Spring 5-2017

The Effect of Voice Gender and Spoken Messages in Augmented Interactions

Katrina Lapham
University of Maine

THE EFFECT OF VOICE GENDER AND SPOKEN MESSAGES IN AUGMENTED INTERACTIONS

by

Katrina R. Lapham

A Thesis Submitted in Partial Fulfillment
Of the Requirements for a Degree with Honors
Communication Sciences and Disorders

The Honors College

University of Maine

May 2017

Advisory Committee:
Jennifer Seale, PhD, CCC-SLP
Amy Booth, MA, CCC-A
Elizabeth DePoy, PhD
Christopher Grindrod, PhD
Joshua Roiland, PhD

Abstract

A speech-generating device is often implemented to aid communication for those with limited ability to produce mouth speech. Although these devices have come a long way since their initial development, there are still pervasive problems regarding augmentative and alternative (AAC) technology. These problems include communication rate, intelligibility of the synthesized voice, and the effectiveness of the synthesized speech to transfer information for a variety of interactions. Additionally, the device is responsible for portraying unique information about the augmented speaker, including their competence, individuality and identity. This investigation sought to contribute to efforts aimed at understanding the impact of computer-generated voice output in routine social interactions. Using an iPad and an AAC mobile application, the primary investigator approached 6 novel communication partners and engaged in an interaction under 3 conditions. These conditions included female speech output, male speech output, and a speech-off function. Findings suggest limited differences between gendered speech output and suggest that the speech-off condition is more efficient for information seeking interactions. More research is needed on synthesized voices to address these issues and determine future directions for AAC technology.

Acknowledgments

First, I would like to thank my advisor Dr. Jennifer Seale for going above and beyond the average advisor—meeting with me every week, responding to my seemingly endless and hourly emails, and continuing to offer the best advice. I definitely would not have made it through this process without her guidance and support, and I could not imagine a more exceptional advisor.

I would additionally like to thank my committee: Professor Amy Booth, Dr. Elizabeth DePoy, Dr. Christopher Grindrod and Dr. Joshua Roiland, for their thoughtful questions and encouragement throughout this project.

Lastly, I sincerely thank my friends and family who have always pushed me to do my best. I'm so grateful to have friends who listened to my nearly continuous complaints, while also being so involved in this thesis that they could almost present it themselves. Thank you for always allowing me to give you a very detailed play-by-play of my life.

Table of Contents

List of Tables and Figures

Introduction

Over 4 million Americans have complex communication needs and can benefit from the use of an alternative method, or modality for communication (Beukelman & Mirenda, 2013). Communication impairments can co-occur with congenital, developmental, intellectual, acquired and/or degenerative disabilities (Beukelman & Mirenda, 2013; Higginbotham, 2010). Augmentative and alternative communication (AAC) is a specialized area of practice in speech language pathology that aims to support effective communication for individuals with communication impairments. AAC methods range from supplemental strategies (e.g., letter boards, picture symbols) to high technology, sophisticated devices that are intended to serve as a replacement for mouth speech (ASHA, 2005; Beukelman & Mirenda, 2013).

For some augmented speakers, AAC solutions may consist of low-technology options such as picture communication boards or lip-reading (Beukelman & Mirenda, 2013). These low-technology options may be used as a primary mode of communication for those who are unable, or prefer not, to operate high-technology devices. At a minimum, low technology systems are commonly used as a back-up plan in the event of technological failure (Beukelman & Mirenda, 2013; Rackensperger, 2005; Robillard, 1994; Robillard, 1996; Williams, Krezman & McNaughton, 2008). The current investigation focused on use of high technology AAC in a routine social interaction.

Technological advancements have produced communication aids that are increasingly more customizable in terms of voice selection, speech generating features, and vocabulary (Williams, Krezman & McNaughton, 2008). Additionally, mobile applications that facilitate communication as an alternative to specialized devices have

become more available. Commercially available, mainstream mobile technology is often smaller and less expensive than specialized AAC devices (McNaughton & Light, 2013; Williams, Krezman & McNaughton, 2008). Communication apps are becoming increasingly available for use on iPads and smartphones, making widely used, multi-purpose devices viable options for AAC solutions. Researchers have shown that access to AAC through these mainstream devices has decreased stigma associated with AAC and increased adoption rates with augmented speakers (McNaughton & Light, 2013).

In conjunction with technological advancements, the field of AAC has made great strides since its inception (Higginbotham, 2007; Higginbotham, 2010; Williams, Krezman & McNaughton, 2008). Many improvements have led to smaller devices with more features. In addition to decreased size, the technology itself has become more efficient for communication. For example, increased storage on the device allows more messages to be saved and prevents common messages from being repeatedly retyped. Finally, speech intelligibility has improved and become less robotic, and multiple input methods have been developed to meet the needs of individuals with varying degrees of physical impairment (i.e., eye gaze technology) (Higginbotham, 2010; Williams, Krezman & McNaughton, 2008).

Yet, problems surrounding AAC technology continue to persist. Specifically, its availability, and usability in the AAC community (Higginbotham, 2010; Williams, Krezman & McNaughton, 2008). Problems commonly discussed are related to the message presentation affordances offered by AAC devices. Individuals who use AAC have suggested that communication rate, intelligibility, and lack of personalization of synthesized voices are problematic when interacting in face-to-face contexts (Beukelman

& Mirenda, 2013; Portnuff, 2006; Robillard, 1994; Robillard, 1996). Interestingly, message presentation style has been shown to influence perceptions of AAC users' social competence (Bedrosian et al., 1992; Bedrosian et al., 2003).

Bedrosian and colleagues (1992) investigated how variables such as message length, observer background and partner reauditorization affected the perceived competence of the AAC user. Third party observers, who were mouth speakers, filled out a questionnaire immediately after viewing a videotaped augmented interaction. Observers judged competence based on the following qualities: (1) the grammatical completeness of the augmented speaker's message, (2) message intelligibility, (3) the rate and accuracy of augmented messages, (4) the sociolinguistic or pragmatic skills of the AAC user, and (5) the compensatory strategies employed by both the AAC user and partner (Bedrosian et al., 1992). Researchers found that message length had an effect on perceived competence when interactions were evaluated by SLPs familiar with AAC, but novice observers offered different perceptual data. The observers unfamiliar with AAC use for communication were not as impacted by the length of the message, but appeared to be more impressed that the AAC user was able to communicate through the use of a computer (Bedrosian et al., 1992).

