Leveraging Communication Partner Speech to Automate Augmented Input for Children on the Autism Spectrum who are Minimally Verbal: Prototype Development and Preliminary Efficacy Investigation

Christine Holyfield[a], Stephen MacNeil[b], Nicolette Caldwell[a], Tara O'Neill Zimmerman [a], and Elizabeth Lorah[a]

[a]University of Arkansas, [b]Temple University

Christine Holyfield, Department of Communication Disorders and Occupational Therapy, University of Arkansas; Stephen MacNeil, Department of Computer Science, Temple University; Nicolette Caldwell, Department of Rehabilitation, Human Resources, and Communication Disorders, University of Arkansas; Tara O'Neill Zimmerman, Department of Communication Disorders and Occupational Therapy, University of Arkansas; Elizabeth Lorah, Department of Curriculum and Instruction.

Correspondence concerning this article should be addressed to Christine Holyfield, Department of Communication Disorders and Occupational Therapy, University of Arkansas, Fayetteville, AR 72701. E-mail: ceholyfi@uark.edu

Abstract

Purpose: Augmentative and alternative communication (AAC) technology innovation is urgently needed to improve outcomes for children on the autism spectrum who are minimally verbal. One potential technology innovation is applying artificial intelligence (AI) to automate strategies such as augmented input to increase language learning opportunities while mitigating communication partner time and learning barriers. Innovation in AAC research and design methodology is also needed to empirically explore this and other applications of AI to AAC. The purpose of this report was to describe (a) the development of an AAC prototype using a design methodology new to AAC research, and (b) a preliminary investigation of the efficacy of this potential new AAC capability.

Method: The prototype was developed using a Wizard-of-Oz prototyping approach that allows for initial exploration of a new technology capability without the time and effort required for full-scale development. The preliminary investigation with three children on the autism spectrum who were minimally verbal used an adapted alternating treatment design to compare the effects of a Wizard-of-Oz prototype that provided automated augmented input (i.e., pairing color photos with speech) to a standard topic display (i.e., a grid display with line drawings) on visual attention, linguistic participation, and (for one participant) word learning during a circle activity.

Results: Preliminary investigation results were variable, but overall participants increased visual attention and linguistic participation when using the prototype.

Conclusion: Wizard-of-Oz prototyping could be a valuable approach to spur much needed innovation in AAC. Further research into efficacy, reliability, validity, and attitudes is required to more comprehensively evaluate the use of AI to automate augmented input in AAC.

Keywords: Augmentative and alternative communication; Autism; Minimally verbal

Leveraging Communication Partner Speech to Automate Augmented Input for Children on the Autism Spectrum who are Minimally Verbal: Prototype Development and Preliminary Efficacy Investigation

The autism spectrum includes children who are at varying stages of communication and language development and use. Children on the autism spectrum who are minimally verbal are at the early stages of language development that can include learning social engagement and interaction, the use of intentional communication, and the learning and use of words. Children who are minimally verbal demonstrate restricted social and joint attention, use no more than a small number of gestures and words (through any modality), and use those gestures and words with limited frequency (DiStephano & Kasari, 2016; Mundy, 2018).

Benefits and Challenges of AAC for Children who are Minimally Verbal

Augmentative and alternative communication (AAC) technologies (e.g., speech generating devices, mobile technologies with communication apps) are effective to enhance communication and language skills as well as participation for children who are minimally verbal on the autism spectrum (Ganz et al., 2021; Ganz et al., 2023). With little use of speech and gestures, children who are minimally verbal require access to AAC to fulfill their human right to effective communication.

Still, children on the autism spectrum who are minimally verbal can continue to face significant barriers to participation and language learning even with access to today's standard AAC technologies and services (DiStefanio & Kasari, 2016). Demands of existing AAC technologies that can be particularly challenging to children who are minimally verbal often reflect the separateness from and lack of responsivity to the user and the immediate communication environment from which early language learning typically emerges (Holyfield, Caron et al., 2019; Light, 1997; Light & McNaughton, 2015). Thoughtfully designed AAC technologies (Light et al., 2019; O'Neill & Wilkinson, 2019) and the use of evidence-based interaction strategies (Lorah, Holyfield, Griffen et al., 2022) are critical to AAC effectiveness for children on the autism spectrum.

Benefits of Augmented Input and Challenges in Implementation

Augmented input is an evidence-based strategy (see Allen, Schlosser, Brock, & Shane, 2017; Biggs, Carter, & Gilson, 2018; O'Neill, Light, & Pope, 2018) that involves communication partners' use of aided AAC while communicating with an individual who uses AAC. Specifically, the partner points to and/or activates key symbols on the person's AAC device within the context of an ongoing interaction. For children who are minimally verbal, the pairing of speech with aided AAC serves to map spoken language to symbols, while also increasing the salience and meaning of the incoming speech (Wood et al., 1998). Furthermore, mapping speech to aided symbols serves to increase both receptive and expressive language skills by adding an external visual support and modeling appropriate use of the system for the communicator (e.g., Beukelman & Garrett, 1988; Romski & Sevcik, 1988).

Despite the benefits of augmented input for children who are minimally verbal, there are challenges for implementing this strategy clinically that create a need for innovation. Primarily, it places a number of demands on communication partners. Partners must be knowledgeable about the operation of the aided AAC system (e.g., where symbols are located), and they must consider a number of factors related effective use of augmented input including, among others: (a) the frequency of augmented input ; (b) the symbols for which aided input is provided (e.g., key word or full phrase); (c) whether the input is provided synchronously with or asynchronously from speech; and (d) the use of other strategies in conjunction with augmented input (e.g., time delay, open-ended questions, direct prompting, expansions or recasts). Partners must consider these factors in conjunction with general partner interaction skills for maintaining an interaction with the child using AAC (e.g., use of engaging materials and contexts).

Given the demands of implementing augmented input, partners may not use this strategy even though it is beneficial. In their survey of caregivers', teachers', and assistants' use of partner interaction strategies, Tegler et al., (2021) found that aided augmented input was the *least* used interaction strategy, compared to other strategies such as responsive strategies (e.g., wait time), environmental arrangements, and open-ended questions. Maximal benefit is derived from augmented input only when it is provided by multiple partners in various contexts throughout the communicator's everyday life. However, this requires training, and speech-language pathologists (SLPs) have reported varying success when attempting to train teachers and other communication partners in the use of aided AAC input, citing challenges related to limited time and opportunities for training (e.g., De Bortoli et al., 2014; Leatherman & Wegner, 2022). Additional macro-level factors such as disparities in access to AAC technologies (e.g., Pope, Light, & Franklin, 2022) may also impact the implementation of augmented input.

Addressing Technology Gaps with the Current Report

Clearly, the current standard of care for using augmented input and training partners places significant demands on communication partners. Furthermore, the current standard technologies used today place significant demands on children who are minimally verbal. Innovation of AAC technology is urgently needed to better meet the needs of children on the autism spectrum who are minimally verbal and their communication partners. Advances in artificial intelligence (AI) have the potential to increase the capacity of AAC technologies (Sennott et al., 2019), including by moving from contextual isolation to contextual awareness

(Holyfield et al., 2023). Applied to augmented input, the process could be automated through AI-engines that encode communication partner speech into visual input that is provided on the device in realtime. Automating augmented input through AI-powered, context-aware AAC could provide the high-meaning visual input critical to children who are minimally verbal while also reducing the effort required by communication partners.

Promise for this approach is supported by work outside AAC in which contextual information has been effectively integrated into technology. For instance, recent research in the field of human-computer interaction has demonstrated how context can be used to scaffold learning (Head et al., 2021; Head et al., 2015; MacNeil et al., 2022) by identifying contextually relevant opportunities to provide suggestions or learning materials. Additionally, researchers have previously developed a creativity support tool which listens to verbal design discussions and automatically displays contextually relevant design inspiration on a digital wall (Shi et al., 2017). These advances suggest that AI-enabled AAC technology could identify contextually relevant information from the communication partner's spoken input and automatically provide visual augmented input to the communicator. AI-supported integration of the communication partner linguistic context has the potential to automate the process of augmented input, thus reducing demands on partners and increasing the likelihood that the AAC communicator will receive aided input to improve communication skills. This report represents what is, to the authors' knowledge, the first step toward development of AAC technology that can provide automated augmented input through the power of AI and contextual awareness.

Because this report represents only an initial step, a prototype rather than a fully functional technology was developed. The approach to prototyping used, borrowed from the field of user experience design, offers a novel approach to AAC technology design and development.

To the authors' knowledge, this report represents the first use of this design approach within the field of AAC, addressing another gap relative to AAC technology design and development.

Addressing Research Gaps with the Current Report

In addition to the innovative nature of this work as described in the previous section, this study also seeks to fill gaps in the current literature in three ways. Primarily, this report includes a description of a preliminary investigation of the prototype developed that, to the authors' knowledge, represents the first evaluation of AAC technology that automatically provides augmented input through awareness of a critical aspect of the immediate communication context, communication partner speech.

