HoloAAC: A Mixed Reality AAC Application for People with Expressive Language Difficulties

Liuchuan Yu[1(✉)], Huining Feng[1], Rawan Alghofaili[1,2], Boyoung Byun[1], Tiffany O'Neal[1], Swati Rampalli[1], Yoosun Chung[1], Vivian Genaro Motti[1], and Lap-Fai Yu[1]

[1] George Mason University, Fairfax, VA 22030, USA
{lyu20,hfeng2,ralghofa,cbyun2,toneal,srampall,ychung3,vmotti,craigyu}@gmu.edu, rawan@utdallas.edu
[2] University of Texas at Dallas, Richardson, TX 75080, USA

Abstract. We present a novel AAC application, HoloAAC, based on mixed reality that helps people with expressive language difficulties communicate in grocery shopping scenarios via a mixed reality device. A user, who has difficulty in speaking, can easily convey their intention by pressing a few buttons. Our application uses computer vision techniques to automatically detect grocery items, helping the user quickly locate the items of interest. In addition, our application uses natural language processing techniques to categorize the sentences to help the user quickly find the desired sentence. We evaluate our mixed reality-based application on AAC users and compare its efficacy with traditional AAC applications. HoloAAC contributed to the early exploration of context-aware AR-based AAC applications and provided insights for future research.

Keywords: Augmentative and alternative communication · Mixed reality · Assistive technology · Object detection · Text-to-speech

1 Introduction

Augmentative and alternative communication (AAC) [5] is a communication mechanism for those with complex communication needs (CCN) [33], and existing AAC devices are forms of assistive technology (AT) comprising hardware and software that can support or replace natural speech entirely. On the other hand, augmented reality (AR), a user's visual perception supplemented with additional computer-generated sensory modalities, is rising in its ability to support AT through rehabilitation therapies that support people with disabilities.

While immersive learning applications in AR have greatly supported individuals with disabilities, current AAC devices do not carry the contextual intelligence to prompt appropriate conversation choices or phrases based on a user's environment. This is particularly concerning for emergency situations where real-time communication is important for supporting AAC users who have to not

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2024, LNCS 14708, pp. 304–324, 2024.
https://doi.org/10.1007/978-3-031-61047-9_20

only consider their accommodations but also navigate a crisis with heightened emotions. This prompts the need for an AI-driven AAC system aware of the environmental situation and demand. Because of the monumental shift in the nature of AAC, AAC has expanded its reach to include more people with a wider range of CCN [39].

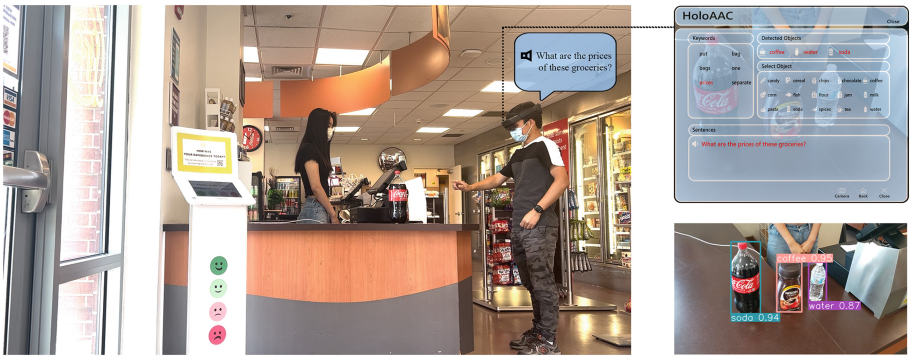

Fig. 1. When the user wears HoloLens 2 and stands by the side of the cashier, the user clicks the camera button to capture current objects on the desk. In this scenario, there are three objects on the desk: soda, coffee, and water. After the captured image is processed on the PC, the detected objects, the generated keywords, and the generated sentences will be shown in front of the user via an AAC interface visualized by HoloLens 2. As the user clicks the *prices* keyword, the sentence "what are the prices of these groceries?" is shown. The user clicks this sentence to trigger our application to speak it accordingly.

AR is becoming popular in various fields such as teaching [42], learning [29], entertainment [19], defense [44], and marketing [36]. As an immersive technology, AR opts to observe the user's surroundings, understand the context, and synthesize context-aware content with the aid of computer vision and artificial intelligence algorithms. Moreover, head-mounted AR headsets feature egocentric vision, referring to being capable of seeing what the user sees. These factors make head-mounted AR headsets promising vehicles for delivering AAC applications in the future. Compared to current AAC devices that require users to operate an AAC application on a phone or tablet, an AAC app running on a head-mounted AR headset could be less distracting and more intuitive to provide in-situ conversation help, thanks to the advantage of AR in being able to incorporate a user interface into the physical environment, which reduces gaze switch and enhances eye contact.

To explore this direction, we propose HoloAAC, a computer vision-guided mixed-reality AAC application that helps AAC users in grocery shopping scenarios as shown in Fig. 1. First, we devise a computational approach to generate shopping-related sentences. Second, we use a mixed reality device to capture an image of the current context, based on which an object detection algorithm

is applied. Third, we propose a natural language processing (NLP) based algorithm to help the user quickly find the desired sentence. Fourth, a text-to-speech engine will translate the entire sentence into speech upon the user's selection. To the best of our knowledge, HoloAAC is the very first application that explores using mixed reality and contextual awareness to provide AAC for users who have expressive language difficulties but possess good control over hand movements. The major contributions of this work include:

- Proposing a novel AAC interface that can be used on a mixed-reality headset;
- Devising an interactive approach based on object detection and text retrieval techniques to help AAC users quickly retrieve and speak desired sentences via text-to-speech;
- Evaluating our approach through experiments that mimic grocery scenarios and case studies conducted with people who have expressive language difficulties.

HoloAAC code is available at https://github.com/luffy-yu/HoloAAC.

2 Related Work

There are needs for just-in-time communication and context-aware technologies in the AAC community. In fact, this is an area of need that has been prevalent. We review some existing works.

2.1 Context-Aware AAC

Communication depends on context. People talk about things that are rooted in their environments [31]. A context-aware system decides what information and which service should be presented to the user [38].

TalkAbout [21] is a context-aware and adaptive AAC system that tailors its users' word list based on their current surroundings and the person they are conversing with. TryTalk [14] operates similarly, considering the user's location obtained through GPS or building QR code, as well as the day and time. Chan et al. [7] employed Bluetooth Low Energy beacons for precise indoor tracking and a micro-location context-aware AAC system to minimize the cognitive burden of user interaction. Moreover, Chan et al. [8] proposed a context-aware AAC system to enhance daily communication for nonverbal schoolchildren with moderate intellectual disabilities. On the other hand, Shen et al. [40] devised KWickChat for nonspeaking individuals with motor disabilities, which leverages a GPT-2 language model and context information to improve the quality of the generated responses. Rocha et al. [37] introduced a system to assist individuals with aphasia to achieve two-way communication.

