they wanted to click.

These results overall show more variation of users' responses. While the users had a generally positive reaction towards the interface, users appeared to have slightly more neutral or negative reactions towards using the gaze tracking device.

5.3 Comparing Text Prediction Methods

The results of the short, task-based survey has been organized by question to allow for better visualization and comparison of the different interface nuances between tasks. The first question of the short survey presented to participants prompted the user to respond to the following: *I felt I could effectively communicate what I felt I needed to for the given situation.* The number of responses to this

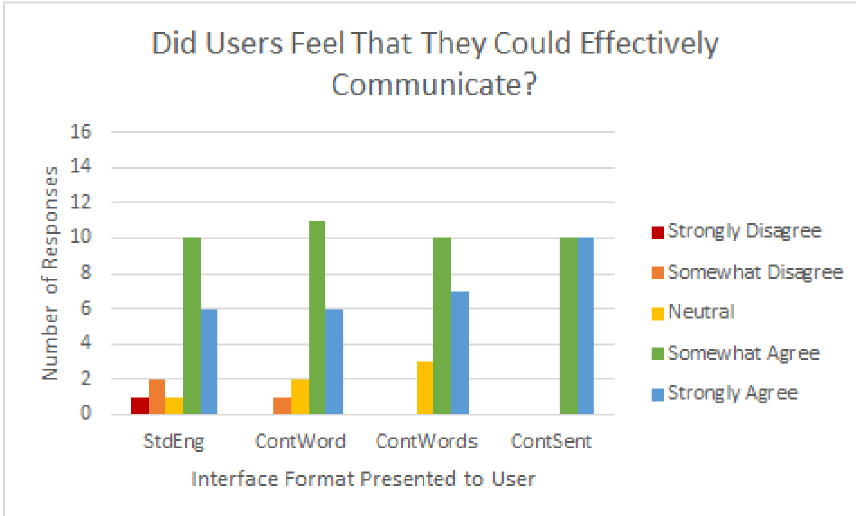

Figure 5.4: These are the responses to the question 1, "I felt I could effectively communicate what I felt I needed to for the given situation," of the surveys administered after each scenario. The interface format shorthand labeled along the horizontal axis corresponds to those outlined in Table 3.1.

questions can be seen in Fig. 5.4. For all tasks, a majority of participants responded that they either strongly agreed or somewhat agreed with the statement. For the task presented to the participant where the interface provided context driven, single word, multiple word, and complete sentence and phrase suggestions, all users responded that they either strongly agreed or somewhat agreed with the statement. In general, participants felt that they were able to communicate effectively for all tasks. There was a slight increase in the number of people who agreed with the statement with the task that presented users with a context driven, single word, multiple word, and complete sentence and phrase suggestions as compared to other tasks.

The second question of the short survey prompted the participants to respond

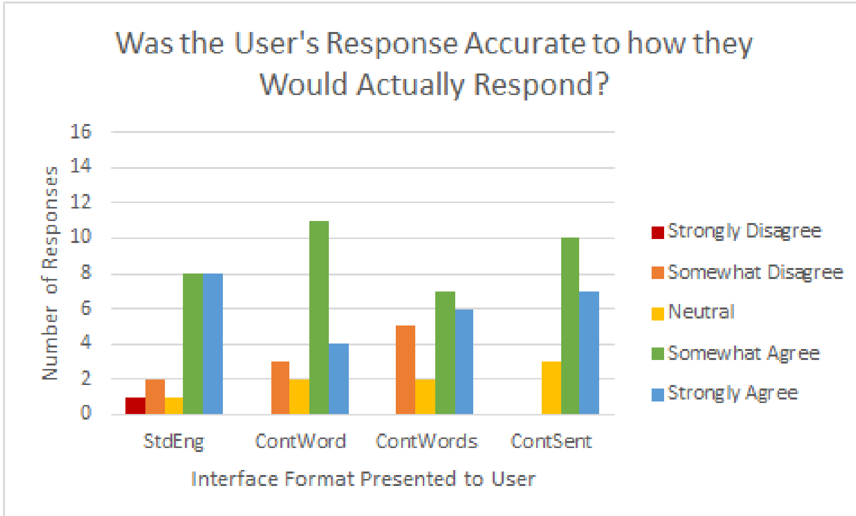

Figure 5.5: These are the responses to the question 2, "The response I provided is exactly what I would have said in the given situation," of the surveys administered after each scenario. The interface format shorthand labeled along the horizontal axis corresponds to those outlined in Table 3.1.