Augmented speakers also report other's perceptions of them change in the presence/absence of AAC use for communication. Colin Portnuff (2006), an augmented speaker, stated that a critical social mission for people with a visual disability is to "establish his or her credentials." Due to his lack of mouth speech and use of a wheelchair, his audience often automatically generalized his disability, and assumed he was mentally incompetent. For him, his laptop brought a sense of authority and normality

to the situation because his communication partners were familiar with the device and associated it with more cognitive skill (Portnuff, 2006). Although the physical presence of AAC changed communication partner perceptions, Portnuff indicated that the AAC software had room for improvement to achieve the goal of supporting communication in typical interaction settings.

One pervasive and well-documented problem that contributes to perceived competence in augmented interactions is that of communication rate. It often takes more time to compose messages externally, on an AAC device in comparison to mouth speech (Bedrosian et al., 2003; Beukelman & Mirenda, 2013; Kim, 2001). Mouth speaker speech rates can range from 150 to 250 words per minute, while augmented communication ranges 15 to 25 times slower (i.e., 2-25 WPM) (Beukelman & Mirenda, 2013; Kim, 2001). This slow communication rate influences the perception and comprehension of the synthesized speech in part due to processing demands. The phonetic makeup of synthesized speech takes more resources to process at the word level, which in turn limits the resources available to decode sentences and higher level comprehension processes (Duffy & Pisoni, 1992). Increased processing demands and slow rate of speech sometimes leads to communication breakdowns, which influences the perception of the augmented speaker (Higginbotham, 2010; Kim, 2001; Robillard, 1994). Specifically, when augmented speakers take longer to communicate, communication partners may assume that they don't know what they're talking about or that they haven't heard their message (Clark & Brennan, 1991).

Portnuff referred to email as "the great leveler" in the communication arena. The rate of typing during e-mail is of little concern because it is composed on the speaker's

own time (Portnuff, 2006). In addition, elements that often contribute information during face-to-face interactions such as mutual gaze, body positioning and gesture, are no longer at play when communication is via email (Gold, 2000). Still, Portnuff also expressed concern that the rising popularity of instant messaging and video chatting through social media has reduced the leveling effect of emailing (Portnuff, 2006). Continued research on the impact of increasing technology in various interaction contexts is needed to explore how these changes affect augmented speakers.

In addition to communication rate, message relevance also contributes to perceived AAC user competence. In another study by Bedrosian and colleagues (2003) the importance of accuracy, speed, and delivery of augmented messages relative to perceptions of augmented speakers were explored. The project involved scripted videotaped interactions between an AAC user and a sales clerk. Again, third party observers filled out a survey immediately after viewing a videotaped augmented interaction. The conditions involved: (1) using prestored messages that were delivered relatively fast but only partly relevant and (2) slowly delivered messages with completely relevant context. In both conditions messages were delivered with and without a conversational "floor holder", such as "please wait while I construct my message."

Using prestored messages is often a strategy employed by augmented speakers to reduce the communication rate gap (Beukelman & Mirenda, 2013). However, there are trade-offs if the saved messages do not exactly match the interaction at hand (Clark & Brennan, 1991). Results from Bedrosian and colleagues' (2003) study, showed a mouth speaker preference for augmented speakers using a slowly delivered message with a conversational floor holder (Bedrosian et al., 2003). This may have been due to observers

not having time constraints while watching the videos, or actively participating in the interaction. In natural interactions, it may benefit the augmented speaker to use a combination of both prestored and generative messages. The current investigation will utilize this combination strategy in a structured task that reflects a common everyday interaction.

Conversational floor holders are commonly used to introduce the device and its purpose at the beginning of interactions, whether it is in face-to-face contexts or on the telephone (Portnuff, 2006; Rackensperger et al., 2005). Beginning an interaction by introducing the device notifies the communication partner that the device is used as a communication tool, which is helpful because the general public often lacks experience with AAC technology use for face-to-face interactions (Rackensperger et al., 2005; Williams, Krezman & McNaughton, 2008).

Intelligibility

Intelligibility of AAC devices is comparable to mouth speech in quiet environments, but noisy environments lead to decreased intelligibility of synthesized speech (Duffy & Pisoni, 1992). Research on intelligibility of AAC devices began in the mid-1980s, and several broad results were discovered (Higginbotham, 2010). Research suggested that mouth speech was only slightly more intelligible than synthesized speech. In some instances, AAC devices were found to be more intelligible when using the male voice rather than female or child voice. Finally, intelligibility of synthesized voices is highly affected by environmental conditions (Higginbotham, 2010).

Drager & Reichle (2001a) investigated how divided attention tasks impacted comprehension of mouth speech and synthesized speech. Prior research indicated that

listeners were able to comprehend synthesized speech in quiet, non-distracting environments. However, in interactions in which the environment was noisy and distracting, the comprehension of synthesized speech was negatively impacted (Duffy & Pisoni, 1992). Drager and Reichle showed that when listeners were distracted, comprehension of the synthesized speech was more negatively impacted than in divided attention conditions with mouth speech (Drager & Reichle, 2001a). This may be due to mouth speech having far more nonverbal qualities that contribute to comprehension. For example, during mouth speech the listeners may lip-read, evaluate facial expressions, interpret tone and prosodic elements, and recognize gestures to better understand the speaker (Drager & Reichle, 2001a; Drager & Reichle, 2001b).

Many visual and suprasegmental cues are not available with synthetic speech (Drager & Reichle, 2001a; Drager & Reichle, 2001b). However, comprehension of high-quality synthesized speech was equal to mouth speech in ideal (i.e., quiet) listening situations, which suggests that synthesized speech is intelligible but processing demands become too great for the listener in noisy, distracting environments. In noisy environments, it may be beneficial for augmented speakers to forgo the use of the voice function and share information through text messages displayed on their device display screen. This strategy would also circumvent the perceptions that result from intelligibility of synthesized speech. Additionally, text based information exchange may reduce the need for repetition of an uncomprehended message (Portnuff, 2006; Rackensperger, 2005).