Unlike the previous works, our application offers full sentences for users to select instead of a single word or a single phrase. Inspired by TryTalk [14], our application also prioritizes frequently clicked sentences relevant to the detected objects. Our application leverages the image capturing, hand tracking, visualization, and audio capabilities of the HoloLens 2 to realize a novel and integrated AAC interface in augmented reality.

2.2 Computer Vision-Based AAC

Computer vision has been applied for AAC. The computer vision-based AAC applications primarily lie on eye tracking, blink recognition, head tracking, facial detection, and sign language recognition [31].

Raudonis et al. [35] proposed an affordable eye-tracking system that uses a webcam and artificial neural classifiers to achieve precise eye-tracking. Al-Rahayfeh et al. [2] surveyed eye-tracking and head movement technologies, demonstrating their potential for enhancing the accuracy and reducing the costs of assistive technologies. Jen et al. [20] proposed a wearable, highly accurate and robust eye-gaze tracking system, which only required one single webcam mounted on the glasses. On the other hand, Al-Kassim et al. [1] designed an eye-tracking scanning keyboard to help individuals with paralysis. Moreover, Zhang et al. [45] developed GazeSpeak, an eye gesture communication system that operates on smartphones and benefits individuals with motor impairments. Fiannaca et al. [13] presented AACrobat, a Gaze-Based AAC to lower communication barriers and provide autonomy using mobile devices. For more recent research on eye-tracking, please refer to a recent review [23].

For other AAC applications based on blink recognition, head tracking, facial detection, and sign language recognition, please refer to this review [31].

Disparate previous AAC research that used computer vision for communication purposes, we leverage computer vision to drive our application: object detection analyzes the context in a scenario, and the detection result hints what items the user is probably concerned about, helping the user quickly generate context-aware sentences.

2.3 Augmented Reality for AAC

Augmented reality (AR) for AAC is a relatively new research field. Ramires et al. [34] proposed a system that utilizes AR and integrates AAC along with Applied Behavior Analysis (ABA) for aiding interventions with children diagnosed with Autism Spectrum Disorders (ASD). Kerdvibulvech et al. [22] proposed a three-dimensional augmented reality-based human-computer interaction application to assist children with special problems in communication. Also, other research works [3,9–11,17,26,27] show that AR can be used to improve language and communication skills in individuals with ASD and has positive outcomes such as increased motivation, attention, and learning new tasks.

The direction of using HoloLens for AAC applications is relatively unexplored. Zhao et al. [46] proposed an AAC application that runs on HoloLens to use eye-gaze technology to select words and make sounds. Krishnamurthy et al. [24] introduced HoloType, a prototype system aimed at enhancing communication outcomes for individuals with nonspeaking autism to deliver interactive educational content, enabling users to concurrently enhance their pointing skills.

Compared to previous works, HoloAAC not only aims to help AAC users in daily grocery shopping scenarios but also aims to speak a meaningful sentence rather than a word or a phrase.

2.4 User Interface and Interaction for AAC

Several design efforts focused on user interfaces and interactions to support AAC applications. Sobel et al. [41] found that higher-resolution displays enhance AAC applications. Gibson et al. [15] extracted design requirements from a clinical AAC tablet application. Kristensson et al. [25] proposed a design engineering approach for quantitatively exploring context-aware sentence retrieval. Besides, Obiorah et al. [30] developed meal-ordering prototypes for people with aphasia to dine in restaurants. Mitchell et al. [28] investigated a custom-designed optimized keyboard alongside the widely used QWERTY keyboard for three individuals experiencing dexterity impairments caused by motor disabilities.

Wearable devices, like smart glasses, offer convenience compared to handheld devices. In contrast to using a handheld AAC device, using an AAC app visualized through an AR headset allows the user to maintain better eye contact in face-to-face conversations. Additionally, an AR headset can sense the surroundings and provide scene-aware conversational assistance. Devices like HoloLens have provided a glimpse into the future prevalence of such devices. Thus, it is valuable to integrate AR with AAC. We devised HoloAAC as a prototype to explore realizing an AAC interface on smart glasses. We believe that combining computer vision, natural language processing, and text-to-speech technologies in AAC shows promise as an emerging research direction.

3 Interview with AAC Users

To devise a friendly, accessible, and practical application for AAC users, we interviewed 2 professional AAC users who have been using AAC devices for more than 3 years and also teaching people to use AAC devices. We obtained the following insights about the design of this application.

- This application should be portable and the device running the application should be untethered (A1).
- This application should be easy to use with minimal configurations and intuitive operations (A2).
- Considering that some AAC users are used to symbol-based or text-based AAC tools, it is preferable to use similar symbols in this application (A3).
- This application should be friendly to those AAC users with listening disabilities (A4).
- For the grocery shopping scenario, it would be convenient to automatically detect items and support the user in selecting items (A5).

We devise our augmented reality AAC application, HoloAAC, based on the above observations. The application runs on the Microsoft HoloLens 2 (A1). It comprises three windows: an entry window, a network setting window, and a main window (A2). In this application, we support setting voice speed, volume, and voice type (male voice/female voice). Besides that, since computer vision can be used in context-aware AAC to determine what objects of interest are in

the environment [31], we use computer vision techniques to detect groceries and provide an optional way to select/deselect groceries (A5). In addition, the application also tracks the user's sentence selection history to prioritize previously selected sentences (A2). Our application employs the wireless network to realize the portable goal (A1). We choose mid-air tapping as the interaction method. To enhance intuitiveness, we add symbols in front of nouns (A3). In order to make this application more accessible, we use red color to denote being selected. What's more, we set the pressed sentence's color to red to indicate that it is being spoken, which is more friendly for people with listening disabilities (A4).

4 Technical Approach

4.1 Overview

Figure 2 shows our application workflow. First, the user wearing a HoloLens 2 takes a picture of the groceries in front. The picture is then sent to the server (a PC in our experiments) for processing: semantic segmentation, object detection, sentence retrieval, and text-to-speech. When HoloLens 2 gets the response from the server, the user can select one or more keywords to quickly locate the desired sentence and trigger the device to speak it.