Second, the existing research on the use of augmented input to improve communication skills for children who are minimally verbal has been limited. For example, in their review of augmented input, O'Neill et al. found that a small portion of participants were minimally verbal (i.e., only 8 of 88 total participants, or 17%, had a receptive language age between 1-24 months, with most falling on the upper end of the range). Also, previous research included important naturally occurring contexts (e.g., play, storybook reading), but augmented input within many frequently occurring contexts for young children (e.g., circle time activities) is needed.

Third, this study seeks to begin to investigate visual attention towards augmented input, a variable of interest that has received little attention to date (see O'Neill et al., 2018). While basic research has begun to investigate the relationships between visual cognitive processing and communication via AAC in a controlled environment (e.g., Wilkinson, Zimmerman, & Light, 2021), and other research has considered visual attention measured via eye-tracking technologies as a method to assess receptive language (e.g., Plesa-Skwerer et al. 2016), intervention research examining this variable is lacking (with some exceptions, e.g., Holyfield, 2019; Wilkinson & Bennett, 2021). This is likely due to the time intensive nature of hand coding visual attention data, and/or the cumbersome nature of using advanced eye tracking research technologies within the sphere of intervention research. However, visual attention is a critical variable to consider in augmented input research given that visual input is only useful insofar as it is attended to. Considering visual attention is particularly important for minimally verbal children on the autism spectrum. Specifically, children on the autism spectrum who are minimally verbal often have attentional limitations (DiStefano & Kasar, 2016; Mundy, 2018; Venker et al., 2018). Limited specific and sustained visual attention to referents as corresponding words are being spoken by communication partners creates an auditory-visual misalignment for children on the autism spectrum (Venker et al., 2018). This misalignment is disruptive to word learning because "…looking at the right thing at the right time allows children to build up correct associations between words and their references, and prune away incorrect or inconsistent ones, which supports learning in the moment and gradually produces a functional vocabulary" (Venker et al., 2018; p. 1622). Visual attention can also be measured in a short time frame while providing meaningful information regarding long-term language learning. Given the importance of visual attention to word learning, research on this important variable is critical.

Wizard-of-Oz Prototype Development

Established methodology in the field of user experience design was applied to the AAC prototype developed in this brief report, as described below. This work resulted in the development of a prototype of AI-powered, context-aware AAC with automated augmented input. Described in more detail below is the approach to prototyping used, the prototype that was developed, the inclusion of AAC users and other stakeholders throughout the development process, and an initial assessment of the technical feasibility of the innovation prototyped.

Wizard-of-Oz Prototyping Approach

There is an urgent need to develop technologies that genuinely consider the perspectives and experiences of end-users (Buchanan, 2001). Adopting a human-centered design approach entails deep empathy with users, ensuring that technology designers are grounded in users' lived experiences rather than relying on designer-made assumptions. This becomes particularly vital when creating products for neurodiverse users, given the vast range of abilities within this group (Gibson, 2020).

Designing for young children with cognitive and communication limitations also presents unique challenges (Light & McNaughton, 2015; Hendriks et al., 2015) as explicit linguistic feedback on designs may not be possible. For users with communication limitations, design researchers have developed on tangible design methods (Wilson et al., 2015). For users with cognitive limitations, design researchers tend to instead include friends, families, and caregivers in the design process (Bereton et al., 2015). Though our end-user group has both cognitive and communication limitations, we understood it to be critical to study their perspectives directly by creating a prototype with which they could interact (see "Current Prototype Technology"). However, given the important role that communication partners play in interactions with children who are minimally verbal, we also included their perspective in the design process (see "Stakeholder Engagement").

This approach introduces some risks because the design introduces more speculative elements, as it's not solely driven by end-user input. We were able to partially mitigate this risk by using a "wizard-of-oz" prototyping method (Dahlback et al., 1993; Maulsby et al., 1993).

This method involves constructing partially functional technology prototypes operated by a "wizard" who controls functionalities not yet implemented. In our case, we applied this method to AAC by employing behind-the-scenes human intervention to simulate machine learning and audio processing, enabling the system to listen to a communication partner and provide augmented input automatically. This approach helped us avoid the costly development of machine learning models and effectively prototype an intelligent input augmentation system to learn whether this approach would be worth further investigation.

Current Prototype Technology

The current prototype, developed by the second author, included a wizard interface where the wizard (i.e., the human completing the action that allowed a new technological capability to be mimicked) could select previously uploaded color photo images with corresponding text on one device (e.g., a cell phone or tablet) that would then appear in near real-time (i.e., under 1 s from selection) on a connected AAC user interface on a different device. That is, the automatic augmented input appearance of photos occurred by human action “wizarding” (i.e., creating an innovative technology experience through hidden human action rather than full-scale technology development) the automatic effect using the wizard’s user interface. After the background action of photo selection by the wizard, the selected photos were pushed in near real-time to appear on a separate, digitally connected AAC technology. During the preliminary investigation described in the next section, the wizard would click on a pre-loaded image and it would appear on a touchscreen as a communication partner spoke the corresponding referent. For instance, when the interventionist said, “Uh oh! Grass! Long, wavy grass!” a photo of tall grass was selected in the background by the wizard and appeared instantly on the AAC device seen by the participants.

The prototype was developed with careful consideration to the theory- and evidence-based in AAC technology design. For example, the prototype used motion to attract visual attention to a specific visual referent of a word at the same time the word was spoken by the interventionist (Jagaroo & Wilkinson, 2008). This design addressed attentional limitations common in children on the autism spectrum who are minimally verbal (Mundy, 2018), in particular, visual-auditory misalignment (Venker et al., 2018). The design also addressed both the input-output asymmetry experienced by children who use AAC (Light, 1997; Smith & Grove, 2003) and the limited provision of augmented input from communication partners (De Bortoli et al., 2014; Tegler et al., 2021). Color photos were used given the engaging and meaningful nature of this representation option for individuals who are minimally verbal (Holyfield, Brooks et al., 2019; Light et al., 2019; Muttiah et al., 2022; O'Neill & Wilkinson, 2019). If selected, the photos had corresponding voice output automatically generated through text-to-speech based on pre-determined text labels for each photo, another important component of AAC technology for children on the autism spectrum (Lorah, Holyfield, Miller et al., 2022).

Stakeholder Engagement

It is important to note that each step of the prototype development process occurred collaboratively with people who use AAC and other stakeholders. Before any development occurred, people who use AAC, parents of people who use AAC, and clinicians who work regularly with people who use AAC were included in interview discussions and provided formative feedback about the evolving design. Stakeholders were compensated with a gift card for their involvement in these formative needfinding sessions. The approach was developed in response to this input from stakeholders and AAC users about the need for AAC technology that is responsive to the current communication context. Additionally, people who use AAC were paid consultants on this project and provided guidance and feedback throughout the development process from initial conceptualization to final fine-tuning. The paid consultants were invited to all research meetings focused on the prototype, and they attended most of them. This collaboration was critical to the development process. AAC technology innovations should be driven by the priorities and needs of AAC users, and AAC users should have self-determination over the technology they use every day to communicate with the world.

Assessment of Technical Feasibility

Given the importance of research-to-practice development in the field of AAC, our team has investigated the technical feasibility of implementing the prototype described earlier in this section. This prototype is currently fully functioning regarding presenting the visual input and converting text to speech. However, natural language processing was wizarded for this prototype. Previous work has demonstrated the ability to perform automatic speech recognition (50-100ms), key phrase extraction (10-100ms), and image retrieval (300ms) in less than one second (Shi et al., 2017). This latency is faster than the performance of the wizard in this study. In addition, the recent introduction of large language models (LLMs; AI algorithms using very large data sets) has already demonstrated immense improvements in automatic speech recognition (ASR; i.e., OpenAI Whisper) and entity recognition (i.e., GPT-4). In previous work, heuristic approaches were used to extract key phrases to identify design inspiration to show on the “idea wall” with 93.5% recognition rate (Shi et al., 2017). However, LLMs will be able to extract even more relevant keywords from more complex sentences and phrases. Therefore, it is highly feasible to implement the prototype deployed in this study with low latency and high performance. The primary remaining challenge would be to build up a corpus of images to support the image retrieval task. This would be necessary because relying on images from the web could have safety and relevancy risks. One possibility is to rely on existing grid picture symbol line drawing sets, though ideally a set of photorealistic images would be crowdsourced and curated. This process could be bootstrapped using existing image repositories such as ImageNet (Russakovsky et al., 2015).