Offering contextualized information is another strategy that may increase comprehension of synthesized speech (Drager & Reichle, 2001b). Research shows that

establishing context may make up for the processing demands of synthesized speech and would require fewer resources to comprehend the message. The intelligibility of sentence length utterances is higher than single-word utterances due to contextual cues provided in the surrounding words (Drager & Reichle, 2001b). Context builds throughout an information-seeking interaction, as each communicative contribution prepares the communication partner for understanding content provided in later contributions (Brennan, 1998; Clark, 2002; Clark & Brennan, 1991; Schegloff, Jefferson, & Sacks, 1977). As a result, the need for additional resources decreases (i.e., processing, attention) and ultimately increases intelligibility (Drager & Reichle, 2001b).

Attention demands imposed by the device for the augmented speaker is another problem. It can be difficult to tell when an augmented speaker is preparing a message due to the restricted eye contact and the communication partner's limited knowledge of what the AAC user is doing on their device. This is especially true for transient communication contexts, such as passing somebody in a store or a fast-paced public setting (Robillard, 1996). Many AAC users work around this difficulty by storing frequently used messages on their devices. Pre-stored phrases allow the AAC user to produce utterances at a faster rate and maintain eye contact. When pre-stored messages are not relevant, feedback offered in the form of audible beeps while the AAC user types a generative message could indicate to the communication partner that a message is being constructed. After the AAC user has shared their message, they may resume eye contact in order to signal communication turns and better prepare the communication partner to receive a message.

Identity & Voice

Intelligibility and communication rate are not the only concerns with voice output amongst augmented speakers. The issue of identity and individuality is another area that has received more attention in recent years and continues to require more research (Mills, Bunnell, & Patel, 2014). Synthesized speech has been reported to lack "naturalness" when compared with mouth speech, and is devoid of individual variations in voice quality, tone, volume and emotional prosody (Higginbotham, 2010; Jreige, Patel, & Bunnell, 2009; Portnuff, 2006; Wickenden, 2011). Oftentimes, a classroom with multiple students who use AAC devices use the same adult synthesized voice (Jreige, Patel, & Bunnell, 2009; Mills, Bunnell, & Patel, 2014). This mismatch may limit the adoption of the device as an extension of the individual, both by the user and communication partners (Jreige, Patel, & Bunnell, 2009). The inability to have an individual voice and unique form of expression furthers the divide between the AAC user and mouth speakers. Although there are more options for voice selection than in the past, the vast majority of synthesized voices are not representative of the user's individual vocal qualities or personality (Jreige, Patel, & Bunnell, 2009). Limited research has been done to determine how voice qualities impact communication.

Anecdotal evidence that suggests the gender of the synthesized speech may affect intelligibility. In situations where an AAC user will be frequently communicating with a person with even a mild hearing loss, a male voice may suit their needs better even if they are female (Dietz et. al, 2013; Portnuff, 2006). Many people develop a hearing loss in the higher frequencies as they age (Stach, 2008). Women's voices are typically a higher

frequency than male voices, leading to a decreased ability to understand female speech in comparison to male speech in the case of a hearing loss.

Recent efforts to personalize synthesized voices for augmented speakers using residual vocal abilities combined with a matched healthy donor speaker have gained ground (Jreige, Patel, & Bunnell, 2009). For example, it is common practice for people with progressive diseases, such as ALS, to bank their own voice for future AAC use, preserving the identity qualities portrayed through their voice. Vocalid, a company founded in 2014, uses this technique to create unique, personalized voices for AAC users. Researchers found that listeners could transcribe samples of the personalized speech with an accuracy rate of 94%, and were also able to identify the samples with a specific speaker about 80% of the time (Jreige, Patel, & Bunnell, 2009). Not only is this individualized option highly intelligible, it also allows the augmented speaker to have a unique voice that represents their age, gender and personality.

Summary

The development of synthesized speech has a long way to go in emulating mouth speech. Adjusting communication rate to resemble mouth speech may be difficult or impossible when communicating through a computer. However, voice quality is a solvable obstacle that could lower processing demands, enhance comprehension and offer AAC users a greater sense of individuality through their AAC device. Presentation style, both in terms of message formatting and voice selection, should be studied to determine which styles are most successful in various situations. AAC users, practitioners, and product developers should be aware of strategies that have been successful such as using a conversational floor holder or combining prestored and generative messages.

Several factors of message presentation have contributed to perceptions of augmented speakers. AAC devices contribute to other's presumptions of character, competence and identity of augmented speakers. Devices should allow the user to take part in interactions as effortlessly as they would using mouth speech. Synthesized voice should be as authentic as possible, and represent an individual rather than a group of people who use the same voice. The AAC device voice output should reflect the individual user, allow access to prestored and generative message input, have message output strategies that match the environment (speech volume, legible text, font size) and be intelligible to communication partners. This investigation seeks to contribute to efforts aimed at understanding the impact of manipulating computer-generated voice output features in routine social interactions. Specifically, this project aims to address the following questions:

1. To what extent does interaction duration, message repetition, or repair frequency change relative to the gender of the voice output, and/or absence of speech output, during an interaction between an augmented speaker and a library clerk?
2. To what extent do library clerk's perceptions of interaction success and augmented speaker's performance change relative to gendered voice output or speech-off conditions?

Methods

To address the previous questions, structured interactions involving the primary participant acting as an augmented speaker were analyzed. The independent variables in this study included three voice conditions: female speech output, male speech output, and speech off. The dependent variables included: duration of interaction, augmented speaker contributions, need for repetitions, whether partner's read off the iPad screen, abandoned utterances and perceptual data collected from survey results. Each independent variable was tested two separate times during scripted interactions with novel communication partners. Data was collected throughout the interaction using screen-recording technology; following the interaction participants completed a survey.

Participants

The primary investigator was the main participant, who represented herself as a student who used AAC and was seeking information for a research project. In this role, the primary investigator communicated as an augmented speaker, using only the AAC device to request information. When the interaction was complete, the primary investigator returned to her natural communication modality as a mouth speaker to debrief communication partners and collect survey data. The primary participant was 21-year-old female with an undergraduate level education in the Department of Communication Sciences and Disorders at the University of Maine. She had no hearing, vision, language or cognitive deficits.