to the following statement: *The response I provided is exactly what I would have said in the given situation.* The number of responses to this questions can be seen in Fig. 5.5. Once again, a majority of users responded that they either somewhat agreed or strongly agreed with the statement for every task. And once again, in general, participants believed that the response they provided was what they would have actually said in the situation. There was a slight decrease in the number of participants who disagreed with the statement with the task that presented users with a context driven, single word, multiple word, and complete sentence and phrase suggestions as compared to other tasks.

The third question of the short survey prompted the participants to respond to the following statement: *I felt frustrated using the interface to complete this task.*

The number of responses to this questions can be seen in Fig. 5.6. For the task presented to users where the interface provided standard text prediction with single and multiple word suggestions, users had varied responses. For that task, one user responded that they strongly agreed with the statement, seven users responded that they somewhat agreed with the statement, six users responded that they are neutral about the statement, four users responded that they somewhat disagreed with the statement, and two users responded that they strongly disagreed with the statement. For all other tasks, a majority of the users responded that they either somewhat disagreed or strongly disagreed with the statement. When comparing the responses to each task, there is an increase in the number of participants who disagreed with this statement when going from the standard text prediction to the context driven text prediction. There was, once again, an increase in the people who disagree with the statement when moving to the task that presented users with a context driven, single word, multiple word, and complete sentence and phrase suggestions as compared to other tasks.

The fourth question of the short survey prompted the participants to respond to the following statement: *I found this task easy to complete.* The number of responses to this questions can be seen in Fig. 5.7. Once again, for the task presented to users where the interface provided standard text prediction with single and multiple word suggestions, users had varied responses when compared to other tasks. For that task, one user responded that they strongly disagreed with the statement, three users responded that they somewhat disagreed with the statement, eight users responded that they were neutral towards the statement, six users responded that they

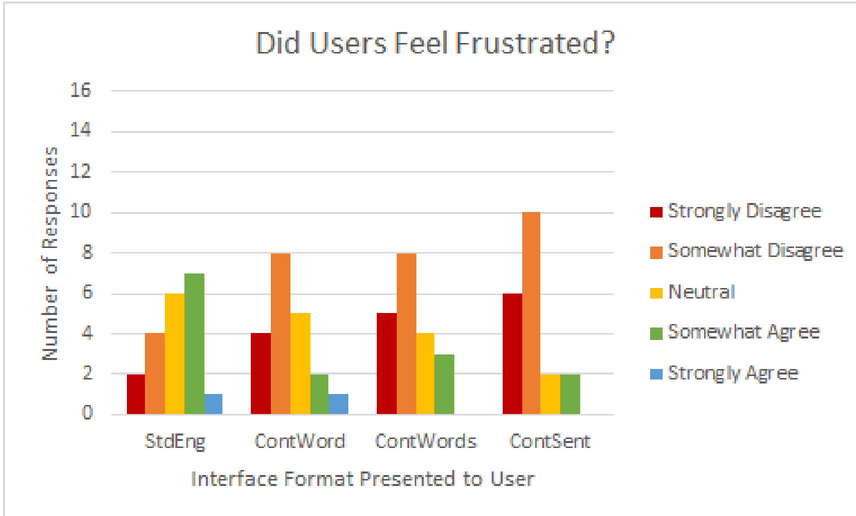

Figure 5.6: These are the responses to the question 3, "I felt frustrated using the interface to complete this task," of the surveys administered after each scenario. The interface format shorthand labeled along the horizontal axis corresponds to those outlined in Table 3.1.

somewhat agreed with the statement, and two users responded that they strongly agreed with the statement. For the remaining tasks, a majority of users responded that they either strongly agreed or somewhat agreed with the statement. There was an increase in the number of participants who claimed they found the task easy to complete when looking at the tasks that provided context driven predictions when compared to the tasks that provided standard text prediction.