Preliminary Investigation of Prototype Efficacy

After the prototype was developed, it was utilized successfully in a small, preliminary investigation. The investigation was focused on the efficacy of the technological approach rather than the feasibility of the technology because evaluating feasibility of use of the technology would require-full scale development. Instead, the preliminary investigation served to provide a starting point of exploring whether future, larger scale research and eventual full-scale development is warranted. The preliminary investigation addressed the following research questions:

1. What are the comparative effects of AAC technology with automated augmented input compared to existing, standard AAC technology on visual attention toward the technology from preschoolers on the autism spectrum who were minimally verbal during an action song activity? We hypothesized that automated color photo augmented input would result in comparatively higher visual attention when compared to a standard AAC grid topic display with color line drawings due to the power of motion as a visual attractor (Jagaroo & Wilkinson, 2008) and the developmental appropriateness of color photo representations (Holyfield, Brooks et al., 2019; Light et al., 2019; Muttiah et al., 2022; O'Neill & Wilkinson, 2019).
2. What are the comparative effects of AAC technology with automated augmented input compared to existing standard AAC technology on linguistic participation from

preschoolers on the autism spectrum to participate in an action song activity? We hypothesized that the automated augmented input AAC technology would result in comparatively higher linguistic participation due to the hypothesized increased attention toward the technology and the immediate relevance of the word presented for input and communication, though we anticipated this difference to be less likely to be demonstrated given a lack of instruction for selecting the technology in either condition (Light et al., 2019; Lorah, Holyfield, Griffen et al., 2022).

3. What are the comparative effects of AAC technology with automated augmented input compared to existing standard AAC technology on the receptive language of preschoolers on the autism spectrum? We hypothesized that, given the short time frame of the study, increases in receptive language would be unlikely, but possible in the automated augmented input condition because the technology paired a spoken word to a high-salience referent – a process key to language learning (Light, 1997; O'Neill et al., 2018; Romski & Sevcik, 1996; Rose et al., 2016).

Method

Prior to recruitment, this study received Institutional Review Board approval from the University of Arkansas (approved protocol #2211433922). Prior to the start of the study, a guardian for all participants was informed about the study and provided consent for participation.

Design

The study used an adapted alternating treatment design (AATD) to compare two technology conditions while minimizing the possibility of learning carryover effects across the two conditions (Sindelar et al., 1985). The two technology conditions were: (1) a wizard-of-oz AAC technology prototype (see Materials section for further discussion) featuring automated color photo augmented input of keywords from a communication partner, and (2) a standard topic display with color line drawings on a popular AAC application. Technology condition type served as the independent variable in this study.

Participants

Participants were eligible to participate in the study if they met the following inclusion criteria based on parent/guardian interview and further assessed through structured observation as relevant: (a) were 4-5 years old; (b) were formally diagnosed on the autism spectrum; (c) were minimally verbal (DiStephano & Kasari, 2016), having restricted joint attention, using no more than a small number of gestures, and using no more than a small number of words (including AAC, speech, or manual sign); (d) could make a selection on a touch screen to the accuracy of at least a quarter of the screen; and (e) had hearing and vision within normal limits. Laura, Damien, and Jared (all pseudonyms) participated in the preliminary investigation. The participants were four (Damien and Jared) and five (Laura) years old at the start of the study. All participants lived in a home with family or a temporary guardian at the time of the study. Two participants (Laura and Damien) lived in a bilingual home where English and another language were both regularly spoken, but all parents/guardians reported English as each participants' primary language. The participants in this study were attendees at a small campus preschool month-long summer program designed to promote AAC and language learning for children on the autism spectrum. For two of the participants (Damien and Jared), the preschool was their first formal educational or clinical experience. One participant (Laura) received regular services, including behavioral services and speech-language services, and had received these services for about a year prior to attending the preschool summer program. During their time at the preschool, all participants gained opportunities to use AAC though, at the time of this study, Damien and Jared were not consistently using any words to communicate. Laura had a year of previous experience with AAC and was able to consistently select about 20 words to request, and also approximated fewer than five words vocally to request, though her vocalizations overall were limited. Damien and Jared did not approximate any words vocally and did not frequently vocalize beyond laughing and crying. Though Laura had some consistent and emerging use of formal linguistic communication (i.e., the use of a small number of words through AAC selection and vocal approximation), all three participants communicated primarily through prelinguistic means. Experience at the preschool clinic with AAC for all participants included both the use of tablets housing grids with line drawings and color photo scenes. Laura's experience with AAC in the year prior to the clinic was with a tablet housing words for requesting favorite activities, foods, and people. Table 1 provides more information about the participants, gathered through parent/guardian interview.

Table 1

Descriptive information about participants

Participant[a]	Age	Diagnosis	Gender	Race	Primary communication	Hearing and vision	Fine motor	Receptive language	History of services
Laura	5	Autism	Female	White	Prelinguistic; linguistic high-tech AAC use for requesting; emerging use of vocal approximations (e.g., "pi" to request "spin")	WNL	Consistent use of isolated point for selection	Responds behaviorally to communication partner questions and directions	1+ year receiving clinical services, including AAC
Jared	4	Autism	Male	White	Prelinguistic behaviors with unclear communicative intentionality, including body movements (e.g., leading others by hand, pushing away), nonspecific vocalizations (e.g., crying), and facial expressions	WNL	Whole hand selection; emerging use of isolated point	Responds behaviorally to communication partner questions and directions	No previous formal educational or clinical services
Damien	4	Autism	Male	Pacific Islander	Prelinguistic behaviors with unclear communicative intentionality, including body movements (e.g., reaching, walking away), facial expressions, and nonspecific vocalizations (e.g., laughing)	WNL	Whole hand selection; emerging use of isolated point	Responds behaviorally to communication partner questions and directions	No previous formal educational or clinical services

Note. WNL = Within normal limits.

[a]Participant names all pseudonyms.

Interventionist and Context

The first author, a licensed speech-language pathologist, served as the interventionist in the study. The third and fifth authors served as the “wizards” who controlled the prototype to create the automated augmented input effect using their mobile phones. While at the preschool, the participants engaged in a number of group circle time and one-on-one activities. The “Going on a Bear Hunt” action song was a group circle time activity in which all three children engaged. All study sessions occurred in this action song circle time context within the on-campus preschool setting. The participants and interventionist stood together in a circle with the song playing in the background while the interventionist sang the song and performed the actions with exaggerated expression. The song was adapted to include lyrics for all target words (see Word Lists section), and the activity lasted approximately 5 min. Participants danced, imitated actions, and had the opportunity to participate linguistically using AAC. This activity served as the context for the current pilot study due to the frequency of such activities in preschool, the importance of highly engaging AAC intervention contexts for children who are minimally verbal (Caldwell et al., in press; Griffen et al., in press; Holyfield, 2019), and the opportunities within the activity for new language learning and use for all participants.

Materials

The study compared two AAC software technologies. Both software were housed on the same hardware – an iPad® Pro with a 12.9” screen (measured diagonally). Both technologies as well as the word lists and receptive language probe used are discussed below.

Prototype Technology. The first condition in this study utilized the wizard-of-oz prototype developed prior to the start of the study. Please see the "Wizard-of-Oz Prototype Development" section for detailed information about the prototype.

Standard Technology. The second condition was designed to evaluate the current standard of technology commonly be used with children with developmental disabilities, including those on the autism spectrum who are minimally verbal (O'Neill & Wilkinson, 2019; Thistle & Wilkinson, 2021). A large topic grid display with four color line drawing icons was used. The four concepts represented on the grid were the three target words for the standard technology condition and a fourth non-target word that also appeared in the song. The topic display was created using the TouchChat® HD AAC app due to its ease of programming and popularity. Color line drawings used as representations on the display were the first icons suggested when searching for each of the words for button programming within the app.

Word Lists. Two word lists were created for the study based on the "Going on a Bear Hunt" action song activity. Six nouns that could appear in the story as a type of obstacle were selected after being reported by a professional or family members to be unknown words to each participant. Once selected, target words were randomly assigned to create two lists of three words. The two lists were: (a) grass, snow, forest; and (b) mud, river, cave. The lists were then randomly assigned to one of the two technology conditions. A different word list for each technology condition reduced the risk of carryover effects across conditions, particularly in the case of receptive language. Word list assignments could not be counterbalanced across participants because the study occurred in a group setting. Word lists contained three target words per condition to allow for the possibility of new word learning while restricting the learning demands on the participant given the short timeframe of the study.

Procedure

At the start of each session, the participants were gathered together in a circle with the interventionist and the other adults. The "Going on a Bear Hunt" song was played through a speaker on the third author's phone. This allowed the third author to pause the song at times when a target word appeared in the song and it also allowed control over the order in which each target word appeared in the song. The word list order was counterbalanced each session. That is, the three words from Word List A appeared before the three words from Word List B in one session, then the order flipped in the next session with the words from Word List B appearing first before the words from Word List A. Within each word list, words were chosen pseudorandomly by the third author who controlled the speaker by starting the first session with the target words within each list appearing in a randomly selected sequence, then rotating that sequence by one word each session.