Communication partners in this study were information desk clerks in three public libraries (Orono Public Library, Bangor Public Library, Fogler Library). For the purposes of this study, partner's age, gender and familiarity with AAC was not recorded. No

interaction data was recorded for communication partners (i.e., number of turn contributions), but they did offer perceptual ratings of their experiences interacting with AAC technology. Information desk clerks were ideal candidates for their approachability in public settings, and the naturally occurring interaction possibility. Data collection sessions occurred in a university library, urban public library and a rural public library. Data analyzed in this study was collected from six interactions with seven different communication partners; one interaction involved two desk clerks. Hearing and vision of the information desk clerks were judged to be functional for conversational speech in quiet environments.

Materials

Data for this analysis was collected through a structured communication task with unfamiliar communication partners using an iPad Pro with Predictable Speech application (version 5.0.3) and attached Logitech case which provided access via keyboard input. The augmented speaker for this investigation used the Predictable Speech application, and implemented two pre-stored phrases. When pre-stored phrases were exhausted, she used the keyboard to generate the unprepared, or novel, responses.

A Likert scale survey was developed to collect follow-up data after the interaction. Each communication partner filled out the anonymous survey and placed it in the envelope provided by the primary investigator. The survey was adapted from the surveys used by Bedrosian and colleagues (1992, 2003), in which third party observer perceptions of augmented speakers were analyzed. However, the current study differed from previous studies in that it requested participants who engaged in the interaction themselves, to rate their experience. The Likert scale survey in the current study included

8 statements and asked the desk clerk to rate the statements on a 7-point scale from "strongly disagree" to "strongly agree" (see Appendix B). The survey aimed to evaluate the communication partner's assessment of the interaction. Questions prompted responses about computerized voice quality, speed of communication, voice volume, font size, and the success of the interaction overall.

Procedure

In order to prepare for the experimental interactions, the primary investigator practiced communicating via AAC in private (5 hours) and public settings several times (i.e., 3-5 different times) prior to the study. Practice interactions included places such as pizza shops, cafes, and at home practice with familiar partners. Different techniques such as using primarily pre-stored phrases, using the attached keyboard versus the touch screen keyboard, predictive text, and/or timing of message output were tested. During these practice sessions, several voice options were used, and pronunciation of words expected to come up in the experimental sessions were tested. Operational proficiency was gained through four weeks (6-8 cumulative hours) of practice with different device settings both at home and in public settings. Skills gained included using the attached keyboard, turning off word prediction, turning on/off volume, running screen capturing software in the background of speech application, and using the "speech" button to speak the message. In addition, the primary investigator became familiar with nuances of the Predictable Speech application. Specifically, learning that the messages could remain on the page until purposefully deleted by the primary investigator, rather than be deleted as a function of the application after speaking the message. The primary investigator chose

this option because it allowed for quicker repetition of the phrase if requested by the communication partner.

Prior to data collection the primary investigator entered the library and connected to Wi-Fi in order to access the screen capturing application (AirShou). No interaction with library staff was required to connect to the public Wi-Fi. Additionally, the speech output condition (male/female, or speech-off) was selected, and the volume was adjusted to an appropriate level. Next, recording the iPad display was initiated with the screen capturing application. In the role of augmented speaker, the primary investigator approached an information desk clerk. The interaction was initiated by the primary investigator issuing two prestored messages; one introducing the device and requesting assistance, the second specifying the assistance needed (see Appendix A). If the conversation about the desired book required less than three contributions from the augmented speaker, additional questions were asked such as, additional assistance in finding related research articles, and/or clarifying the library's business hours. After the interaction was complete, the primary investigator "broke character," and debriefed the communication partner. After obtaining consent, the investigator asked them to complete the Likert scale survey; survey data was obtained from 6 of 6 experimental sessions.

The post interaction debriefing included an explanation that the actual study was to explore the effects of gendered voice and speech output during augmented conversational interactions. The communication partner was also notified that the data would primarily consist of the duration of the interaction, and the AAC users repairs and repetitions. They were told that their responses were not recorded, as their primary role was to provide a vehicle to collect data about the use of AAC to mutually construct an

interaction. The analysis consisted of the number of turns the AAC user took in the typical interaction, the repairs or repetitions provided, and the duration of the interaction; recording and transcription of the responses from the communication partner was beyond the scope of this investigation. The entire experimental session, beginning with entering the library and ending with the survey collection, averaged 5-7 minutes per session.

Immediately following the experimental session, screen recordings were exported from the application to the iPad photo album, and subsequently downloaded to a laboratory computer in room 304 Dunn Hall on the UMaine campus for transcription and further analysis. The primary investigator also noted relative noise level in different libraries, and if the communication partners appeared to read off the iPad display while she typed.

Data Analysis

The screen capturing software recorded interaction time and augmented speaker contributions; abandoned utterances were deduced from the message displayed on the iPad screen. The video recordings were transcribed to examine the number of turn exchanges, generative phrases, and abandoned utterances. Survey results were analyzed to determine if gendered output or text-only communication influenced partner's perceptions of: intelligibility, speech rate, text size, and overall success of the interaction (see Appendix B). Raw counts of participant ratings for the individual questions obtained from the survey data were analyzed.

Interaction duration was determined as the time from which the primary participant approached the information desk clerk to when the primary investigator broke "character." Augmented speaker contributions were recorded as the number of times the

augmented speaker shared a message. Abandoned utterances were not counted towards this total. For the purpose of this study, an abandoned utterance symbolized an instance when the augmented speaker would write a message, and then choose not to share the message with the communication partner. These objective measures were visually inspected for differences across conditions.

Reliability

Initial practice sessions aimed to ensure operational competence and comfort level while using the device for the primary investigator. In order to determine intra-rater reliability the primary investigator re-watched and re-transcribed 33% of the interaction videos. After comparison between the first and second transcriptions reliability was determined to be 100%. Intra-reliability was high because the video recordings primarily involved the written display of the augmented speaker's message during the interactions; transcription included simply rewriting the text.

Results

Results recorded from the structured interactions are reported in Figure 1, Table 1 and Table 2. Variables reported in this investigation include: interaction time, augmented speaker contributions, repetitions, abandoned utterances, and survey responses. Due to the small sample size, results reported below reflect observations from inspection of raw data; no statistical analyses were performed.