The fifth question of the short survey prompted the participants to respond to the following statement: *I felt I could communicate quickly using this interface.* The number of responses to this questions can be seen in Fig. 5.8. Similar to the previous two questions, for the task that presented users with an interface that provided standard text prediction with single and multiple word suggestions, users had varied

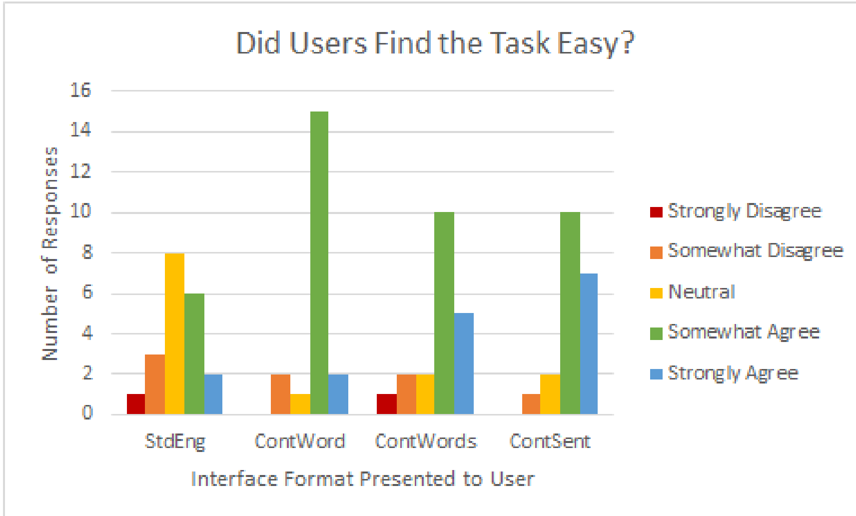

Figure 5.7: These are the responses to the question 4, "I found this task easy to complete," of the surveys administered after each scenario. The interface format shorthand labeled along the horizontal axis corresponds to those outlined in Table 3.1.

responses when compared to other tasks. For that task, two users responded that they strongly disagreed with the statement, six users responded that they somewhat disagreed with the statement, six users responded that they were neutral towards the statement, 4 users responded that they somewhat agreed with the statement, and two users responded that they strongly agreed with the statement. For the other tasks, once again, a majority of users responded that they either somewhat agreed or strongly agreed with the statement. There was an increase in the number of participants who felt they could communicate quickly using the interface for tasks that provided context sensitive predictive text over the standard predictive text. There was an increase in the users who somewhat agree with the statement when comparing the task that provided context driven text prediction with sin-

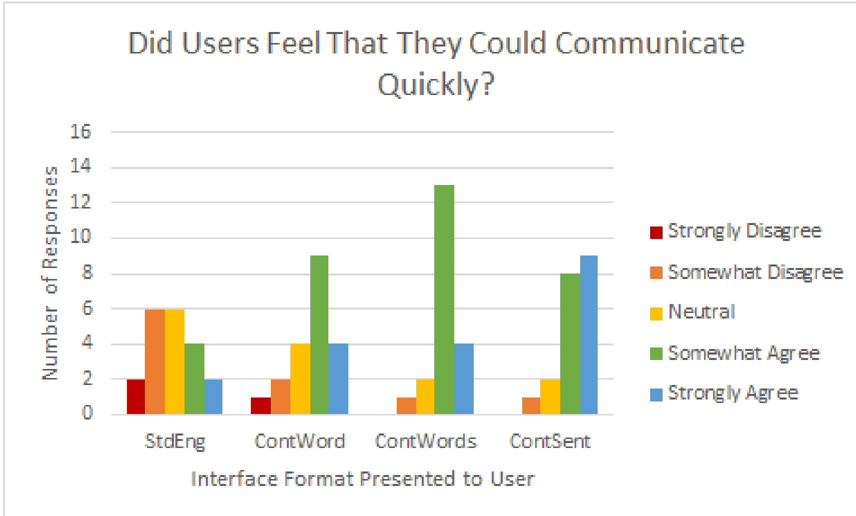

Figure 5.8: These are the responses to the question 5, "I felt I could communicate quickly using this interface," of the surveys administered after each scenario. The interface format shorthand labeled along the horizontal axis corresponds to those outlined in Table 3.1.

gle and multiple word suggestions over the task that provided context driven text prediction with single word suggestions. There was an increase in the participants who strongly agree with the statement when comparing the task that provided users with a context driven, single word, multiple word, and complete sentence and phrase suggestions as compared to other tasks.