As the song played, the interventionist and other adults sang along and modeled the actions that accompanied actions or repeated statements in the lyrics (e.g., stomping feet when the lyrics were "Stomp, stomp, stomp, stomp!"; making faux binoculars with hands around eyes when the lyrics were "I've got my binoculars."). Actions were not included for the target word nouns. The use of salient social behaviors by the investigator meant that her singing, speaking, gestures, and facial expressions were exaggerated to promote engagement from the beginning communicator participants (Caldwell et al., in press; Holyfield, 2019). The interventionist also held the AAC device throughout the song. While each child had their own AAC device present and available, the dancing, movement, and circle formation during the action song context meant the participants were not holding their devices nor could they be positioned in front of them without being held by someone. Despite being available, participants only used the AAC technology held by the interventionist throughout the study and never used their other devices.

The AAC device held by the participant featured one technology condition for one half of the song, then the other technology condition for the other half of the song in the counterbalanced order dictated by the word list as each word list was assigned to one technology condition for the entirety of the study. That is, in one session the interventionist used the prototype during the first half of the song (with Word List A) then the standard display during the second half of the song (with Word List B). In the prototype condition, when the target word appeared in the song, the third author used her phone to "wizard" it to automatically appear by selecting the corresponding photo on her phone that then pushed it to the AAC device seen by the participants. The screen was blank other than at times when a target word was being spoken by the interventionist. In the standard topic display condition, the grid display containing all target words for that condition remained present throughout the entirety of the half of the song in which that technology condition was being evaluated, reflecting a current AAC display.

When the music was paused for each target word, the interventionist spoke the word while positioning the AAC device in front of each participant. The device was positioned in front of each child for at least 5s. No prompting beyond the initial positioning of the device to attend to or use the device was provided as the purpose of the study was to evaluate the comparative effects of the technologies alone rather than comparing the technologies + teaching as a package. After each child had the opportunity to attend to and use the AAC device for that target word, the song was played and the singing and actions continued until the next target word appeared and the song was again paused. These procedures were repeated throughout the remainder of the song. Procedures were the same across study conditions and study sessions. Figure 1 outlines the procedural flow of the study across conditions.

Procedural Fidelity. The above procedures resulted in eight steps that were repeated across the six target words every session. Procedures were identical across technology conditions. The steps, repeated for each target word, were: (a) the song played with singing and actions delivered enthusiastically by the interventionist, (b) the song was paused at the target word, (c-e) the device was positioned in front of each participant at midline, under chin level, angled to allow for looking downward for at least 5 s while the target word was spoken by the interventionist, (f-h) no prompting beyond the initial device positioning or other teaching by the interventionist was provided. These eight steps per each of the six target words became the 48 procedural step checklist for fidelity each session. The third author monitored procedural fidelity live every session, providing in-the-moment reminders as needed to the interventionist. Immediately following each session, the third author reported to the interventionist any of the steps that were violated and this information was recorded by the interventionist. Steps completed were divided by the total number of steps for each session (step numbers were lower for one session when Laura was absent) then multiplied by 100 to yield a percentage. Sessions were also video recorded to allow for video review if needed, but this option was never needed. Mean procedural fidelity across the sessions was 95% (range: 92% - 100%).

Measures and Analyses

The independent variable in this study was technology condition type (i.e., prototype or standard). This study had two main dependent variables that were measured during every study session. The first main dependent measure of the study was sustained visual attention to the AAC technology; a behavior critical to eventual language learning that could be measured in the short study timeframe (O'Neill et al., 2018; Romski & Sevcik, 1988; Rose et al., 2016; Venker et al., 2018). Visual attention was operationally defined as 3 or more s of uninterrupted gaze directed toward the device screen or, if a selection was made on the device in fewer than 3 s, uninterrupted gaze directed toward the device screen until a selection was made on the device. The 3 s cutoff point provided a straightforward approach to measuring gaze that was sustained but not so long as to dismiss sustained gaze that was occurring, given that any sustainment of gaze is a meaningful opportunity for learning for children with limited engagement (Holyfield, 2019). The second main dependent measure of the study was accurate linguistic participation in the action song. Accurate linguistic participation was defined as the participant independently communicating a word that corresponded with the current point of the action song. Words communicated through AAC use, speech (i.e., vocal approximations), or formal signs/gestures (e.g., producing a sign from American Sign Language) would fulfill this definition, though none of the participants approximated a word vocally in this study nor did they produce a formal sign or gesture. Data were collected live by the interventionist who recorded it on a printed form then transferred it to a digital form after each session. To collect sustained visual attention data, the interventionist watched the eyes of each participant as the technology was positioned in front of them for each target word and marked sustained gaze as present for that target word if it met the operational definition above. To collect linguistic participation data, the interventionist recorded accurate linguistic participation as present if a word (e.g., a target word on the AAC device) was used at a time that corresponded to that point in the action song. Prior to this publication of the data, a student lab member with no previous involvement in the study or study discussions reviewed the printed and digital data forms and confirmed the data were a complete match.

Data Analyses. Primary data were graphed and visually analyzed per single subject methodological guidelines (Kratochwill et al., 2013). Visual analysis of the alternating treatment design included analyzing for differences in level, trend, and variability across the two technology conditions. The effect size of the technology type was also estimated using an approach with a strong relationship to visual analysis, Nonoverlap of All Pairs (NAP) (Parker & Vannest, 2009). Data were copied into an online single subject calculator to determine NAP values that were then interpreted as representing a strong effect (0.93 or higher), a moderate effect (0.66-0.92), or a weak effect (below 0.66) (Parker & Vannest, 2009).

Interrater Reliability. The third author, present at every session, also tracked participant data live during study sessions. Following each session, the interventionist and third author discussed their observed data to confirm interrater reliability. Interrater reliability was 100% for all dependent variables after each brief discussion. Prior to discussion, independent interrater reliability was on average 93% for visual attention across all sessions (range: 88%-100%), 100% for linguistic participation across all sessions. All sessions were also video recorded to allow for offline review in the case of disagreements, but such review was never needed.

Receptive Word Learning Probe

Secondary to method described above, a limited pre-post probe was also completed to evaluate any receptive learning of the target words across conditions to address the third, exploratory research question of the investigation. The probe measured an exploratory dependent variable secondary to the two main dependent variables described above: receptive vocabulary knowledge of key words from the action song, measured through accurate identification of a corresponding photo representation out of a field of four across two trials (to reduce the likelihood of correct identification by chance). This variable was measured immediately before the first study session and immediately after the last study session. This measure was probed pre-post as opposed to being measured after every session because it was an exploratory rather than primary measure due to the limited time frame of the study and the time required for children who are minimally verbal to acquire new words (Romski & Sevcik, 1988). Additionally, the pre-post measure reduced testing demands on the young participants who were minimally verbal.

The probe used two PowerPoint slides with four photo choices with each slide dedicated to one of the two word lists. The slide contained four color photos representing the three target words from the corresponding word list and a photo of a non-target word. All photos and non-target words from the probes did not appear elsewhere in the study. Photos differed from representation used elsewhere in the study to ensure the probe measured any growth in receptive vocabulary rather than a one-to-one correspondence. One non-target word was added for each condition to reduce levels of chance performance or reliance on process of elimination.

The interventionist sat with each participant one-on-one to administer the probe immediately prior to the first session and immediately after the last session, but only Laura completed the probes. Damien and Jared indicated an unwillingness to complete both probes by walking away from the interventionist and technology. Upon their leaving, the probe was concluded, so no information on word learning was gathered for two of the participants. Formal education and clinical experiences were new to both Damien and Jared at the time of the study, and they were not yet accustomed to completing educational testing of any kind. For the receptive language probe with Laura, the interventionist placed the iPad with the probe slides on the table directly in front of her sitting upright on a case. The interventionist used identical procedures for both the pre- and post- probes. For each target word, the interventionist: navigated to the corresponding slide, said “Point to [target word].”, and provided no corrective or confirmative feedback about performance. The interventionist did provide praise throughout the probe for participation (e.g. “Good working!”). These procedures were repeated twice each probe so that Laura was required to identify each target word twice each probe, and only those words that she accurately identified across both trials were counted as correct. Laura’s results on the probe were analyzed by comparing pre-post scores per condition. Interrater reliability for the probes, completed independently by the third author, was 100% for both sessions.

Results

This preliminary investigation explored the comparative effects of two types of AAC – (a) a prototype of AI-powered, contextually aware AAC with automated augmented input that provides color photo visual scaffolds of communication partner speech, and (b) existing, contextually isolated AAC with a standard topic display with isolated line drawing picture symbols. The comparative effects of the technologies on visual attention and linguistic participation from the three young child participants on the autism spectrum are discussed below. For one participant, the comparative effect of the technologies on receptive language pre- and post-study are also discussed. Figure 2 shows the results for the two primary measures for all participants.