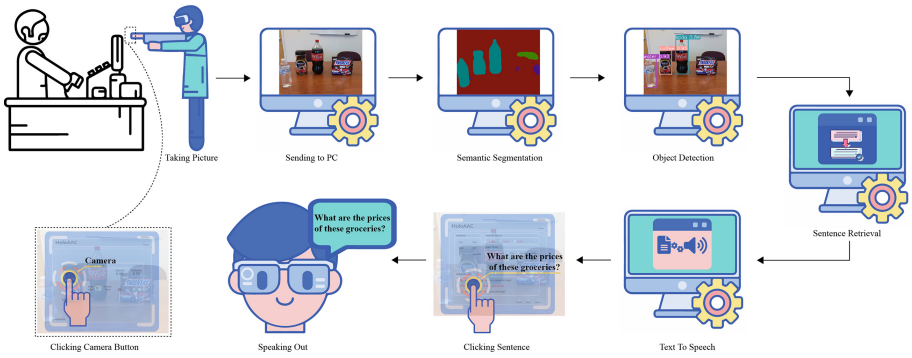

Fig. 2. Our application's overview

4.2 AR Tool and User Interface

As aforementioned, our application runs on Microsoft HoloLens 2. We use the Unity and the Mixed Reality Toolkit (MRTK) to develop the application. HoloLens 2 supports hand tracking so the user interface (UI) is movable in the 3D space. The UI primarily includes three parts: entry UI (Fig. 3(a)), network setting UI (Fig. 3(b)), and main UI (Fig. 3(c)).

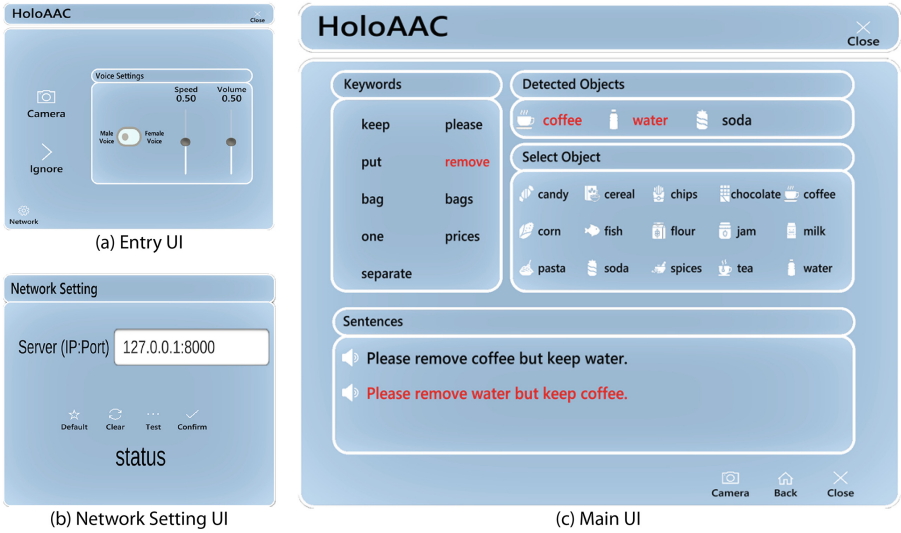

Fig. 3. HoloAAC UI

The entry UI will be displayed after launching our application on Microsoft HoloLens 2. The *Camera* button is for starting the prototype by capturing an image. It will trigger the main UI. The *Ignore* button allows the user to start the prototype without capturing an image. In the *Voice Settings* panel, the user can choose the voice type: male voice or female voice, set the speed of the speech, and change the volume of the sound. On the left bottom, we design the *Network* setting button to support network configuration, which will trigger the network setting UI.

In the network setting UI, the *Default* button is for setting the server input field with the default value. The *Clear* button is for clearing the content in the server input field. The *Test* button is for testing whether the server is accessible. If the server is accessible, the *status* will change to *OK*. Otherwise, the *status* will change to *FAIL*. The *Confirm* button is for setting the server configuration and closing this panel.

The main UI is where the detected objects, keywords, and sentences are shown. The top *Detected Objects* panel shows the detected objects in the captured picture. The left *Keywords* panel displays keywords related to the selected objects in the *Detected Objects* panel and the *Select Object* panel. The central *Select Object* panel lists all the objects that are supported. In case the object detection fails and therefore no object is detected and automatically selected, the user can still select any object in this panel manually. To enhance understanding, we add a symbol in front of each object's name. The bottom *Sentences* panel shows relevant sentences retrieved according to objects and keywords. When the user presses one sentence, the application will speak the sentence.

At the bottom right, there are three buttons: *Camera*, *Back*, and *Close*. This *Camera* button performs the same action as the camera button in the entry UI. The *Back* button is used to go back to the entry UI. The *Close* button is used to quit this application. We use a red color to denote the selected objects, keywords, and sentences in the *Detected Objects* panel, the *Select Object* panel, *Keywords* panel, and *Sentences* panel. Note that the *Detected Objects* panel, the *Select Object* panel, and the *Keywords* panel support multi-selection.

4.3 Object Detection

As aforementioned, we take the image captured by the HoloLens 2 as the input. The next step is to detect possible objects in the image.

Semantic Segmentation. Although object detection can be directly employed on grocery items, detection failure may happen in practice due to the difference between training images and captured images by HoloLens 2. Therefore, we apply semantic segmentation as a preprocessing step to improve object detection accuracy. In our approach, we first apply a semantic segmentation method (Deeplabv3+). It generates object masks that will be applied to crop the captured image into multiple smaller segments for object detection.

Object Detection. Inspired by the GroceEye[1], to perform object detection of grocery items, we fine-tune a YOLOv5 model with the Freiburg Grocery dataset. We use the processed Freiburg dataset which can be downloaded from Github.

4.4 Relevant Sentence Retrieval

We interviewed two professional AAC users, who have been using AAC devices for more than 3 years and also teaching people to use AAC devices, for their opinion regarding common conversations in grocery shopping scenarios. We abstracted them and made them extensible to support adding other sentences easily. We devise a sentence database to construct object-relevant sentences. Since the number of sentences with regard to every object is large, it is hard for a user to locate the target sentence. Therefore, we tokenize and stem sentences to get keywords, which are used to group sentences. Hence the user can select the target sentence through selecting keywords. We also consider historical data, that is, which sentences are selected by the user before, to sort the sentences. As a result, the more times one sentence is selected, the higher the precedence of showing that sentence is. After the sentences are confirmed, the text-to-speech engine will synthesize the corresponding audio of speaking the sentences.

[1] http://students.washington.edu/bhimar/highlights/2020-12-18-GrocerEye/.