The interface program automatically timed and recorded the amount of time it took each user to complete each task. The statistics for the recorded values are visualized in Fig. 5.9. When simply observing the graph of the statistics, there is a visible decreasing trend in the times that is required to complete tasks as the interface becomes more "helpful," meaning there are context-based predictions and the suggested text gets longer. Table 5.1 lists the mean and median required times to

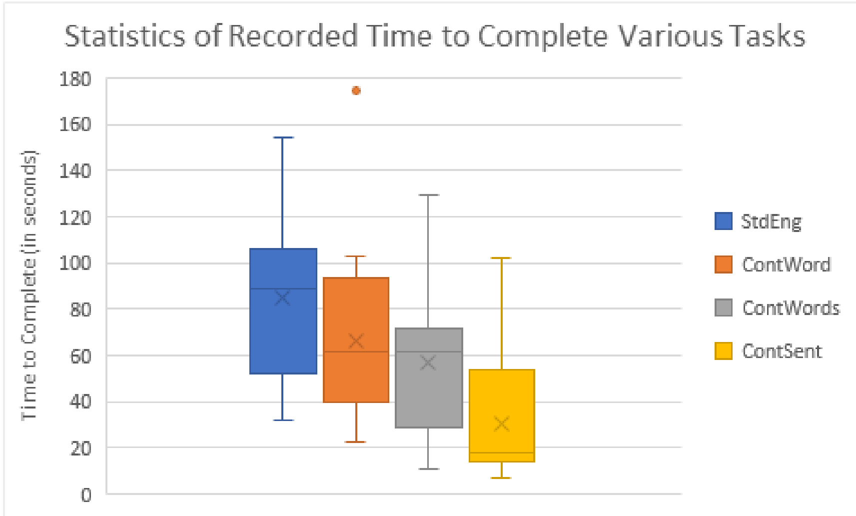

Figure 5.9: These are box and whisker plots showing the time required to complete each method as presented to the user. The 'x' in each plot represents the median value of time required by all users. The interface format shorthand labeled in the legend corresponds to those outlined in Table 3.1.

complete a task for each prediction method. From this data it can be seen that there is a decrease in the amount of time required to complete tasks that provide context driven text prediction. It can also be seen that the time required to complete a task that provides context-based text prediction with complete sentence and phrase suggestions is nearly one third the time of tasks that use standard text prediction with single and multiple word suggestions.

The interface program also recorded the number of clicks each user required to complete each individual task. The statistics for the recorded values are visualized in Fig. 5.10. Similar to the times required to complete tasks, observing the graph of the statistics, there is a visible decreasing trend in the number of clicks that is required to complete tasks as the interface becomes more "helpful." Table 5.2 lists the mean

Table 5.1: The average times (in seconds) to complete tasks with the four text prediction methods tested (as outlined in Table 3.1).

Method	Mean	Median
StdEng	84.57	88.61
ContWord	65.87	61.18
ContWords	57.19	61.33
ContSent	29.99	17.76

and median number of clicks needed to complete a task for each prediction method. Once again, from this data it can be seen that there is a decrease in the number of clicks required to complete tasks that provide context driven text prediction. It can also be seen that the number of clicks required to complete a task that provides context-based text prediction with complete sentence and phrase suggestions is over one third the number of clicks need for tasks that use standard text prediction with single and multiple word suggestions.

These recorded results, combined with the results from the survey administered after each task, show trends towards a preference in a system that can provide context-based text prediction, specifically one that can provide complete sentences as suggested text. There is an obvious increase in performance from users seen in the recorded results as the interface uses a language model that provides more help (context-based text prediction, more suggested words). Results from the surveys also shows a similar trend in the responses and reactions from the users perspective.