Comparison of Visual Attention

Visual analysis of the data suggests that for Laura, visual attention was similar across the two technology types, with attention level higher in the automated input technology condition for just one session (NAP = 0.62, suggesting a weak effect). For Jared and Damien, however, sustained visual attention to the prototype with automated augmented input was higher than attention to the traditional topic display in every session (NAP = 0.96 for Jared and 1.00 for Damien, both suggestive of a strong effect). For Jared, variability in visual attention was observed in both conditions with no clear trend observed across sessions. Damien similarly demonstrated no clear trends in visual attention in either condition. However, his data were less variable; Damien demonstrated no instances of visual attention to the traditional topic display, only sustaining attention to the automated input prototype throughout the entirety of the study.

Comparison of Linguistic Participation

Visual analysis revealed that, though Laura demonstrated similar visual attention across the technology types, she was able to accurately use language to participate in every opportunity across every session when using prototype with automated input. That is, at every point in the song that a target word from the prototype condition was being featured in the lyrics, Laura not only attended to the device visually but selected the photo to communicate the word. Conversely, she demonstrated low levels of accurate linguistic participation using the traditional topic display (NAP = 1.00, indicating a strong effect). Jared and Damien demonstrated only emerging use of linguistic participation when provided automated augmented input, but this contrasted their having not once used the traditional topic display to participate linguistically across all study sessions (NAP = 0.90 for Jared and 0.70 for Damien, indicative of a moderate effect).

Comparison of Word Learning

For two participants, Jared and Damien, word learning could not be assessed because of their indication of a lack of willingness to complete the receptive probe by walking away, which was communication honored by the researchers. For one participant, Laura, comparative word learning across the two conditions could be explored within the limited pre-post probe context. When her receptive vocabulary knowledge of the 6 target words was tested prior to the first study session, Laura correctly identified none of the target words correctly across two trials. When tested after the last study session, Laura correctly identified across two trials all three target words assigned to the automated input condition. She identified none of the target words from the traditional topic display correctly across two trials. This suggests that Laura may have learned the words from the automated augmented input condition, but not the traditional topic display condition.

Discussion

Innovation of AAC technology is needed to better support communication and language growth from young children on the autism spectrum who are minimally verbal. Many advances in AI could allow for innovative change in AAC. Possibilities are limitless. AAC development resources, in contrast, are quite limited. Thus, innovative research and design methodologies are also needed to ensure that research and development resources maximize efficiency and effectiveness and are driven by the needs and priorities of end users. This brief report described a process for prototyping technology innovation that was new to AAC – wizard-of-oz prototyping – then described a preliminary investigation of the efficacy of a novel technological approach – AI-powered, contextually aware AAC with automated augmented input – using the prototype developed. What follows is a discussion of the implications and limitations of the work reported.

Implications for AAC Technology Innovation

The prototype development described in this preliminary report was, to the authors' knowledge, the first application of the wizard-of-oz design approach to AAC research and development. The prototyping approach allowed our research team to quickly begin to evaluate the effects of a novel AAC technology approach without requiring us to first devote the intensive time and resources required to fully develop the technology (Dahlback et al., 1993; Maulsby et al., 1993). The prototype allowed participants to experience the innovative technology as they would if it were fully developed. Thus, this report shows that prototyping can be an appropriate mechanism for efficient and effective efficacy evaluation with end users in order to determine the value and/or direction of future, more in-depth development. Resources for developing and evaluating AAC technology are significantly limited as compared to resources for mainstream technology development. Yet, recent AI advances that could be as transformative in AAC as they have been in the mainstream are innumerable (Holyfield et al., 2023; Sennott et al., 2019). As such, wizard-of-oz prototyping and other prototyping design approaches that allow for exploration of various applications of AI in AAC could provide the opportunity to explore a wider range of AI applications using fewer resources. Future research and development could apply wizard-of-oz prototyping as well as other user experience design approaches to accelerate AAC innovation.

The prototyping reported represented a collaborative effort between AAC researchers, user experience design researchers, AAC users, and other stakeholders (e.g., parents, clinicians). This collaboration was critical to the design process. The second authors' user experience design expertise allowed for successful implementation of the wizard-of-oz design approach. The AAC users and stakeholders who collaborated on this work ensured that the work was driven by the needs, priorities, and expert insights of people who use AAC every day. The prototyping process from the current report shows that collaboration among people with unique expertise can spur technology innovation in AAC.

Furthermore, the prototype reported represents, to the authors' knowledge, the first attempt toward the development of AAC that provides automated augmented input through awareness of a critical component of the immediate communication context, communication partner speech, using the power of AI. Today, AAC technologies are contextually isolated until manual selection by a communication partner connects them to the immediate communicative context (e.g. through the interaction strategy of augmented AAC input) (Holyfield, Caron et al., 2019; O'Neill et al., 2018), a task requiring effort and knowledge (De Bortoli et al., 2014; Kent-Walsh et al., 2015; Leatherman & Wegner, 2022; Tegler et al., 2021). The current report suggests along with prior literature (Holyfield et al., 2023; Sennott et al., 2019) suggests future research is warranted to explore this and other uses for AI in AAC technology for individuals who are minimally verbal. However, more research evaluating the existing prototype is warranted before implications for full-scale development can be drawn. In addition to more comprehensively evaluating efficacy, further evaluation of the existing prototype could allow for iteration on its design to maximize effectiveness.

Implications for AAC Research

The preliminary investigation in this report was, to the authors' knowledge, also the first study to evaluate the efficacy of integrating automated augmented input into AAC technology by leveraging real-time contextual information, specifically communication partner speech. The effects of automated input on visual attention, linguistic participation, and comprehension from young children on the autism spectrum within a common preschool context – action songs during circle time – were preliminarily explored. All participants showed heightened visual attention and linguistic participation toward the automated augmented input prototype as compared to the standard topic display, and one participant showed possible receptive learning of target words from the automated augmented input, but not the standard display, condition. Given the importance of visual attention and participation on communication and language learning (Holyfield, Caron et al., 2019; O'Neill et al., 2018; Rose et al., 2016), these preliminary findings provide an initial direction for future research into AI-automated augmented input for supporting language in young children on the autism spectrum who are minimally verbal.

Prior research has comprehensively demonstrated the positive impact of augmented AAC input on receptive and expressive language outcomes for individuals with developmental disabilities, including children on the autism spectrum who are minimally verbal (Allen et al., 2017; Beukelman & Garrett, 1988; Drager et al., 2006; Harris & Reichle, 2004; O'Neill et al., 2018; Romski & Sevcik, 1988). The purpose of the current report was not to evaluate augmented input as an intervention strategy, though the findings from the preliminary investigation do provide limited extensions of previous evidence by measuring visual attention from young children who are minimally verbal in response to augmented input and by providing augmented input within a circle time action song activity. Nor was the purpose to generate implications for the use of automated augmented input in clinical practice given the limited scope and exploratory nature of the investigation. Rather, the purpose the investigation completed was to begin to evaluate if and what future research and development is warranted to comprehensively explore automated augmented input as an AAC technology feature for young children on the autism spectrum who are minimally verbal.

The current findings suggest the possibility that the previously established positive impact of augmented AAC input could maintain when it is automated by technology rather than completed by a communication partner. If further research of automated augmented input shows comparable benefits to partner-produced augmented input, such a technology innovation stands to address the largest barrier to augmented input: communication partner implementation (De Bortoli et al., 2014; Kent-Walsh et al., 2015; Leatherman & Wegner, 2022; Tegler et al., 2021). Automating augmented input could allow for the same rich language learning opportunities while both: (a) alleviating existing knowledge, skill, and time barriers to manual augmented input implementation by communicators using technological automation of the input (b) scaffolding the visual-auditory alignment critical to word learning (Venker et al., 2018), by pairing a specific and salient visual referent in real-time to its spoken word representation. Automation could also correct input-output asymmetry – a major barrier to communication development for children who use AAC – by increasing augmented input frequency (Light, 1997; Smith & Grove, 2003). Could automation elevate augmented input from the least commonly used support to the most common? If so, what impact would such a change have on outcomes? Future research is certainly warranted.