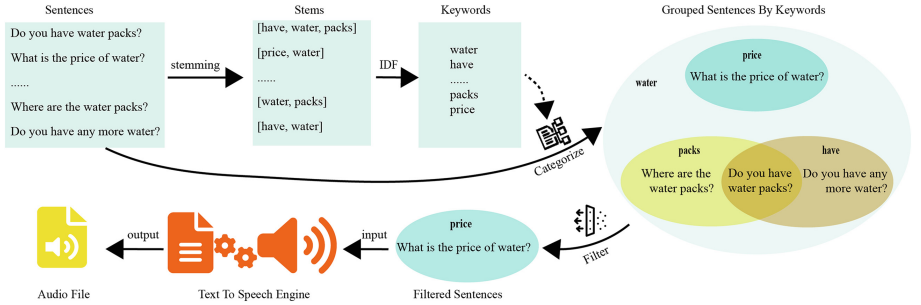

Fig. 4. Sentence retrieval overview

Overall Workflow of Sentence Retrieval. After detecting the objects on the image, our approach retrieves relevant sentences that the user may want to speak. As illustrated in Fig. 4, our method first retrieves all sentences containing the detected grocery names. After removing stop words and punctuations, it extracts the stems of each sentence. We use the IDF algorithm to obtain keywords. The sentences will then be categorized by the keywords. The user could click further keywords on the UI, which will then trigger our approach to filter out any irrelevant sentences. Those sentences that pass the filter will be processed by the text-to-speech engine to generate the audio files of the spoken sentences.

Keywords Generation for Locating Sentences. As we have the sentences of one or more items, the next step is to enable the user to quickly select the target sentence. First, for every sentence, we tokenize the sentence, removing punctuations and stopwords. In NLP, stopwords refer to those words that do not add much meaning to a sentence, such as "a" and "the". After that, we get the stem for every sentence. Then, we vectorize the sentences based on the occurrences of words. The result will be a count matrix. We apply the IDF algorithm to get the words with high frequency. In NLP, IDF means inverse document frequency. IDF is a common term weighting schema in information retrieval. A token with a higher IDF weight has a lower frequency, and vice versa. In our approach, we use the top-ten lower IDF weight tokens as the keywords. It will split sentences into several groups.

Sentence Filtering. After we get both the object name(s) and the keywords, we are able to filter the sentence database. First, we filter the subset of the entire sentence dataset using the object name(s). Sentences irrelevant to the objects will be removed, while those relevant will be kept. Then, we filter the subset again with the keywords. After that, we obtain several target sentences that the user may prefer. To adapt to the user and personalize our approach, we record the sentences the user has selected before. This data is a kind of prior knowledge. When the user selects the same objects and the same keywords next time, the

Table 1. Our seven participants (2 females and 5 males) had different years of experience using different AAC devices. Note that for P6 and P7, the years of AAC experience are counted as zero as they only used the conventional typing approach on cellphones/iPads for communication.

Participant	Gender	AAC Device	Years using AAC	VR/AR Experience
P1	F	Proloquo4Text	5	No
P2	M	ASL Interpreter	5	VR
P3	F	EZKeys	20	No
P4	M	Proloquo2Go	11	VR
P5	M	NovaChat 8	5	No
P6	M	Cellphone/iPad	0*	VR
P7	M	Cellphone	0*	No

sentences will be sorted according to this data. The more times a sentence has been selected, the higher precedence of appearance the sentence is given.

4.5 Error Handling

Computer vision techniques such as semantic segmentation and object detection could fail in some circumstances, for example, due to motion blur caused by the user's head movement or varying light conditions. We devise our application to tolerate such situations if semantic segmentation or object detection fails. In such situations, our approach leaves the *Detected Objects* panel empty and fills the sentences panel with sentences with no object specification. The user can select listed objects in the *Select Object* panel to retrieve relevant sentences.

5 Case Studies

As disability simulations might introduce negative stereotypes and fail to highlight infrastructural and social challenges [4], we recruited people with expressive language difficulties for case studies. According to the American Speech-Language-Hearing Association, about 0.60% of the population use AAC[2]. Inspired by AACrobat [13], we formed case studies where we observed a small group of people with expressive language difficulties who used HoloAAC to complete tasks. We then obtained the users' feedback. According to the local standards for sample size in computer-human interaction studies [6], considering the COVID-19 pandemic, the study setting, and the availability of participants, we recruited 7 participants. This sample size follows the highly expert recommendations ranging from 4 ± 1 to 10 ± 2 [6]. P1, P2, P3, and P4 are local, while P5, P6, and P7 are non-local. P1 is blind in her right eye. P2 is deaf. P4 has a lower-limb disability. P6 has aphasia. P7 has aphasia and hemiplegia. Table 1 shows their demographics.

[2] https://www.asha.org/njc/aac/.

Fig. 5. Proloquo2Go Symbol (PS)

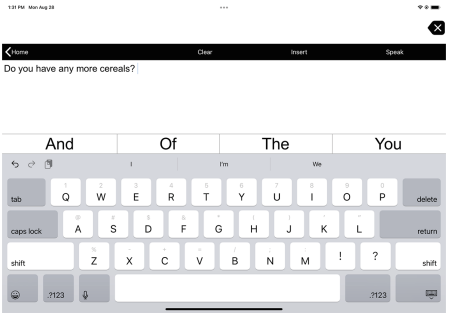

Fig. 6. Proloquo2Go Typing (PT)

Since Proloquo2Go[3] is a popular AAC application on iPhone and iPad for people with expressive language difficulties [12], we let participants complete the same tasks using it as a baseline to investigate the usability and feasibility of our application. Considering the comfort, IRB regulation, safety, convenience, and privacy of AAC users, we conducted the case studies in a simulated environment for P1, P2, P3, and P4. We used a private room inside a lab and set up an environment similar to a grocery store cashier. P5 lives 400 mi away, so we traveled to his home for the case study. Similarly, we drove to an aphasia rehabilitation center 100 mi away to conduct case studies for P6 and P7.

5.1 Implementation

We developed HoloAAC using a PC equipped with a Nvidia GTX 3070 GPU, running Unity 2020.3.20, Microsoft Visual Studio 2019, Anaconda3, and PyCharm 2021.2.3. The backend services such as image processing and sentence retrieval also run on this PC. The prototype runs on a Microsoft HoloLens 2. For fine-tuning the YOLOv5 object detection model, we used a PC with a Nvidia GTX 3090 GPU.

5.2 Procedure

Control Groups. We used two control groups: Proloquo2Go Symbol (Fig. 5) and Proloquo2Go Typing (Fig. 6) since these two modes are frequently used by AAC users. In our case study, Proloquo2Go runs on an iPad.