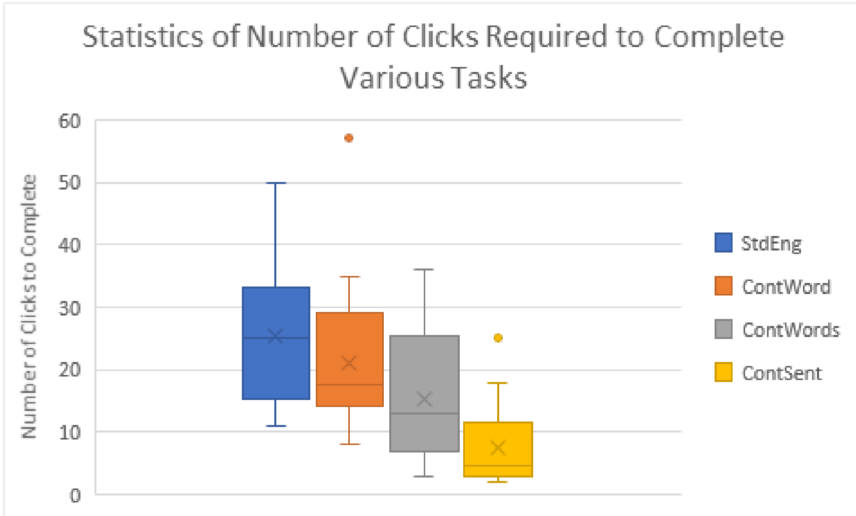

Figure 5.10: These are box and whisker plots showing the number of clicks required to complete each method as presented to the user. The 'x' in each plot represents the median value of number of clicks by all users. The interface format shorthand labeled in the legend corresponds to those outlined in Table 3.1.

Table 5.2: The average number of clicks (in clicks) to complete tasks with the four text prediction methods tested (as outlined in Table 3.1).

Method	**Mean**	**Median**
StdEng	25	25
ContWord	21	18
ContWords	15	13
ContSent	8	5

5.4 Study Observations

As the study was wrapping up for individuals, participants were also asked if there were any other comments or suggestions they had in regard to any of the technology or the study that they had completed. Study administrators collected the following responses:

- No place to rest your eyes - contemplation / creative
- I felt the keys were spaced too close. Maybe spacing them out better.
- Need more time to read longer phrases - time on short phrases is good
- The word phrase suggestions were all relevant, but not necessarily the options that my usual speaking choices would be
- I was able to control the cursor by adjusting the angle of my head and squinting, but with my lazy eye, this interface was difficult even after calibration
- If I squinted, the cursor wouldn't move as much - would prefer it if looking away and looking back restarted the timer instead of continuing from where it was
- The only problem I had when using this interface was sometimes it was hard for my eyes to stop at the right place because they moved too quickly and the interface didn't wait long enough for them to be clicked.
- My eyes tired quickly. You need to continue working out the kinks. This

> technology is for disabled people, not lazy able bodied folks. The scenarios were bizarre for quadriplegics.

Some of these items might be attributed to unfamiliarity of the interface and, more specifically, the gaze tracking device. Several of these items can be carried forward into the future development of the interface and communication system.

Administers of the study also had several observations while running it. They noticed that once participants switched to the QWERTY keyboard page of the interface, they were very unlikely to return to the main suggestions page. They also noticed that participants in the study preferred to not permit typos when inputting text into the interface; Participants would often utilize the [Backspace] button and [Clear Word] button. It was also noted that most participants would create similar phrases; few would create original phrases. For the task in which the interface provided context driven text prediction with single word, multiple word, and complete sentence and phrase suggestions, participants commented that they chose a full phrase because it was there but otherwise would have said something different. Lastly, administrators of the study noted that as the study progressed and the interface began using context driven text prediction, two participants quickly noticed and commented that the suggested text was "much better" and asked if the data set had been changed.

Chapter 6: Discussion and Future Work

6.1 Conclusions

From the results of the user study, in general, users did feel that they could communicate easily and effectively using the communication system provided to them. The responses to the comprehensive survey administered to study participants supports the claims that the users would be positive towards the interface; specifically, users did not find the interface unnecessarily complex, users thought that most other people would be able to learn how to use the interface quickly, and that they thought they did not need to learn a lot prior to using the interface. All of these things suggest that the interface is easy to use.

Participants did have some slight adverse responses to several of the questions of the comprehensive survey. It appears that the implementation of the gaze tracking device was not as refined as one would hope. Users responses to the survey suggested that some had difficulty concentrating on areas of the interface long enough to interact with them and that the mouse cursor was not as responsive as they would have preferred. These could possibly be attributed to a lack of familiarity with the gaze tracking device that could be overcome with more usage. There could also be settings in the communication system or calibration to be completed with the gaze tracking device that be better fine-tuned.