While effective instruction is key to intervention for children on the autism spectrum (Lorah, Holyfield, Griffen et al., 2022), this study supports previous literature suggesting that theory- and evidence-based changes to technology design alone can be impactful for minimally verbal communicators (Light et al., 2019; O'Neill & Wilkinson, 2019); participants demonstrated differing engagement and participation based only on the difference in technology available to them. The current evaluation was theory- and evidence-informed by using accessible representation (Holyfield, Brooks et al., 2019; Light et al., 2019; Muttiah et al., 2022; O'Neill & Wilkinson, 2019) and motivating communication partner behavior and contexts (Caldwell et al., in press; Griffen et al., in press; Holyfield, 2019). Still, the results from this study were variable. While all three participants were minimally verbal, the one participant with comparatively more linguistic skills and more AAC experience seemed to benefit the most from access to the automated augmented input prototype. This is true despite her previous AAC experience more closely reflecting the standard topic display used in the alternate condition. Mapping AAC input onto communication partner speech can happen in different ways for different AAC users and toward different ends (O'Neill et al., 2018). Therefore, future research is warranted to evaluate alternate approaches to automated augmented input for different users and purposes. Can automated augmented input be made more accessible to minimally verbal communicators who are in the earliest stages of language learning? For example, could powerful visual attracters like movement and luminance (Jagaroo & Wilkinson, 2008) be used to highlight one object/person/event within a color photo scene to support specific vocabulary learning within a broader visual context? Should extended reality (XR) be leveraged to embed automated input deep within the physical environment to reduce the demand of attentional shifting and coordinating that can serve as a language learning bottleneck (Mundy, 2018)? Conversely, can automated augmented input be extended to support more advanced communication skills such as combining words, locating words on a robust communication system, or by modeling literacy? Movement and luminance could also be used to highlight one symbol on a grid or an entire selection sequence to support operational competence for semantic-syntactic communicators. Future research should also explore these and other automated augmented input approaches for individuals with different diagnostic and language profiles across different contexts.

Limitations of Technological Innovation

Of course, no approach is without barriers. Much knowledge, time, and energy on the part of technology developers would be required to fully develop and implement context-aware, AI-powered AAC that can automatically provide augmented input based on communication partner speech. Significant development and decision making would be required, such as determining which spoken words were augmented for whom. Working side-by-side with people who use AAC and their communication partners, as this team has done, will be critical to implementation success. The size of the required investment calls for more research with prototypes to continue evaluating efficacy before the required resources are invested (Dahlback et al., 1993; Maulsby et al., 1993).

Relatedly, while we considered technical feasibility during the prototype development process, the wizard-of-oz prototyping approach does not allow for the evaluation of feasibility because it only mimics the technological process. Mimicking rather than fully developing technology bypassed a number of possible complications that may emerge through actual AI-enabled technology implementation (e.g., background noise in noisy environments; interpreting different types of speech). As such, a major limitation of the current report is that feasibility of the automated augmented input technological innovation was not evaluated. Future research is needed to evaluate technological feasibility as well as feasibility of use of the approach by naturally occurring communication partners within a real-world context. For instance, to determine how communication partner speech could be clearly captured inside a bustling preschool setting, and to ensure communication from partners who use synthesized speech is effectively integrated.

Furthermore, automated augmented input is made possible through technology that is tracking words spoken by communication partners. Though safeguards could be implemented, this privacy compromise is one that some people will surely refuse. Processes would be required to ensure the technology is only used in specific contexts with people who explicitly agree to it. Additionally, because of the risk involved with computers selecting photos, oversight would be needed to ensure the photos were developmentally, culturally, and age appropriate. Such oversights could take the form of the computer drawing from a large, curated, pre-approved bank of photo options (such as described in the "Evaluation of Technical Feasibility" section), or human-in-the-loop selection of personalized or computer-offered photo representations by communication partners (e.g., parents) for each word the first time it is detected for augmentation.

Limitations of the Preliminary Investigation

Beyond the potential technological limitations, the preliminary investigation in this report had important limitations in size and scope that must be considered when interpreting its results and planning future research. The study included only three participants who were all young children on the autism spectrum who were minimally verbal. The study also occurred over a short period of time in only one context – an action song circle activity – led by a trained speech-language pathologist and researcher with no measures of generalization to other contexts and communication partners. Future research should include more participants with a range of ages and diagnoses, should occur over multiple contexts, and should include naturally occurring communication partners to being to more comprehensively evaluate the use of context-aware, AI-powered AAC with automated augmented input. While one participant was a Pacific Islander and two participants in the current study lived in bilingual homes where English and another language were spoken regularly, this study was completed in English only. Future research should continue to include participants from underrepresented races, and should evaluate automated augmented input use within languages other than English.

The paper was further limited by the measures. While visual attention is an important measure in augmented AAC input research (O'Neill et al., 2018), it was measured in a very limited way by the investigator watching the gaze of the participant and documenting whether or not that gaze sustained for 3 s or longer. Higher technology approaches to measuring visual attention are available and should be utilized in future research (e.g., cameras that can be added to a device screen). Additionally, despite receptive language being a major goal of supplementing communication partner speech with visual scaffolds as was done in the current investigation (Beukelman & Garrett, 1988; Drager et al., 2006; Harris & Reichle, 2004; Romski & Sevcik, 1988), receptive measures were only completed with one participant, and measured only in a single pre-post probe. The measure also relied on live speech from the interventionist (e.g., "Point to ____") rather than recorded audio that could ensure consistency. The probe also was generated by the researchers for the specific purposes of this study and was not assessed for validity or reliability. Future research should also include more standardized measures of receptive vocabulary. The receptive measure was also not effective for use with two of the children in the study and future research should more effectively evaluate receptive one-word vocabulary from individuals who are minimally verbal.

Finally, the two technologies used in the current study differed in multiple ways in order to compare an ideal automated technology to the current standard. However, this creates a limitation in that the technologies differed on several features including the provision of automatic augmented input, the presence of a different number of symbols, and the representation use. As a specific example, the prototype utilized color photos based on the evidence of their power for individuals who are minimally verbal (Holyfield et al., 2019; Light et al., 2019; Muttiah et al., 2022; O'Neill & Wilkinson, 2019) while the comparison technology used the picture symbol line drawings that represent the standard of AAC technology often used today. The differences in technologies makes it difficult to determine which specific differences were most impactful, though the differences probably combined to have an effect. Future research is needed to fully understand which aspects of the automated AAC technology may or may not be supportive of engagement, communication, and language, and for whom.

Conclusion

Today, AI is being leveraged at an accelerated pace and toward constantly expanding ends. Yet, technology innovations in the AI space have largely focused on mainstream technologies and neurotypical users. AI innovations in AAC technology have been far less accelerated and far from constant. Deliberate research and development is needed to ensure that AI innovation is more equitable and the needs of all technology users are considered, including those who use AAC. This report described the application of a technology design approach novel to AAC development, wizard-of-oz prototyping. This application showed the potential value of the approach in AAC development and research for allowing initial evaluation of innovations without requiring full-scale development. In using wizard-of-oz prototyping, future work could efficiently explore a wider range of innovation options without treating limited resources for AAC development with frivolity. The wizard-of-oz approach was applied to prototype AI-powered, context-aware AAC technology that could automatically provide real-time augmented input based on spoken input from communication partners. With augmented input being a powerful learning support (Allen et al., 2017; O'Neill et al., 2018) that is infrequently implemented by communication partners (De Bortoli et al., 2014; Tegler et al., 2021), it is ripe for automation. The preliminary evaluation in this report suggests that some young children on the autism spectrum who are minimally verbal may demonstrate increased visual attention toward and linguistic participation with AAC technology that automatically pairs color photo augmented input with communication partners speech as compared to a standard topic display, warranting future research. Further, the positive exploratory results on word learning with one participant suggest the possibility that automatic color photo augmented input should also be considered in more comprehensive future research on receptive language. More research and development is critical to comprehensively determine the efficacy and feasibility of automated augmented input in AAC technology. As was done in this study, such research and development must include people who use AAC and other stakeholders, and must pay them for their invaluable expertise. In particular, more robust research in to AAC user and communication partner perspectives about this approach is needed. Automated augmented input will only be a useful technological support if it is adopted. If adopted, automation could transform augmented input from a powerful language support too effortful to regularly implement into a near effortless language support rendered more powerful than ever by the consistency of its implementation.

Acknowledgements

This project was funded by a Track H Convergence Accelerator grant (Award Number 2236352) from the National Science Foundation to which we are grateful. We are also grateful to this study's participants and their families for allowing us to learn from them. We are grateful too to the paid consultants who used AAC who provided important insights into prototype development and helped guide our work.

Data Availability Statement

All data generated or analyzed during this study are included in this published article.

References

Allen, A. A., Schlosser, R. W., Brock, K. L., & Shane, H. C. (2017). The effectiveness of aided augmented input techniques for persons with developmental disabilities: a systematic review. *Augmentative and Alternative Communication*, *33*(3), 149-159. https://doi.org/10.1080/07434618.2017.1338752

Beukelman, D., & Garrett, K. (1988). Augmentative and alternative communication for adults with acquired severe communication disorders. *Augmentative and Alternative Communication, 4*(2), 104–121. https://doi.org/10.1080/ 07434618812331274687

Biggs, E. E., Carter, E. W., & Gilson, C. B. (2018). Systematic review of interventions involving aided AAC modeling for children with complex communication needs. *American Journal on Intellectual and Developmental Disabilities*, *123*(5), 443-473. https://doi.org/10.1352/1944-7558-123.5.443

Brereton, M., Sitbon, L., Abdullah, M. H. L., Vanderberg, M., & Koplick, S. (2015). Design after design to bridge between people living with cognitive or sensory impairments, their friends and proxies. *CoDesign, 11*(1), 4-20. https://doi.org/10.1080/15710882.2015.1009471

Buchanan, R. (2001). Human dignity and human rights: Thoughts on the principles of human-centered design. *Design Issues, 17*(3), 35-39.