The recorded interaction duration averaged three minutes across conditions. Interactions were longest with the female voice with an average of 215.5 seconds (3.6 minutes) (see Figure 1). The second longest interaction time was the male condition with an average interaction time of 187 seconds (3.1 minutes). Shortest was the speech off condition with an average interaction time of 170.5 seconds (2.8 minutes). The Orono Public Library interaction had the longest interaction time at 264 seconds (4.4 minutes). The Bangor Public Library averaged 176 seconds (2.9 minutes) over two interactions, and the University library averaged a time of 176 seconds (2.9 minutes) over three interactions.

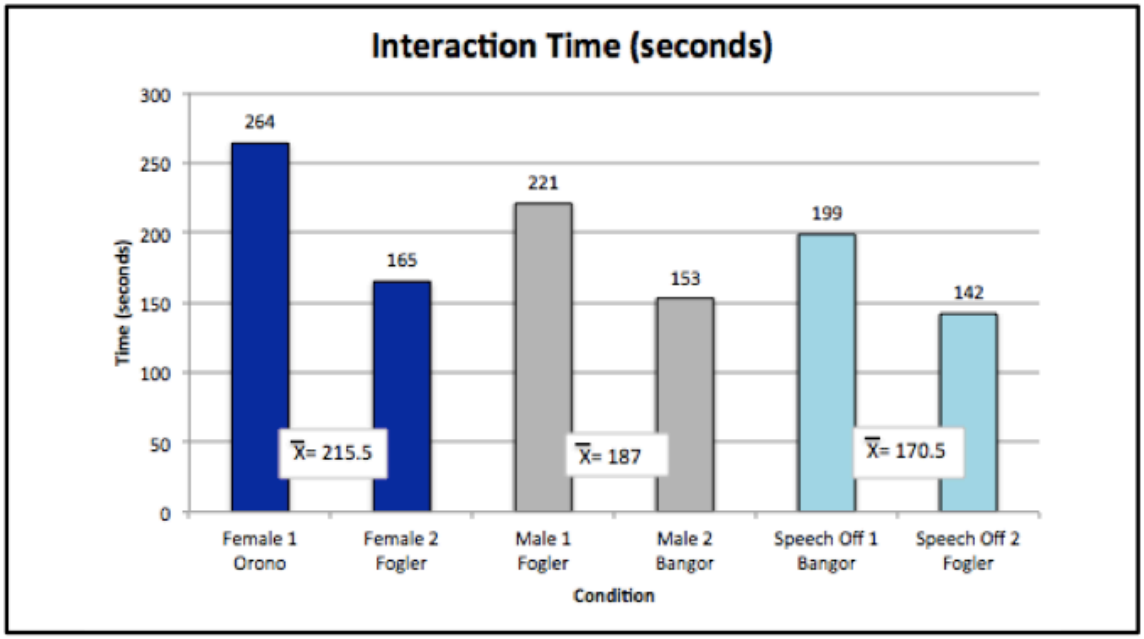

Figure 1. Changes in interaction times across conditions in seconds. Female 1 condition at the Orono Public Library was longest; speech off 2 at the University Library was shortest.

The number of contributions from the augmented speaker revealed that the male (6) and female voice condition (5.5) involved the most contributions. The speech off condition had the fewest amount at an average of 5 (see Table 1). Each interaction had a minimum of two contributions because of the two pre-stored messages; the total contribution count is two plus the additional messages typed by the augmented speaker. The remaining dependent variables in this investigation achieved no differentiable results regardless of the condition. Specifically, repetition did not occur in any of the structured interactions, and communication partners appeared to read the screen regardless of the voice condition.

Table 1. The frequency of contributions, including the novel messages produced by the augmented speaker across conditions.

Condition	**Total Contributions**	**Augmented Speaker Generative Contributions**
Female 1	5	3
Female 2	6	4
Male 1	6	4
Male 2	6	4
Speech Off 1	5	3
Speech Off 2	5	3

Abandoned utterances occurred just 3 times across all experimental sessions. Two of these abandoned utterances took place in the University library, and one took place at the Bangor Public Library. This primarily took place when the communication partner changed the subject, answered the partially typed question or the communication partner walked away, unaware that there was an impending message. Two abandoned utterances occurred in the male speech output condition, and the only other abandoned utterance occurred in the speech off condition. Interestingly, the female conditions did not have any abandoned utterances.

While just six interactions were officially recorded for interaction analysis, survey data included eight total respondents. The additional surveys resulted from multiple desk clerks participating in one structured interaction. For example, during one interaction two desk clerks participated in the interaction and both filled out the Likert scale survey. In a different exchange, technological issues led to a survey being collected without the accompanying video. This data collection session had to be repeated at a later date in order to collect quantitative data for the specific condition. Qualitative information was

retained from all interactions in an effort to collect data that was representative of partner's perceptions of the primary investigator as an augmented speaker.

The results of the survey were favorable across conditions. Partner ratings for each of the questions ranged from slightly agree to strongly agree (Table 2). Participants indicated minimal misunderstanding as a result of AAC use, and 100% "strongly agreed" that the augmented speaker was a successful communicator. In addition, participants indicated that the speed of communication was neither too fast nor too slow, that the voice was easy to hear, and that the text/font size was easy to read (see Table 2).

Table 2. Survey response data.

	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree
1. Katrina expressed herself clearly						2	6
2. Katrina was easy to understand						2	6
3. The speed of communication was just right (neither too fast nor too slow)					1	4	3
4. Katrina's computer voice was easy to hear						2	3
5. Katrina's text/font size was easy to read						3	4
6. Katrina was successful in getting her point across							8
7. The way Katrina communicated resulted in some misunderstandings; I needed her to clarify her message	4	1		1			
8. Katrina was easy to talk to					1	1	6

Summary of Results

The speech off condition resulted in the least amount of interaction time and included the fewest augmented speaker contributions. The male and female conditions involved the most augmented speaker contributions and longest interactions. The female voice interaction took 45 seconds longer than the speech-off condition, and nearly 30 seconds longer than the male condition. Interaction times and augmented speaker contributions were the only observed differences across conditions. No repetitions or repairs occurred in any of the interactions. Communication partners appeared to read off the iPad screen regardless of speech conditions. All participants indicated they had positive experiences with and perceptions of the primary participant as an augmented speaker.