Although there are some slight trends seen from the responses to the surveys administered after each task to participants, it is difficult to draw conclusions from them alone. In general, users for most tasks felt that they could effectively, accurately, easily, and quickly communicate without feeling frustrated. When comparing the responses based on the various prediction methods, there is a slight preference towards an interface that provides context driven text prediction with single word, multiple word, and complete sentence and phrase suggestions, with the most significant trend seen in the frustration level of users.

When observing the statistics from the recorded values of time required to complete tasks and the number of clicks required to complete tasks there is noticeably strong trend comparatively between tasks. When the interface implements a context-based text prediction method we see a decrease in the time and number of clicks required to complete tasks. If the interface utilized context-based text prediction and uses single word, multiple word, and complete sentence and phrase suggestions, the time and number of clicks required is nearly one third of that required with an interface that uses standard text prediction. Several of the users even commented on their preference for the context-based text prediction without the knowledge of the difference in text prediction corpus was. This strongly supports my hypothesis that an interface that provides context driven text prediction, and one that provides complete sentence suggestions, will results in faster communication by users.

Considering the comments made by participants about the study, there is still room for improvement with the AAC system. The interface itself is full of click-able areas, meaning that a user does not have place to rest their eyes while completing communication objectives. Users mentioned that there was no way to rest and that buttons seemed too close together. These are certainly factors to keep in mind for the future. A user also commented that the amount of time granted before a click was activated by the interface was too short for being able to read longer phrases. This should be a setting that is easily configured on a per user basis moving forward.

I believe that some of the items observed from the user study would be resolved with continued use of the communication system. I believe that once users build a habit of using the suggested text screen of the interface, they will want to return to it more often after using the QWERTY keyboard screen. As users of the interface gain confidence in their usage, I expect them to be less critical of errors. As for users selecting a complete sentence or phrase just because it was there, the intention is to have the language model learn from each unique user. This would result in text suggestions that are closer to what that user would actually say.

The results of this study demonstrated very similar results to an earlier study performed in which users interacted with the interface without the use of of a gaze tracking device as input. Participants in the study used a mouse to interact with the interface. This early study reflected an overall positive reaction towards the interface itself. The study also mirrored the trend in quicker times and less interaction required when completing tasks that provide context-based text prediction, specifically, when the interface suggests complete sentences.

Overall, the results from this study are very encouraging. The results support that the interface that was developed was easy to use and efficient to communicate with. It was also found that when the interface provides context driven text prediction that includes single word, multiple word, and complete sentence and phrase suggestions, users are able to communicate faster and with less interaction with the interface. Some of the results from the study do suggest that there is more work to be done in regard to the gaze tracking device integration into the AAC interface.

6.2 Future Work

The language model that would support an AAC interface as outlined in this paper has plenty of opportunities to expand upon its domain. Language grounding plays a pivotal role in this language model. There is need for a system that can see the environment surrounding a user, and it needs to be able to interpret that into some type of meaning. We would start with determining what type of sensor platform is capable of gathering the information we need. A platform can be developed initially to gather training data in realistic organic settings to begin to better determine in what type of situations communication is required.

Once the system can see its environment, it needs to be able to process the data into some type of meanings. The system needs to be able to pick out the important things in any environment. This would need to be coupled with some type of learning system. With a supervised learning model, we could train the model to pick out the most relevant objects or events from an array of scenarios. The goal being to have the system interpret context automatically so as to invoke the right model. Ideally, once the structure of the language and learning model is developed, the system would be trained by a user while they still have the ability to speak verbally. This would allow for a custom-tailored solution, having the system select the context that is most important to each individual user.

Provided the results of the user study, it suggests that the gaze tracking device requires further study. It may be the case that with more use, users of the device will feel more confident, and will find that the responsiveness and interacting with the graphic interface becomes better and more fluid. It may also be a matter of reconfiguring certain settings, such as the time it takes to recognize a cursor click with the device. As gaze tracking technology advances, there may be other options on the market to investigate as well. The goal is to find a budget friendly, non-invasive gaze tracking device in order to produce an AAC system that users not only feel comfortable with using but want to use.