Caldwell, N., Holyfield, C., & Zimmerman, T.O. (in press). Effects of interaction approaches and added high-tech augmentative and alternative communication on prelinguistic and linguistic communication from school-age beginning communicators. *Folia Phoniatrica et Logopaedica*.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993, February). Wizard of Oz studies: why and how. In Proceedings of the 1st international conference on Intelligent user interfaces, 193-200.

De Bortoli, T., Arthur-Kelly, M., Mathisen, B., & Balandin, S. (2014). Speech-language pathologists' perceptions of implementing communication intervention with students with multiple and severe disabilities. *Augmentative and Alternative Communication*, *30*(1), 55-70. https://doi.org/10.3109/07434618.2014.881916

DiStefano, C., & Kasari, C. (2016). The window to language is still open: Distinguishing between preverbal and minimally verbal children with ASD. *Perspectives of the ASHA Special Interest Groups, 1*(1), 4-11. https://doi.org/10.1044/persp1.SIG1.4

Drager, K. D., Postal, V. J., Carrolus, L., Castellano, M., Gagliano, C., & Glynn, J. (2006). The effect of aided language modeling on symbol comprehension and production in 2 preschoolers with autism. *American Journal of Speech-Language Pathology*, *15*(2), 112-125. https://doi.org/10.1044/1058-0360(2006/012)

Gibson, R. C., Dunlop, M. D., Bouamrane, M. M., & Nayar, R. (2020, April). Designing clinical AAC tablet applications with adults who have mild intellectual disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13). https://doi.org/10.1145/3313831.3376159

Griffen, B., Holyfield, C., Lorah, E., & Caldwell, N. (in press). Increasing Linguistic and Prelinguistic Communication for Social Closeness During Naturalistic AAC Instruction with Young Children on the Autism Spectrum. *Augmentative and Alternative Communication*.

Harris, M. D., & Reichle, J. (2004). The impact of aided language stimulation on symbol comprehension and production in children with moderate cognitive disabilities. *American Journal of Speech-Language Pathology*, *13*(2), 155-167. https://doi.org/10.1044/1058-0360(2004/016)

Head, A., Lo, K., Kang, D., Fok, R., Skjonsberg, S., Weld, D., & Hearst, M. (2021). Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 413, 1–18. https://doi-org.libproxy.temple.edu/10.1145/3411764.3445648

Head, A., Appachu, C., Hearst, M. & Hartmann, B. (2015) Tutorons: Generating context-relevant, on-demand explanations and demonstrations of online code. IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 3-12. https://doi.org/10.1109/VLHCC.2015.7356972.

Hendriks, N., Slegers, K., & Duysburgh, P. (2015). Codesign with people living with cognitive or sensory impairments: a case for method stories and uniqueness. *CoDesign, 11*(1), 70-82. https://doi.org/10.1080/15710882.2015.1020316

Holyfield, C. (2019). Preliminary investigation of the effects of a prelinguistic AAC intervention on social gaze behaviors from school-age children with multiple disabilities. *Augmentative and Alternative Communication, 35*(4), 285-298. https://doi.org/10.1080/07434618.2019.1704866

Holyfield, C., Brooks, S., & Schluterman, A. (2019). Comparative effects of high-tech visual scene displays and low-tech isolated picture symbols on engagement from students with multiple disabilities. *Language, Speech, and Hearing Services in Schools, 50*(4), 693-702. https://doi.org/10.1044/2019_LSHSS-19-0007

Holyfield, C., Caron, J., & Light, J. (2019). Programing AAC just-in-time for beginning communicators: The process. *Augmentative and Alternative Communication, 35*(4), 309-318. https://doi.org/10.1080/07434618.2019.1686538

Holyfield, C., MacNeil, S., Vucetic, S., Dragut, E., Rackensperger, T., Lorah, E., Caldwell, N., Karnes, A. (2023, January). Is the current state of natural language processing the future for AAC? Presentation at the annual conference for the Assistive Technology Industry Association, Orlando, Florida.

Jagaroo, V., & Wilkinson, K. (2008). Further considerations of visual cognitive neuroscience in aided AAC: The potential role of motion perception systems in maximizing design display. *Augmentative and Alternative Communication, 24*(1), 29-42. https://doi.org/10.1080/07434610701390673

Kent-Walsh, J., Murza, K. A., Malani, M. D., & Binger, C. (2015). Effects of communication partner instruction on the communication of individuals using AAC: A meta-analysis. *Augmentative and alternative communication*, *31*(4), 271-284. https://doi.org/10.3109/07434618.2015.1052153

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38. https://doi.org/10.1177/0741932512452794

Leatherman, E. M., & Wegner, J. R. (2022). Augmentative and alternative communication in the classroom: Teacher practices and experiences. *Language, Speech, and Hearing Services in Schools*, *53*(3), 874-893. https://doi.org/10.1044/2022_LSHSS-21-00125

Light, J. (1997). "Let's go star fishing": Reflections on the contexts of language learning for children who use aided AAC. *Augmentative and Alternative Communication, 13*(3), 158-171. https://doi.org/10.1080/07434619712331277978

Light, J., & Mcnaughton, D. (2015). Designing AAC research and intervention to improve outcomes for individuals with complex communication needs. *Augmentative and Alternative Communication, 31*(2), 85-96. DOI

Light, J., Wilkinson, K. M., Thiessen, A., Beukelman, D. R., & Fager, S. K. (2019). Designing effective AAC displays for individuals with developmental or acquired disabilities: State of the science and future research directions. *Augmentative and Alternative Communication, 35*(1), 42-55. https://doi.org/10.1080/07434618.2018.1558283

Lorah, E. R., Holyfield, C., Griffen, B., & Caldwell, N. (2022). A systematic review of evidence-based instruction for individuals with autism using mobile augmentative and alternative communication technology. *Review Journal of Autism and Developmental Disorders*, 1-15. https://doi.org/10.1007/s40489-022-00334-6

Lorah, E. R., Holyfield, C., Miller, J., Griffen, B., & Lindbloom, C. (2022). A systematic review of research comparing mobile technology speech-generating devices to other AAC modes with individuals with autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 34*(2), 187-210. https://doi.org/10.1007/s10882-021-09803-y

MacNeil S, Patel P, Smolin BE. Expert Goggles: Detecting and Annotating Visualizations using a Machine Learning Classifier (2022, October). In Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, 1-3.

Maulsby D, Greenberg S, Mander R. (1993, May) Prototyping an intelligent agent through Wizard of Oz. In Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems, 277-284.

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience, 47*(6), 497-514. https://doi.org/10.1111/ejn.13720

Muttiah, N., Drager, K. D., Beale, B., Bongo, H., & Riley, L. (2022). The effects of an intervention using low-tech visual scene displays and aided modeling with young children with complex communication needs. *Topics in Early Childhood Special Education, 42*(1), 91-104. https://doi.org/10.1177/02711214198448

O'Neill, T., Light, J., & Pope, L. (2018). Effects of interventions that include aided augmentative and alternative communication input on the communication of individuals with complex communication needs: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1743-1765. https://doi.org/10.1044/2018_JSLHR-L-17-0132

O'Neill, T., & Wilkinson, K. M. (2019). Designing developmentally sensitive AAC technologies for young children with complex communication needs: Considerations of communication, working memory, attention, motor skills, and sensory-perception. *Seminars in Speech and Language 40*(4), 320-332. https://doi.org/10.1055/s-0039-1692966

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*(5), 591-604. https://doi.org/10.1177/1362361315600146

Pope, L., Light, J., & Franklin, A. (2022). Black children with developmental disabilities receive less AAC intervention than their white peers: Preliminary evidence of racial disparities from a secondary data analysis. *American Journal of Speech Language Pathology*, *31*(5), 2159. https://doi.org/10.1044/2022_AJSLP-22-00079

Romski, M. A., & Sevcik, R. (1988). Augmentative and alternative communication systems: Considerations for individuals with severe intellectual disabilities. *Augmentative and Alternative Communication*, *4*(2), 83-93. https://doi.org/10.1080/07434618812331274667

Romski, M. A., & Sevcik, R. (1996). *Breaking the speech barrier: Language development through augmented means.* Baltimore, MD: Brookes.

Rose, V., Trembath, D., & Bloomberg, K. (2016). Visual attention and key word sign in children with autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 28*, 33-55. https://doi.org/10.1007/s10882-015-9443-3

Russakovsky, O., Deng., J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge, IJCV.