Warm-Up Session. We conducted a warm-up session to get participants familiarized with the basic operations of Proloquo2Go and our application as well. To let them get ready for the formal case study tasks, the warm-up session comprised of two tasks. The two warm-up tasks were the same, except that

[3] https://www.assistiveware.com/products/proloquo2go.

we assisted them in finishing the first task while they finished the second task independently. For counterbalancing, the participant did the tasks in different orders. For example, if the participant did Proloquo2Go Symbol, Proloquo2Go Typing, and HoloAAC for the first task, the participant would do the second warm-up task in a different order: e.g., HoloAAC, Proloquo2Go Symbol, and Proloquo2Go Typing.

Table 2. Target sentences used for the six tasks. To avoid confusion, we used *bag* in Proloquo2Go Typing and HoloAAC, and *plastic bag* in Proloquo2Go Symbol as the *bag* symbol in Proloquo2Go was not a plastic bag. Also, as Proloquo2Go did not have the plural form symbol of *bag* and *soda*, we used the singular form. Besides, in Proloquo2Go Symbol, we omitted the punctuations of the target sentences for simplicity.

Task	Item(s)	Proloquo2Go Typing and HoloAAC	Proloquo2Go Symbol
1	water	What is the price of water?	What is the price of water
2	soda	Do you have six-packs of soda?	Do you have six-packs of soda
3	coffee	Do you have any more coffee?	Do you have any more coffee
4	soda	Put all the sodas in one bag	Put all the soda in one plastic bag
5	water, soda	Can you put the water and soda in one bag?	Can you put the water and soda in one plastic bag
6	water, coffee, soda	Can you put these groceries in separate bags?	Can you put these groceries in separate plastic bag

Table 3. Task completion times and time analysis (Unit: second) of the participants. HL, PS, and PT denote the HoloAAC, Proloquo2Go Symbol, and Proloquo2Go Typing conditions. SD denotes standard deviation.

Participant	Task 1			Task 2			Task 3			Task 4			Task 5			Task 6			Mean			SD		
	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT
P1	**18**	63	20	**12**	32	19	30	47	**16**	21	90	**15**	**11**	84	20	**18**	81	23	**18**	66	19	7	23	**3**
P2	40	66	**9**	12	65	**8**	40	24	**9**	32	86	**14**	52	102	**18**	27	93	**23**	34	73	**13**	14	28	**6**
P3	**33**	84	34	**15**	137	22	78	48	**23**	39	147	**23**	**15**	147	30	**39**	114	43	37	113	**29**	23	40	**8**
P4	10	92	**7**	35	47	**21**	43	41	**5**	60	113	**7**	14	107	**9**	10	91	**16**	29	82	**11**	21	31	**6**
P5	27	134	**23**	**15**	116	32	33	143	**27**	33	214	**24**	**14**	190	38	**19**	211	51	**24**	168	33	**9**	47	11
P6	**39**	153	81	**26**	75	82	**38**	214	64	**27**	245	47	**7**	256	64	**15**	105	123	**25**	175	77	**13**	75	26
P7	**12**	156	16	31	206	**29**	35	100	**22**	**30**	221	50	41	84	**38**	**17**	119	57	**28**	148	35	**11**	56	16

Case Study Tasks. As shown in Table 2, we designed 6 tasks with different target sentences, which were also given with counterbalancing. Our application tracked the time spent on different operations (e.g., clicking keywords). However, as Proloquo2Go does not have a timing function, we employed an external timer to count the time for the Proloquo2Go Symbol and Proloquo2Go Typing conditions. For Proloquo2Go Typing, we ended the timer once the user had typed the entire sentence. For Proloquo2Go Symbol, we ended the timer once the user had typed the last symbol.

Questionnaire. After the last Proloquo2Go Symbol/Typing and the HoloAAC tasks, we asked the participants to finish a questionnaire to evaluate the workload, using the same iPad which they used to complete Proloquo2Go Symbol/Typing tasks. Besides, we asked them for general feedback. They typed their responses on their AAC devices or phones. We used the NASA Task Load Index (TLX) [16] to assess the subjective workload. It had six questions in total, which were answered using a 7-Likert scale.

Result Analysis. Table 3 shows the task completion times and time analysis of the participants. We can see that P1, P2, P3, and P4 show a more stable ability to type, probably because they type frequently in their daily life. During the case study, they sometimes chose the autocomplete words supplied by the tablet's input keyboard to speed up their input.

Table 4. Mean completion time for each task with HoloAAC.

Task	1	2	3	4	5	6
Mean Time(s)	26	21	42	35	22	21

P1 took similar time using HoloAAC or Proloquo2Go Typing. The completion times with HoloAAC are less than those with Proloquo2Go Typing in 4 out of 6 tasks (Task 1, 2, 5 & 6).

P2 took more time using HoloAAC than Proloquo2Go Typing. We found that it was hard for him to quickly manage to click the target sentence in AR. It took him many attempts to click one sentence to make it speak.

P3 took slightly more time using HoloAAC than Proloquo2Go Typing on average. However, she finished 4 out of 6 tasks (Task 1, 2, 5 & 6) faster using HoloAAC.

P4 took more time using HoloAAC than Proloquo2Go Typing probably due to his many years of experience with Proloquo2Go but no experience with AR.

P5 took less time in 3 out of 6 tasks (Task 2, 5 & 6) using HoloAAC than Proloquo2Go Typing. The mean and standard deviation (SD) show that using HoloAAC is faster than using Proloquo2Go Symbol or Typing. Proloquo2Go and HoloAAC are both new to him. The data shows that he becomes familiar with HoloAAC faster than with Proloquo2Go.

P6 took less time in all 6 tasks using HoloAAC compared to using Proloquo2Go Symbol or Proloquo2Go Typing. From the SD and mean, using HoloAAC is faster than using Proloquo2Go Symbol or Typing.

P7 took less time in 3 out of 6 tasks (Task 1, 4, & 6) using HoloAAC than Proloquo2Go Typing. From the SD and mean, using HoloAAC is faster than using Proloquo2Go Symbol or Typing. Because of hemiplegia, P7 felt hard clicking the sentence precisely and gradually became frustrated as the case study went by. As a result, in the NASA TLX, he gave the same ratings for all questions under

HoloAAC (7), Proloquo2Go Symbol (4), and Proloquo2Go Typing (1) to finish the case study quickly.

We note that the participants generally finished the tasks much faster using HoloAAC than using Proloquo2Go Symbol, even for P1 and P4 who are experienced with Proloquo2Go but not with AR. It seems that choosing keywords/symbols to finish a sentence exactly may take more time than typing especially for experienced typers.