Discussion

Augmentative and alternative communication (AAC) is designed and implemented with the goal of supporting individuals with communication impairments to communicate effectively and efficiently. The relatively consistent interaction times and number of augmented speaker contributions indicates an effective communicative exchange between the primary participant and communication partners across all conditions. The absence of repetitions and repairs further strengthens that AAC was a successful method of communication for the study's interaction setting. In addition to the objective results, favorable survey results confirm that the AAC methods used in the structured interaction context of asking for library assistance were successful.

The speech-off function in the current study was arguably the most efficient because of its short interaction duration and low contribution count. These findings suggest that there may be communicative situations in which AAC use would be more effective with a text-only function. This may be especially true during information seeking interactions where the communication partner can read the sought after material from the AAC device screen. Other examples of informational exchanges may include restaurant settings, business transactions or medical related discussions, where the focus is on the efficient exchange of information rather than the personality of the AAC user. Additional study is warranted to determine if voice customizability may be less important or even exchanged for a text-only communication method in these types of information based interactions.

Communication rate differences were not identified by the communication partner, as indicated by the survey results. The lack of variation in augmented speaker

contributions across conditions further suggests that rate differences were not noteworthy. The strategy of beginning the interaction using two pre-stored messages may have contributed to this rate finding. In addition, the lack of communication rate complaints reflected in the survey may be because the augmented speaker contributions were so contextualized. Information seeking interactions have more structure than interactions for sociability purposes because the quest for information makes it easier to determine the end of the conversation (Bedrosian et al., 1992; Bedrosian et al., 2003; Beukelman & Mirenda, 2013). During the structured interactions the communication partner was aware that the augmented speaker was searching for a specific book, and that the conversation would likely end after a solution was offered. Because there was an end goal to the interaction, the communication partner knew to expect a message until they answered the requests of the AAC user.

Contextualized information also increases comprehension of synthesized speech (Drager & Reichle, 2001b). As the context of the interaction increased, the communication partner was prepared for later contributions from the augmented speaker. The narrow context of the current study may explain the consistent survey results describing the intelligibility of the synthesized voice, along with the lack of requests for repetitions. The communication partners were more able to guess at words they may have had trouble understanding if they did not have surrounding context. These results replicate previous findings that context enhances intelligibility and comprehension.

There were no results that directly suggested different intelligibility levels between gendered voices, despite previous research that suggested female speech might be slightly less intelligible than male speech (Dietz, et al., 2013; Portnuff, 2006). Slightly

longer interaction times and an increase in augmented speaker contributions observed in the female voice condition may have indicated decreased intelligibility, but are not significant enough to definitively attribute these differences to intelligibility problems. The lack of intelligibility issues may be because all communication partners in the study were determined to have functional hearing for conversational speech. In addition to adequate hearing and highly contextualized information, the fact that communication partners read off the iPad screen contributed to their understanding of the message. The longer interaction time for female voice may be because one female condition was tested in the Orono Public Library, and the desk clerk left the interaction to physically look for the book, spending more time than if they had used a computerized system.

Finally, the current study also involved a high-quality voice that enhanced intelligibility. If the structured interactions were completed with a lower quality voice there may have been more difficulties with intelligibility. Synthesized voice has improved dramatically in comparison to earlier technology. More recent synthesized voices have increased intelligibility and sound less robotic than older models (Higginbotham, 2010). Although intelligibility has increased, continued efforts to improve vocal quality to portray emotional content and personal identity are needed.

Identity with the synthesized voice did not appear to be an issue during this study. Interestingly, the communication partners did not indicate problems with the female primary participant communicated using a male synthesized voice. The lack of congruence between the AAC user and synthesized voice limits the communication partner's ability to recognize the device as an extension of the augmented speaker (Jreige, Patel, & Bunnell, 2009). The shortage of response to this mismatch during the current

study suggests that concerns about the synthesized voice matching the user's identity may be restricted to friends, family and the AAC user themselves (Higginbotham, 2010; Robillard, 1996; Robillard, 2003). This may be because the communication partner in this study was primarily concerned with responding to a request for information, rather than trying to connect with the augmented speaker as an individual. Concerns about the individuality of the synthesized voice may have been more prevalent in exchanges other than information seeking interactions. Additional research, in which partners experience both voice output conditions and are asked directly about preferences, is warranted.

Limitations and Future Directions

Findings from this study cannot be generalized to a larger population due to a small sample size. In addition, these observations were based on visual inspection of the data. More data is necessary to generalize these results. A longer interaction would allow for a more thorough investigation of the components of augmented speaker interactions. Future studies should address augmented interaction in a variety of contexts; some of which may be designed to last for more communicative exchanges, be implemented in different environments, and/or use actual augmented speakers to build a more complete picture of the different effects between each voice. In addition future studies could build a more complete picture of the interaction by collecting data from both the augmented speaker and the communication partner.

Unfortunately, each environment was not represented equally. Originally, the conditions were to be tested in each location a total of two times, but due to the Orono Public Library having a small number of desk clerks on staff a decision was made to test the condition in the University location several times rather than completing the

interaction with the same staff member twice at the Orono Public Library. A tradeoff in testing in several different locations is that each location adheres to its own practices of searching for information and answering questions, which may affect the interaction time of each condition.

High competence ratings in this study may be influenced by the primary investigator "breaking character" at the end of the interaction to debrief the desk clerk. By using mouth speech rather than continuing the interaction using augmentative communication, the primary investigator may have influenced perceptions of her that may have been different if she continued using AAC. Future investigations should consider recruiting a third party ally who can provide the debriefing and distribute the survey. Having a third party ally disperse and analyze the participant survey would also eliminate any elements of bias caused by having the primary participant soliciting comments on her performance and also interpreting the results.

Additionally, the fact that the primary participant was not an authentic augmented speaker may have affected from the communication partner perceptions. There were dramatic differences in regards to attitudes about the augmented speaker when contrasted with Bedrosian's study. Participants in Bedrosian's (1992) study commented that they were impressed that the AAC user was able to communicate at all, given their physical limitations. Perhaps the participant's impressions during the current study would have been different if the primary participant appeared "disabled" or if the AAC device were attached to a wheelchair or an object that holds a stereotype of disability.

Although this study suggested that identity with the synthesized voice was not a priority in this particular setting, it doesn't mean that representative voices are not

important. The current study did not reveal any stark differences between male and female synthesized voice. Future research is warranted to determine if these findings would apply to other brands of synthesized voices, and to repair differences if they were to arise. Without differences between male and female voice, augmented speakers would be free to select a voice that best represents their personality and individuality.