More work can be completed on the graphical user interface as well. Further study can be conducted to determine the optimal appearance and layout of the interface. It is also important to ensure the interface will operate on a proper platform that could actually be used by patients in need, such as portable tablet devices. Finally, the system needs to be integrated. The sensor platform, display for the interface, and the gaze tracking device will all need to coexist on some type of other mobile platform such as an electric wheelchair.

6.3 Other Domains

The importance of context sensitive language proven in this study, as well as the future work performed on language groundings can be applied to a wide variety of other domains. Determining the context and salience of various environments could be the keystone of many future artificially intelligent systems.

For instance, in the field of law enforcement, writing incident reports are a standard daily task for officers. If it were possible to develop an intelligent and reliable system that could survey a scene and extract salient parts of the environment, it could greatly aid the officer. The system would be able to provide a resource that could save time and produce more accurate reports.

Provided advancements in the field, on-the-fly language translators could stand to benefit from this technology as well. Providing context to a language translator model could greatly increase the accuracy in the system.

Survey and surveillance systems could also show improvements from development of this technology. Once again, the ability to pull important objects from a scene and automatically provide context for whatever the situation may be could greatly enhance the responsiveness and reliability of such systems.

This technology goes well beyond object detection and classification in images. It is the ability to assign real meaning to a scening; to be able to tell a story. Using language groundings to provide context has a nearly limitless future with a gateway to many other fields.

Appendix A: User Study Surveys

Participants of the user study were administered two types of survey. The first survey type was administered after each scenario, and was only in regards to the recently completed task. The second survey was administered to the participant after all tasks were complete, and was in regards to the overall study and experience with the interface.

A.1 Scenario Survey

1. I felt I could effectively communicate what I felt I needed to for the given situation.

   Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

2. The response I provided is exactly what I would have said in the given situation.

   Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

3. I felt frustrated using the interface to complete this task.

   Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

4. I found this task easy to complete.

Strongly Disagree Disagree Neutral Agree Strongly Agree

5. I felt I could communicate quickly using this interface.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

A.2 Overall Survey

For each of the statements below please select the choice closest matching your feelings about the statement.

1. I think that I would like to use this interface frequently (given the situation where you must communicate via text).

   Strongly Disagree Disagree Neutral Agree Strongly Agree

2. I found the interface unnecessarily complex.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

3. I thought the interface was easy to use.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

4. I think that I would need the support of a technical person to be able to use this interface.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

5. I found the various functions in this interface were well integrated.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

6. I thought there was too much inconsistency in this interface.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

7. I would imagine that most people would learn to use this interface very quickly.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

8. I felt very condent using the interface.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

9. I needed to learn a lot of things before I could start using this interface

   Strongly Disagree Disagree Neutral Agree Strongly Agree

10. I found the interface very cumbersome to use.

    Strongly Disagree Disagree Neutral Agree Strongly Agree

11. I found it hard to look at the buttons long enough for them to be clicked.

    Strongly Disagree Disagree Neutral Agree Strongly Agree

12. I didnt mind the staring at the buttons for the length of time needed to cause a click to occur.

    Strongly Disagree Disagree Neutral Agree Strongly Agree

13. I often felt the mouse cursor wasnt moving to the button I wanted to click.

    Strongly Disagree Disagree Neutral Agree Strongly Agree

14. I could easily control the mouse cursor with my eyes.

    Strongly Disagree Disagree Neutral Agree Strongly Agree

Bibliography

[1] Duffy, Joseph. Motor speech disorders substrates, differential diagnosis, and management. Elsevier. p. 295.

[2] Locked In Syndrome. Retrieved from https://rarediseases.org/rare-diseases/locked-in-syndrome/

[3] What is MND? Retrieved from https://www.mndassociation.org/about-mnd/where-do-i-start/what-is-mnd/

[4] What is ALS? Retrieved from http://www.alsa.org/about-als/ what is als.html

[5] Amyotrophic Lateral Sclerosis (ALS) Fact Sheet. Retrieved from https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets/Amyotrophic-Lateral-Sclerosis-ALS-Fact-Sheet

[6] Augmentative and Alternative Communication. Retrieved from http://vkc.mc.vanderbilt.edu/ebip/augmentative-and-alternative-communication/

[7] Augmentative and Alternative Communication. Retrieved from https://www.asha.org/public/speech/disorders/AAC/

[8] What is AAC? Retrieved from https://www.isaac-online.org/english/what-is-aac/

[9] Chennamma, H. R., and Xiaohui Yuan. "A survey on eye-gaze tracking techniques." Indian Journal of Computer Science and Engineering Vol. 4 No.5 Oct-Nov 2013: 388-393.