Table 4 shows the mean completion time for each task. We can see that Task 3 and Task 4 are the top two in time consumption as they required the participant to click keywords in AR. More AR mid-air interactions generally resulted in more time needed.

5.3 User Feedback

General Feedback. About our HoloAAC application, all participants said that they liked the automatic popping up of relevant keywords and sentences with respect to the objects detected.

Table 5. NASA TLX workload assessment ratings given by the participants. HL, PS, and PT denote the HoloAAC, Proloquo2Go Symbol, and Proloquo2Go Typing conditions. Please refer to Sect. 5.3 for the findings and explanations.

Participant	Mental Demand			Physical Demand			Temporal Demand			Performance Dissatisfaction			Effort			Frustration		
	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT	HL	PS	PT
P1	4	5	1	3	1	1	4	5	2	3	3	1	3	5	1	1	1	1
P2	5	6	1	5	6	1	2	6	1	4	3	1	5	6	1	2	6	1
P3	7	2	1	7	2	2	3	1	1	3	1	1	7	2	1	2	2	1
P4	2	6	2	4	3	2	2	6	5	2	3	2	3	7	3	3	5	4
P5	7	7	1	7	4	1	2	7	1	4	1	1	7	7	1	7	7	6
P6	3	3	3	5	4	3	5	3	3	4	4	3	3	5	4	3	2	4
P7	7	4	1	7	4	1	7	4	1	7	4	1	7	4	1	7	4	1

P1 appreciated the camera feature for quicker expression but found interacting with the AR interface challenging due to her right eye's blindness.

P2 enjoyed automated sentence generation but found sentence clicking in AR challenging. His unfamiliarity with AR glasses presented some task difficulties but ultimately brought a sense of fulfillment upon completion.

P3 appreciated the automatic object detection and sentence generation functionalities but suggested improving user input responses and enhancing mid-air clicking for smoother interactions. During the case study, it took her multiple attempts to click target sentences but she found a sense of accomplishment after completing tasks. She also recommended extending HoloAAC for hospital use.

P4 appreciated HoloAAC's speed and efficiency but had three suggestions. First, he proposed extending the system to distinguish subtle item differences like colors and sizes. Second, he suggested instant picture-to-speech capabilities

for situations like seeing a cute dog on the street. Third, he recommended personalizing response options based on contexts such as retrieving recent personal stories during conversations with friends.

P5 liked the new interaction approach and felt excited when clicking the expected sentence. However, he found the interaction challenging and time-consuming. He suggested using HoloAAC in schools.

P6 was highly enthusiastic about HoloAAC, seeing its potential for communication and time-saving. He expressed a desire for improved interaction accuracy and recommended broader uses in parks, shops, and schools.

P7 struggled with HoloAAC due to hemiplegia as he could only use his left hand for clicking. He recommended improving accuracy and sensitivity to benefit a wider range of users in the workplace.

NASA TLX. We used NASA TLX to measure the workload. It measures the workload from six aspects: mental demand, physical demand, temporal demand, performance dissatisfaction, effort, and frustration. Table 5 shows the original ratings and Fig. 7 shows the rating plots using the box and whisker plot. 1 represents very low and 7 represents very high. The lower, the better.

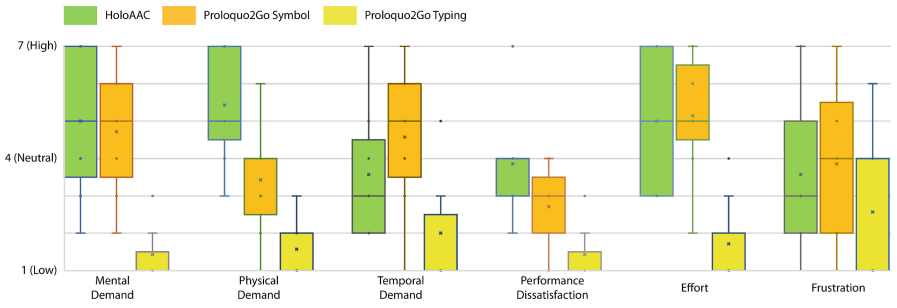

Fig. 7. NASA TLX workload assessment rating plots. Each box and whisker plot comprises six-number summary of the rating: minimum, lower quartile (Q1), median (line), mean (×), upper quartile (Q3), and maximum. Please refer to Sect. 5.3 for the findings and explanations.

P1, P2, and P4 found HoloAAC generally superior to Proloquo2Go Symbol in all aspects. P1 and P2 excelled with Proloquo2Go Typing due to their 5 years of AAC experience. P4 considered HoloAAC on par with Proloquo2Go Typing. P3 rated HoloAAC highly in mental demand, physical demand, and effort due to her 20 years with traditional AAC. P5 gave HoloAAC high ratings in various aspects but struggled due to myopia and the difficulty with sentence selection. P6's ratings were similar for both Proloquo2Go and HoloAAC as both were new to him. P7 assigned the highest ratings to HoloAAC, middle ratings to Proloquo2Go Symbol, and the lowest ratings to Proloquo2Go Typing due to

his limited ability to use only his left hand for tasks. In comparison to other participants, mid-air AR interactions were more physically demanding for him.

Mental Demand. The average rating of HoloAAC is 5, and 4 out of 7 ratings are greater than 4. The reason is that participants needed to focus on the AR panel to be able to interact. On the other hand, all 7 participants hadn't used HoloLens 2 before, but they were more or less experienced in Proloquo2Go or similar devices/applications.

Physical Demand. The average rating of HoloAAC is 5.43, which is even higher than that of the mental demand. 5 out of 7 ratings are greater than 4. The reason is that the task was simple to understand, but the interaction required motion control. Some of the participants had disabilities besides speaking disabilities, which made the physical demand even higher. Another reason is, as Plasson et al. [32] pointed out, the mid-air interaction that HoloLens uses is less accurate than 2D touch and tends to result in physical fatigue.

Temporal Demand. The average rating of HoloAAC is 3.57, a little better than neutral (4); and 5 out of 7 ratings are less than or equal to 4. The reason is that participants didn't feel stressed when performing the tasks. On the other hand, few interactions were needed to complete the tasks using HoloAAC.

Performance Dissatisfaction. The average rating of Holo-AAC is 3.86, a little better than neutral (4). 6 out of 7 ratings are less or equal to 4. Note that only P7 gave a high rating (7) for this aspect. The reason is that P7 did attempt many times to interact with the AR interface because of his hemiplegia. We can say most participants tended to be satisfied with their performance.