Conclusion

In this study, interaction time and augmented speaker contributions were the only differences observed between conditions. The secondary participants rated all interactions favorably. No participants expressed concern or took notice of when the primary participant used a mismatched voice gender. Future research should explore if different interaction settings warrant different levels of identity with the synthesized voice. For example, an interaction more focused on personal characteristics of the individual, such as conversations with friends.

Short interaction times and reduced augmented speaker contributions observed in the speech-off condition suggest text-based interaction may be the most efficient for information seeking tasks. More research regarding the differences between gendered voices, in addition to possible benefits of communicating with a speech-off function is warranted. Augmented speakers have expressed that they feel a lack of identity with their synthesized voice, especially when several people in the same classroom use the same voice (Jreige, Patel, & Bunnell, 2009; Mills, Bunnell, & Patel, 2014; Portnuff, 2006). The inability to have unique vocal qualities to represent an individual's personality limits the adoption of the AAC device as an extension of the individual and creates a divide between AAC users and mouth speakers. More research on gendered speech and qualities

of synthesized speech in general could provide augmented speakers with a personalized way to express themselves.

References

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality progress*, *40*(7), 64.

American Speech-Language-Hearing Association. (2005). *Roles and responsibilities of speech-language pathologists with respect to augmentative and alternative communication: position statement* [Position Statement]. Available from www.asha.org/policy.

Bedrosian, J. L., Hoag, L. A., Calculator, S. N., & Molineux, B. (1992). Variables Influencing Perceptions of the Communicative Competence of an Adult Augmentative and Alternative Communication System User. *Journal of Speech Language and Hearing Research, 35*(5), 1105-1113.

Bedrosian, J. L., Hoag, L. A., & McCoy, K. F. (2003). Relevance and Speed of Message Delivery Trade-Offs in Augmentative and Alternative Communication. *Journal of Speech, Language, and Hearing Research, 46*(4), 800-817.

Beukelman, D. R., & Mirenda, P. (2013). *Augmentative & Alternative Communication: Supporting children & adults with complex communication needs* (4th ed.). Baltimore: Paul H. Brookes Pub.

Brennan, S. E. (1998). The grounding problem in conversations with and through computers. *Social and cognitive approaches to interpersonal communication*, 201-225.

Clark, H. H. (2002). Speaking in time. *Speech communication*, *36*(1), 5-13.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, *13*(1991), 127-149.

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-type scales in the social sciences. *Journal of Adult Education*, *40*(1), 19.

Dietz, A., Thiessen, A., Griffith, J., Peterson, A., Sawyer, E., & McKelvey, M. (2013). The renegotiation of social roles in chronic aphasia: Finding a voice through AAC. *Aphasiology*, *27*(3), 309-325.

Drager, K., & Reichle, J. (2001a). Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication*,

17(2), 109-119.

Drager, K. D., & Reichle, J. E. (2001b). Effects of Discourse Context on the Intelligibility of Synthesized Speech for Young Adult and Older Adult Listeners Applications for AAC. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1052-1057.

Duffy, S. A., & Pisoni, D. B. (1992). Comprehension of synthetic speech produced by rule: A review and theoretical interpretation. *Language and Speech*, *35*(4), 351-389.

Garland, R. (1991). The mid-point on a rating scale: Is it desirable. *Marketing bulletin*, *2*(1), 66-70.

Gold, G. (2000). Meaning of a Disability: The Lived Experience of Paralysis by Albert B. Robillard. *Body & Society, 6*(2), 103-108.

Higginbotham, D. J., Shane, H., Russell, S., & Caves, K. (2007). Access to AAC: Present, past, and future. *Augmentative and Alternative Communication, 23*(3), 243-257.

Higginbotham, D. J. (2010). Humanizing Vox Artificialis: The Role of Speech Synthesis in. Computer Synthesized Speech Technologies: Tools for Aiding Impairment: Tools for Aiding Impairment, 50-66.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: explored and explained. *British Journal of Applied Science & Technology*, *7*(4), 396.

Jreige, C., Patel, R., & Bunnell, H. T. (2009). VocaliD: Personalizing text-to-speech synthesis for individuals with severe speech impairment. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility* (pp. 259-260). ACM.

Kim, K. E. (2001). *Effect of speech rate on the comprehension and subjective judgments of synthesized narrative discourse*.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study. *Educational and psychological measurement*, *31*, 657-674.

McNaughton, D., & Light, J. (2013). The iPad and mobile technology revolution: Benefits and challenges for individuals who require augmentative and alternative communication.

Mills, T., Bunnell, H. T., & Patel, R. (2014). Towards personalized speech synthesis for augmentative and alternative communication. *Augmentative and Alternative Communication*, *30*(3), 226-236.

Portnuff, Colin. "AAC: A User's Perspective". 2006. Presentation.

Rackensperger, T., Krezman, C., Mcnaughton, D., Williams, M. B., & D'silva, K. (2005). "When I First Got It, I Wanted to Throw It Off a Cliff": The Challenges and Benefits of Learning AAC Technologies as Described by Adults who use AAC.*Augmentative and Alternative Communication, 21*(3), 165-186.

Robillard, A. B. (1996). Anger In-the-Social-Order. *Body & Society, 2*(1), 17-30.

Robillard, A. B. (1994). Communication problems in the intensive care unit.*Qualitative Sociology, 17*(4), 383-395.

Stach, B. (2008). Clinical audiology: An introduction. Nelson Education.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 361-382.

Wickenden, M. (2011). Talking to teenagers: Using anthropological methods to explore identity and the lifeworlds of young people who use AAC. *Communication Disorders Quarterly*, *32*(3), 151-163.

Williams, M. B., Krezman, C., & Mcnaughton, D. (2008). "Reach for the Stars": Five Principles for the Next 25 Years of AAC. *Augmentative and Alternative Communication, 24*(3), 194-206.

Appendix A: Sample Script Narrative

Approaching employee at information kiosk in library. Using device with voice output or written text, pending condition.

Katrina (participant): Hello, my name is Katrina and I use my iPad to help get my message across. I'm doing a project for school, would you be willing to help me find some information?

Communication Partner: Yes (continue)/No (discontinue)

Katrina (participant): I'm doing a research project about autism. I'm looking for the book *Thinking in Pictures* by Temple Grandin.