[10] Harnad, Stevan. "The symbol grounding problem." Physica D: Nonlinear Phenomena 42.1-3 (1990): 335-346.

[11] Matuszek, Cynthia, et al. ”A joint model of language and perception for grounded attribute learning.” In 2012 International Conference on Machine Learning (ICML), Edinburgh, Scotland, June 2012.

[12] Steels, Luc, and Manfred Hild, eds. Language grounding in robots. Springer Science & Business Media, 2012.

[13] Communication Guide. Retrieved from http://www.alsa.org/als-care/augmentative-communication/communication-guide.html

[14] Augmentative Communication. Retrieved from http://www.alsa.org/als-care/augmentative-communication/

[15] Electronic Communication Aids. Retrieved from http://www.alsa.org/als-care/augmentative-communication/electronic-aids.html

[16] Liu, S. S., Rawicz, A., Rezaei, S., Ma, T., Zhang, C., Lin, K., & Wu, E. (2012). An eye-gaze tracking and human computer interface system for people with ALS and other locked-in diseases. Journal of Medical and Biological Engineering, 32(2), 111-116.

[17] Majaranta, P., & Räihä, K. J. (2002, March). Twenty years of eye typing: systems and design issues. In Proceedings of the 2002 symposium on Eye tracking research & applications (pp. 15-22). ACM.

[18] Pannasch, S., Helmert, J. R., Malischke, S., Storch, A., & Velichkovsky, B. M. (2008). Eye typing in application: a comparison of two systems with ALS patients. J. Eye Mov. Res, 2, 1-8.

[19] Ball, L., Nordness, A., Fager, S., Kersch, K., Pattee, G., & Beukelman, D. (2010). Eye-gaze access of AAC technology for persons with amyotrophic lateral sclerosis. *Journal Medical Speech Language Pathology*, 18, 11-23.

[20] Ball, L. J., Beukelman, D. R., Anderson, E., Bilyeu, D. V., Robertson, J., & Pattee, G. L. (2007). Duration of AAC technology use by persons with ALS. *Journal of Medical Speech Language Pathology*, 15(4), 371.

[21] Carrington, P., Hurst, A., & Kane, S. K. (2014, April). Wearables and chairables: inclusive design of mobile input and output techniques for power wheelchair users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3103-3112). ACM.

[22] Fried-Oken, Melanie, et al. "Purposes of AAC device use for persons with ALS as reported by caregivers." Augmentative and Alternative Communication 22.3 (2006): 209-221.

[23] Beukelman, David, Susan Fager, and Amy Nordness. "Communication support for people with ALS." Neurology Research International 2011 (2011).

[24] Liu, Shuo Samuel, et al. "An eye-gaze tracking and human computer interface system for people with ALS and other locked-in diseases." CMBES Proceedings 33.1 (2018).

[25] Murphy, Joan. "Communication strategies of people with ALS and their partners." Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders 5.2 (2004): 121-126.

[26] Richter, Melanie, et al. "Attitudes toward communication modes and message formulation techniques used for storytelling by people with amyotrophic lateral sclerosis." Augmentative and Alternative Communication 19.3 (2003): 170-186.

[27] Yu, Haonan, and Jeffrey Mark Siskind. "Grounded language learning from video described with sentences." Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1. 2013.

[28] Dindo, Haris, and Daniele Zambuto. "A probabilistic approach to learning a visually grounded language model through human-robot interaction." Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. IEEE, 2010.

[29] Chen, David L., and Raymond J. Mooney. "Learning to sportscast: a test of grounded language acquisition." Proceedings of the 25th international conference on Machine learning. ACM, 2008.

[30] Abowd, Gregory D., et al. "Towards a better understanding of context and context-awareness." International Symposium on Handheld and Ubiquitous Computing. Springer, Berlin, Heidelberg, 1999.