Effort. The average rating of HoloAAC is 5, which is equal to the mental demand rating. 4 out of 7 ratings are greater than 4. The reason is that some participants had other disabilities in eyes or motion control, which means that it required more effort for them to finish tasks.

Frustration. The average rating of HoloAAC is 3.57, a little better than neutral (4). 5 out of 7 ratings are less than 4. Most participants didn't feel high frustration when performing tasks using HoloAAC.

In all six aspects, participants gave the lowest ratings to Proloquo2Go Typing. That is because 26-key keyboard-based typing is classic and the participants were more or less familiar with it. Compared to tablet-based AAC applications, HoloAAC running on a HoloLens 2 headset was new to the participants and might be rated unfavorably due to the participants' unfamiliarity with mixed reality. On the other hand, tablet-based AAC applications might have been favored due to the participants' familiarity with tablets.

5.4 Limitations and Future Work

Due to the small population size of AAC users and the challenge that few AAC users were willing to sign up for our case study, our study recruited only seven participants. Hence, we are unable to draw statistically significant conclusions.

We only demonstrate HoloAAC for simple grocery scenarios. The generalizability depends on the underlying object detection model. By using a more versatile model, HoloAAC can function in a wider range of scenarios. We chose to experiment with the grocery scenario for two reasons: 1) the conversations at a cashier tend to be more coherent; and 2) we can leverage contextual information based on grocery items recognized using off-the-shelf computer vision techniques. HoloAAC serves as an early prototype to explore and validate the possibility of integrating AR with AAC. The framework can be extended for other applications. It is technically feasible to enhance the generalizability by incorporating a virtual keyboard or replacing the backend models with contrastive language-image pre-training (CLIP) to satisfy the needs of specific users or application scenarios. Furthermore, the seamless integration of large language models (LLMs) such as ChatGPT into our approach is promising as AI-generated phrases can potentially lessen physical and cognitive sentence creation efforts during communication [43].

Another possible extension is to attach a 4G/5G communication module to enable HoloLens to work without Wi-Fi, which would allow our application to be employed in more scenarios such as supporting outdoor activities. Besides, due to the reality that a standard disabled experience rarely plays out in practice [18], it would be helpful to support multi-modal interactions considering multiple disabilities so as to better accommodate AAC users. For example, for those people with both expressive language difficulties and motion control disabilities, an interaction mechanism based on eye-tracking rather than hand-clicking is more accessible.

For those users who have not used HoloLens, it might take them some time to get familiar with the AR interactions. In our case study, some participants experienced difficulty in clicking the keywords or sentences shown in AR. We believe that improving the hand-tracking precision would make AR-based AAC applications more practical and favorable. Alternatively, instead of using mid-air interactions, using a controller (e.g., the clicker of HoloLens 1) could make interaction easier especially for users with body movement disabilities.

6 Conclusion

We presented HoloAAC, a novel mixed reality-based AAC application. We explored its usability and feasibility through a case study, which provided useful insights for future AR-based AAC applications. First, CV and NLP-based functionalities showed promise and were favored by our participants. Second, other disabilities of AAC users may be considered in designing an AR-based AAC application. Moreover, multi-modal interactions can be incorporated to improve the user experience.

Acknowledgments. We are grateful to the participants for their feedback on our application. This project was supported by NSF grants (award numbers: 1942531 and 2128867).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Al-Kassim, Z., Memon, Q.A.: Designing a low-cost eyeball tracking keyboard for paralyzed people. Comput. Electr. Eng. **58**, 20–29 (2017)
2. Al-Rahayfeh, A., Faezipour, M.: Eye tracking and head movement detection: a state-of-art survey. IEEE J. Transl. Eng. Health Med. **1**, 2100212 (2013)
3. Bai, Z., Blackwell, A., Coulouris, G.: Using augmented reality to elicit pretend play for children with autism. IEEE Trans. Vis. Comput. Graph. **21**, 598–610 (2015). https://doi.org/10.1109/TVCG.2014.2385092
4. Bennett, C.L., Rosner, D.K.: The promise of empathy: design, disability, and knowing the "other". In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2019)
5. Beukelman, D.R., Mirenda, P., et al.: Augmentative and Alternative Communication. Paul H Brookes, Baltimore (1998)
6. Caine, K.: Local standards for sample size at CHI. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 981–992 (2016)
7. Chan, R.Y.Y., Bai, X., Chen, X., Jia, S., Xu, X.h.: IBeacon and HCI in special education: micro-location based augmentative and alternative communication for children with intellectual disabilities. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2016, pp. 1533–1539. Association for Computing Machinery, New York (2016)
8. Chan, R.Y.Y., Sato-Shimokawara, E., Bai, X., Yukiharu, M., Kuo, S.W., Chung, A.: A context-aware augmentative and alternative communication system for school children with intellectual disabilities. IEEE Syst. J. **14**, 208–219 (2020)
9. Chen, C.H., Lee, I.J., Lin, L.Y.: Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. Res. Dev. Disabil. **36**, 396–403 (2015). https://doi.org/10.1016/j.ridd.2014.10.015
10. Chen, C.H., Lee, I.J., Lin, L.Y.: Augmented reality-based video-modeling storybook of nonverbal facial cues for children with autism spectrum disorder to improve their perceptions and judgments of facial expressions and emotions. Comput. Hum. Behav. **55**, 477–485 (2016)
11. Cihak, D.F., Moore, E.J., Wright, R.E., McMahon, D.D., Gibbons, M.M., Smith, C.: Evaluating augmented reality to complete a chain task for elementary students with autism. J. Spec. Educ. Technol. **31**(2), 99–108 (2016). https://doi.org/10.1177/0162643416651724
12. DongGyu, P., Song, S., Lee, D.: Smart phone-based context-aware augmentative and alternative communications system. J. Central South Univ. **21**, 3551–3558 (2014). https://doi.org/10.1007/s11771-014-2335-3
13. Fiannaca, A., Paradiso, A., Shah, M., Morris, M.R.: AACrobat: using mobile devices to lower communication barriers and provide autonomy with gaze-based AAC. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 683–695 (2017)
14. Ghatkamble, R., Son, J., Park, D.: A design and implementation of smartphone-based AAC system. J. Korea Inst. Inf. Commun. Eng. **18**(8), 1895–1903 (2014)