Communication Partner: (Partner response)

Katrina (participant): Do you have the book here?

Communication Partner: (Partner response)

Katrina (participant): Is it possible for you to help me access the textbook through interlibrary loan system? Or should I try to purchase it?

Communication Partner: (Partner response)

Katrina (participant): Okay. Thank you. Also, I'm not very familiar with looking up peer-reviewed journal articles. Do you have suggestions for how to get started with that?

Communication Partner: (Partner response)

Katrina (participant): Okay this feels like a great start. Thank you for your help.

Breaking participant "character" Katrina will now debrief information kiosk employee and ask him/her to complete a brief survey about the interaction they just participated in using the following script.

Katrina (researcher): Thank you for taking the time interact with me and assist me with my school project. I'm actually a UMaine student conducting a research study that looks at how this type of technology influences communication, primarily through exploring how individuals respond to somebody communicating using this kind of technology. You were really helpful in answering my questions, and mostly what I am looking at is how long our conversation was and the kinds of information I needed to repeat / clarify. Would it be okay if I used my observations for my research?

Communication partner: Yes / No (discontinue and thank again for time)

Katrina (researcher): I also have a brief 8-question survey asking for your input about your experience interacting with me as a non-speaking individual using technology. The answers you provide will remain anonymous. In addition, I will be destroying the anonymous data in December 2017. Until then, the data will be stored on a secure computer and a locked drawer. Would you be willing to answer the questions?

Communication partner: Yes (give survey)/No (discontinue and thank again for time)

Katrina (researcher): Great thank you. You do not need to put your name on the survey, please only answer the questions that apply to our interaction, and when you're finished you can fold it in half and place it in this envelope for me. I won't be reviewing the surveys until I've collected all of my data; this way you can remain anonymous. If you would like to see the final results of the study you may contact my faculty advisor. Here is a card with contact for myself, my faculty advisor, and the IRB contact person if you have any questions or concerns (see Appendix C).

Give survey and thank again for time when complete.

Appendix B: Survey Questions

Are you at least 18 years old? ☐ YES ☐NO

	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree
Katrina expressed herself clearly	1	2	3	4	5	6	7
Katrina was easy to understand	1	2	3	4	5	6	7
The speed of communication was just right (neither too fast nor too slow)	1	2	3	4	5	6	7
Katrina's computer voice was easy to hear	1	2	3	4	5	6	7
Katrina's text/font size was easy to read	1	2	3	4	5	6	7
	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree
Katrina was successful in getting her point across	1	2	3	4	5	6	7
The way Katrina communicated resulted in some misunderstandings; I needed her to clarify her message	1	2	3	4	5	6	7
Katrina was easy to talk to	1	2	3	4	5	6	7

Appendix C: Contact Information

If you have any concerns or additional questions about participating in this study please do not hesitate to contact the following people.

Principal Investigator:
Karina Lapham
Email: Katrina.lapham@umit.maine.edu

Faculty Advisor:
Jennifer Seale, Ph.D., CCC-SLP
Office: 343 Dunn Hall
Phone: 207-581-2036
Email: jennifer.seale@maine.edu

IRB Contact Person:
Gayle Jones
Office: 418 Corbett Hall
Phone: 207-581-1498
Email: gayle.jones@umit.maine.edu

Appendix D: IRB Approval Letter

APPLICATION FOR APPROVAL OF RESEARCH WITH HUMAN SUBJECTS
Protection of Human Subjects Review Board, 418 Corbett Hall, 581-1498

(Type inside gray areas)
PRINCIPAL INVESTIGATOR: Katrina Lapham
EMAIL: katrina.lapham@maine.edu TELEPHONE: 2073229990
CO-INVESTIGATOR(S):
FACULTY SPONSOR (Required if PI is a student): Jennifer Seale
TITLE OF PROJECT: The Effect of Voice Gender & Spoken Messages on AAC interactions

START DATE: 12/1/16 PI DEPARTMENT: CSD
MAILING ADDRESS: 1 Hudson St. Apt. 421, Orono, ME 04473
FUNDING AGENCY (if any):
STATUS OF PI:
FACULTY/STAFF/GRADUATE/UNDERGRADUATE Undergraduate

1. If PI is a student, is this research to be performed:

- ☒ for an honors thesis/senior thesis/capstone?
- ☐ for a doctoral dissertation?
- ☐ other (specify)
- ☐ for a master's thesis?
- ☐ for a course project?

2. Does this application modify a previously approved project? (No). If yes, please give assigned number (if known) of previously approved project:

3. Is an expedited review requested? YES.

Submitting the application indicates the principal investigator's agreement to abide by the responsibilities outlined in Section I.E. of the Policies and Procedures for the Protection of Human Subjects.

Faculty Sponsors are responsible for oversight of research conducted by their students. The Faculty Sponsor ensures that he/she has read the application and that the conduct of such research will be in accordance with the University of Maine's Policies and Procedures for the Protection of Human Subjects of Research. REMINDER: if the principal investigator is an undergraduate student, the Faculty Sponsor MUST submit the application to the IRB.

Email complete application to Gayle Jones (gayle.jones@umit.maine.edu)

****

FOR IRB USE ONLY Application # 2016-12-12 Date received 12/9/2016 Review (F/E):E
Expedited Category:2

ACTION TAKEN:

- ☐X Judged Exempt; category2 Modifications required?Y Accepted (date) 12/15/2016
- ☐ Approved as submitted. Date of next review: by Degree of Risk:
- ☐ Approved pending modifications. Date of next review: by Degree of Risk:
  Modifications accepted (date):
- ☐ Not approved (see attached statement)
- ☐ Judged not research with human subjects

FINAL APPROVAL TO BEGIN 12/15/2016

Author's Biography

Katrina Lapham was born in Belfast, Maine on August 20, 1995. She was raised in Belfast, Maine and graduated from Belfast Area High School in 2013. Majoring in Communication Sciences and Disorders, Katrina has minors in Interdisciplinary Disability Studies and Psychology. She is a member of Alpha Lambda Delta, Phi Beta Phi and National Student Speech Language Hearing Association. She is also a Mitchell Scholar, Vice President of the Autism Club, and member of the Club Field Hockey team.

Upon graduation, Katrina plans to attend Northwestern University to pursue her Doctorate of Audiology.