Sennott, S. C., Akagi, L., Lee, M., & Rhodes, A. (2019). AAC and artificial intelligence (AI). *Topics in Language Disorders, 39*(4), 389. https://doi.org/10.1097/tld.0000000000000197

Shi Y, Wang Y, Qi Y, Chen J, Xu X, Ma KL. (2017, February). IdeaWall: Improving creative collaboration through combinatorial visual stimuli. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 594-603.

Sindelar, P. T., Rosenberg, M. S., & Wilson, R. J. (1985). An adapted alternating treatments design for instructional research. *Education and Treatment of Children, 8*(1), 67-76.

Smith, M. M. & Grove, N.C. (2003). Asymmetry of input and output for individuals who use AAC. In J. Light, D. Beukelman, & J. Reichle (Eds.) *Communicative competence for individuals who use AAC: From research to effective practice* (pp. 163-195). Baltimore: Brookes.

Tegler, H., Pless, M., Blom Johansson, M., & Sonnander, K. (2021). Caregivers', teachers', and assistants' use and learning of partner strategies in communication using high-tech speech-generating devices with children with severe cerebral palsy. *Assistive Technology*, *33*(1), 17-25. https://doi.org/10.1080/10400435.2019.1581303

Thistle, J. J., & Wilkinson, K. M. (2021). Speech-Language pathologists' decisions when designing an aided AAC display for a compilation case study of a beginning communicator. *Disability and Rehabilitation: Assistive Technology, 16*(8), 871-879. https://doi.org/10.1080/17483107.2020.1745911

Venker, C. E., Bean, A., & Kover, S. T. (2018). Auditory–visual misalignment: A theoretical perspective on vocabulary delays in children with ASD. *Autism Research, 11*(12), 1621-1628. https://doi.org/10.1002/aur.2038

Wilkinson, K. M., & Bennett, R. (2021). Impact of the design of visual communication supports on social interaction by individuals with Down syndrome. *Journal of Interactional Research in Communication Disorders*, *10*(1), 30-46. https://doi.org/10.1558/jircd.19316

Wilkinson, K. M., Zimmerman, T., & Light, J. (2021). Visual attention to cued targets in simulated aided augmentative and alternative communication displays for individuals with intellectual and developmental disabilities. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1726-1738. https://doi.org/10.1044/2021_JSLHR-20-00451

Wilson, S., Roper, A., Marshall, J., Galliers, J., Devane, N., Booth, T., & Woolf, C. (2015). Codesign for people with aphasia through tangible design languages. *CoDesign, 11*(1), 21-34. https://doi.org/10.1080/15710882.2014.997744

Wood, L., Lasker, J., Siegel-Causey, E., Beukelman, D., & Ball, L. (1998). Input framework for augmentative and alternative communication. *Augmentative and Alternative Communication*, *14*(4), 261-267. https://doi.org/10.1080/07434619812331278436

Figures

Figure 1. An illustration exampling the procedural process across technology conditions in the preliminary investigation. As illustrated, in the prototype condition the wizard remotely selected pre-loaded photos that appeared on the AAC device for the child in real-time as the corresponding word was spoken by the interventionist.

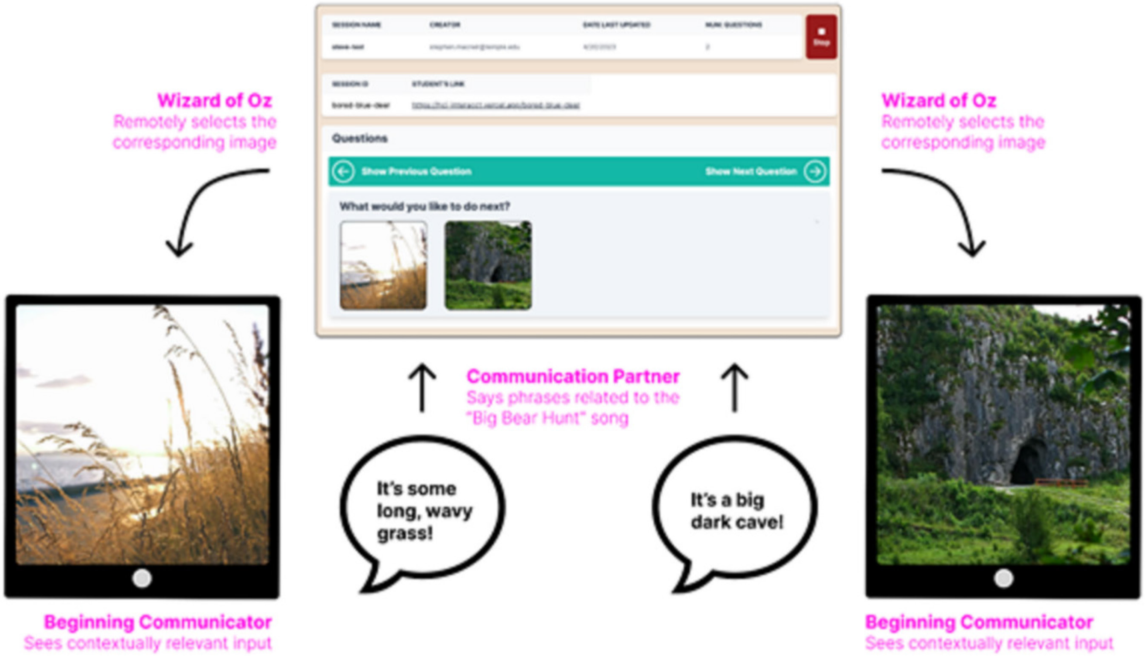
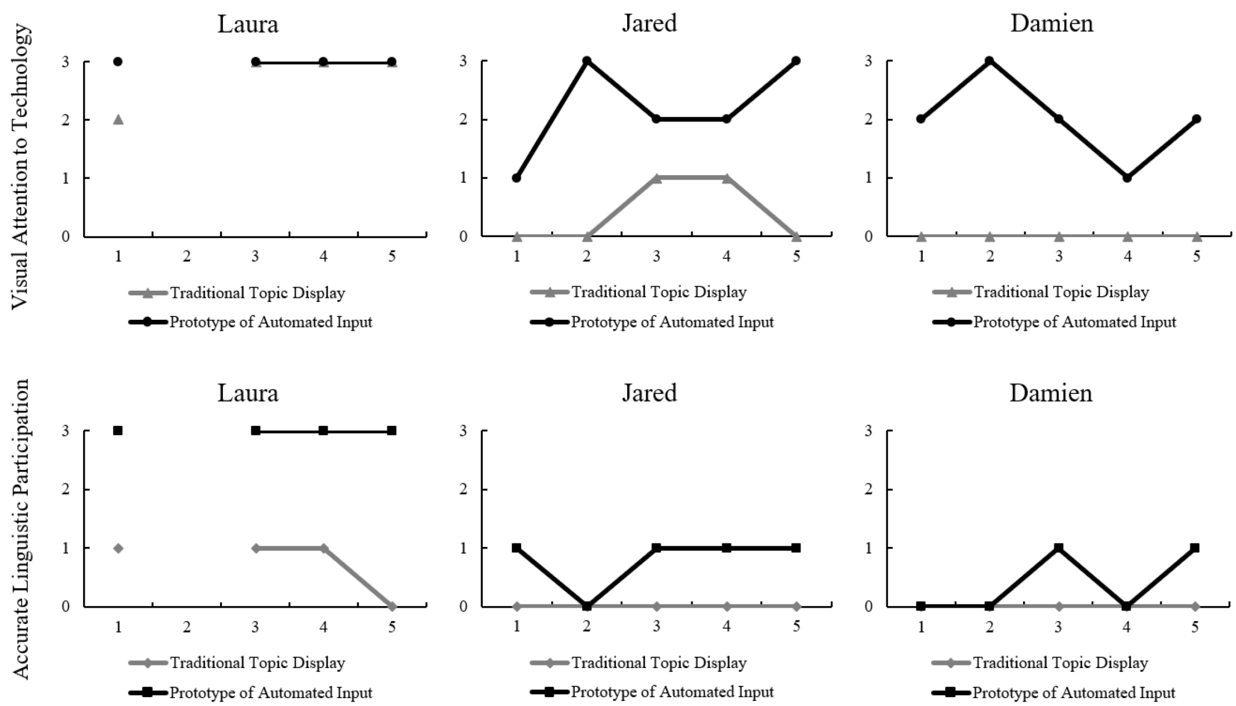


Figure 2. Sustained visual attention to technology across conditions (top row) and accurate linguistic participation in an action song across conditions (bottom row) for the three participants.


Visual Attention to Technology
Laura
Jared
Damien
Traditional Topic Display
Prototype of Automated Input
Accurate Linguistic Participation
Laura
Jared
Damien
Traditional Topic Display
Prototype of Automated Input