15. Gibson, R.C., Dunlop, M.D., Bouamrane, M.M., Nayar, R.: 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 (2020)
16. Hart, S.G.: Nasa task load index (TLX) (1986)
17. Hayden, C.M., et al.: Augmented reality for speech and language intervention in autism spectrum disorder. Ph.D. thesis (2017)
18. Hofmann, M., Kasnitz, D., Mankoff, J., Bennett, C.L.: Living disability theory: reflections on access, research, and design. In: Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility, pp. 1–13 (2020)
19. Hung, S.W., Chang, C.W., Ma, Y.C.: A new reality: exploring continuance intention to use mobile augmented reality for entertainment purposes. Technol. Soc. **67**, 101757 (2021)
20. Jen, C.L., Chen, Y.L., Lin, Y.J., Lee, C.H., Tsai, A., Li, M.T.: Vision based wearable eye-gaze tracking system. In: 2016 IEEE International Conference on Consumer Electronics (ICCE), pp. 202–203. IEEE (2016)
21. Kane, S.K., Linam-Church, B., Althoff, K., McCall, D.: What we talk about: Designing a context-aware communication tool for people with aphasia. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2012, pp. 49–56. Association for Computing Machinery, New York (2012). https://doi.org/10.1145/2384916.2384926
22. Kerdvibulvech, C., Wang, C.-C.: A new 3D augmented reality application for educational games to help children in communication interactively. In: Gervasi, O., et al. (eds.) ICCSA 2016. LNCS, vol. 9787, pp. 465–473. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42108-7_35
23. Klaib, A.F., Alsrehin, N.O., Melhem, W.Y., Bashtawi, H.O., Magableh, A.A.: Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and internet of things technologies. Expert Syst. Appl. **166**, 114037 (2021). https://doi.org/10.1016/j.eswa.2020.114037. https://www.sciencedirect.com/science/article/pii/S0957417420308071
24. Krishnamurthy, D., Jaswal, V., Nazari, A., Shahidi, A., Subbaraman, P., Wang, M.: Holotype: lived experience based communication training for nonspeaking autistic people. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts, pp. 1–6 (2022)
25. Kristensson, P.O., Lilley, J., Black, R., Waller, A.: A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–11 (2020)
26. Liu, R., Salisbury, J., Vahabzadeh, A., Sahin, N.: Feasibility of an autism-focused augmented reality smartglasses system for social communication and behavioral coaching. Front. Pediatr. **5** (2017). https://doi.org/10.3389/fped.2017.00145
27. Mcmahon, D., Cihak, D., Wright, R., Bell, S.: Augmented reality for teaching science vocabulary to postsecondary education students with intellectual disabilities and autism. J. Res. Technol. Educ. **48**, 1–19 (2015). https://doi.org/10.1080/15391523.2015.1103149
28. Mitchell, C., et al.: Ability-based keyboards for augmentative and alternative communication: Understanding how individuals' movement patterns translate to more efficient keyboards: Methods to generate keyboards tailored to user-specific motor abilities. In: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, CHI EA 2022. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3491101.3519845

29. Mystakidis, S., Christopoulos, A., Pellas, N.: A systematic mapping review of augmented reality applications to support stem learning in higher education. Educ. Inf. Technol. **27**(2), 1883–1927 (2022)
30. Obiorah, M.G., Piper, A.M.M., Horn, M.: Designing AACS for people with aphasia dining in restaurants. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2021)
31. Panchanathan, S., Moore, M., Venkateswara, H., Chakraborty, S., McDaniel, T.: Computer vision for augmentative and alternative communication. In: Computer Vision for Assistive Healthcare, pp. 211–248. Elsevier (2018)
32. Plasson, C., Cunin, D., Laurillau, Y., Nigay, L.: 3d tabletop AR: a comparison of mid-air, touch and touch+ mid-air interaction. In: Proceedings of the International Conference on Advanced Visual Interfaces, pp. 1–5 (2020)
33. Porter, G., Kirkland, J., Spastic Society of Victoria: Integrating Augmentative and Alternative Communication Into Group Programs: Utilising the Principles of Conductive Education. Spastic Society of Victoria (1995). https://books.google.com/books?id=weYGPQAACAAJ
34. Ramires Fernandes, A., Almeida da Silva, C., Grohmann, A.: Assisting speech therapy for autism spectrum disorders with an augmented reality application, vol. 3, November 2014
35. Raudonis, V., Simutis, R., Narvydas, G.: Discrete eye tracking for medical applications. In: 2009 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies, pp. 1–6 (2009). https://doi.org/10.1109/ISABEL.2009.5373675
36. Rauschnabel, P.A., Felix, R., Hinsch, C.: Augmented reality marketing: how mobile AR-apps can improve brands through inspiration. J. Retail. Consum. Serv. **49**, 43–53 (2019)
37. Rocha, A.P., et aloward supporting communication for people with aphasia: the in-bed scenario. In: Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2022. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3528575.3551431
38. Sezer, O.B., Dogdu, E., Ozbayoglu, A.M.: Context-aware computing, learning, and big data in internet of things: a survey. IEEE Internet Things J. **5**(1), 1–27 (2018). https://doi.org/10.1109/JIOT.2017.2773600
39. Shane, H.C., Blackstone, S., Vanderheiden, G., Williams, M., DeRuyter, F.: Using AAC technology to access the world. Assist. Technol. **24**(1), 3–13 (2012)
40. Shen, J., Yang, B., Dudley, J.J., Kristensson, P.O.: KWickChat: a multi-turn dialogue system for AAC using context-aware sentence generation by bag-of-keywords. In: 27th International Conference on Intelligent User Interfaces, pp. 853–867, IUI 2022. Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3490099.3511145
41. Sobel, K., et al.: Exploring the design space of AAC awareness displays. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 2890–2903 (2017)
42. Tzima, S., Styliaras, G., Bassounas, A.: Augmented reality applications in education: teachers point of view. Educ. Sci. **9**(2), 99 (2019)
43. Valencia, S., Cave, R., Kallarackal, K., Seaver, K., Terry, M., Kane, S.K.: "The less i type, the better": how AI language models can enhance or impede communication for AAC users. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2023)

44. Wang, W., Lei, S., Liu, H., Li, T., Qu, J., Qiu, A.: Augmented reality in maintenance training for military equipment. In: Journal of Physics: Conference Series. vol. 1626, p. 012184. IOP Publishing (2020)
45. Zhang, X., Kulkarni, H., Morris, M.R.: Smartphone-based gaze gesture communication for people with motor disabilities. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 2878–2889 (2017)
46. Zhao, H., Karlsson, P., Kavehei, O., McEwan, A.: Augmentative and alternative communication with eye-gaze technology and augmented reality: reflections from engineers, people with cerebral palsy and caregivers. In: 2021 IEEE Sensors, pp. 1–4 (2021). https://doi.org/10.1109/SENSORS47087.2021.9639819