Interspeech 2024
1-5 September 2024, Kos, Greece

Visual Scene Display Application for Augmentative and Alternative Communication

Karthik Venkat Sridaran[1], Raja Praveen[2], Reuben T Varghese[3], Ajish K Abraham[4], Shankar R[5], Winnie Rachel Cherian[6]

[1, 3, 4, 6]All India Institute of Speech and Hearing, Mysore, Karnataka, India
[2, 5]JAIN (Deemed-to-be University), Bengaluru, Karnataka, India

karthik@aiishmysore.in, p.raja@jainuniversity.ac.in, reuben@aiishmysore.in, ajish68@aiishmysore.in, shankarraj9929@gmail.com, winnierachel6@gmail.com

Abstract

Augmentative and Alternative Communication (AAC) refers to various strategies and tools used to enhance communication for individuals with limited speech or language abilities. Visual Scene Displays (VSDs) have emerged as a promising approach within AAC, leveraging visual cues and contextualized scenes to support communication and language development. This show and tell presentation describes the development of a Visual Scene Display application, implemented as a webpage. The application offers a user-friendly interface that allows individuals to create and customize their own scenes by uploading images or capturing real-life scenes using the device's camera. The system incorporates a robust symbol library, including pictograms, icons, and audio to support comprehension and expression. The development process involved iterative design, usability testing, and feedback from AAC professionals, individuals with communication challenges, and their caregivers. This collaborative approach ensured that the VSD system addressed the diverse needs and abilities of its target users. Preliminary evaluations of the VSD system have shown promising results, indicating its potential to enhance communication and language development among individuals with limited speech or language abilities.

Index Terms: Complex Communication Needs (CCN), Augmentative and Alternative communication (AAC), Visual Scene Display (VSD)

1. Introduction

In the pursuit of enhancing communication and fostering inclusivity for individuals with Complex Communication Needs (CCN), the Augmentative and Alternative Communication (AAC) systems have emerged as valuable tools to bridge communication gaps for those with challenges in expressive language. Research indicates that using AAC improves language development for people with communication difficulties [1], [2], [3]. However, the powerful characteristics of high-tech AAC systems (e.g., dynamic storage, voice output, programmable vocabulary) now make maintenance and setup difficult for AAC support personnel, such as teachers, speech-language pathologists (SLPs), and family [4]. AAC systems come with a variety of communication tools, such as speech-generating devices (SGDs), boards, signs, and gestures [5]. SGDs, graphic communication boards, sign language, and other kinds of AAC can all be useful tools for supporting engagement for people with complex communication [6].

Keeping in view the advantages of VSD, a novel web-based platform has been developed by integrating key features and benefits of the VSDs, highlighting its potential to revolutionize how individuals with communication disorders can connect with their surroundings. This web-based platform provides a user-friendly interface for capturing real-time images and speech samples both real time and pre-recorded files for future references which can be downloaded in video format. The welcome page of the application is as shown in Figure 1.

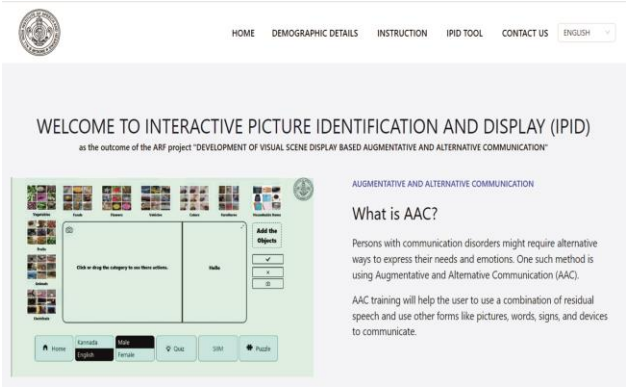

Figure 1: *Home page of the application*

2. Application Overview

The VSD application has a user-friendly interface that enables SLPs, clients, and caregivers to navigate and interact with the web- based application effectively. The application has provision for collecting the demographic details of the client which helps the SLPs in gathering the pre-test information. Anyone can use the application as it does not require the log in credentials. This application supports two languages - English and Kannada.

The application comprises 10 image categories with each category consisting of 10 pre-defined images with an option to add/capture an additional image. A total of 100 images are incorporated which are validated by Speech Language Pathologists (SLPs) to check the appropriateness of the images and text. This application can also add a new category, record audio, captures image using an in-built camera application, language selection (English or Kannada) to read or listen to the content provided, and gender selection in which the client can select the male or female voice. The main view of the application is shown in Figure 2.

The core feature of the application is automatic detection of objects using machine learning technique. COCO-

SSD package has been used for the object detection. This package consists the database with 10 categories and 80 objects which can be identified if captured/uploaded. If the user has captured/ uploaded an image into the application, the COCO-SSD runs in the background and detects few specific objects which are present in its database. Once it detects the objects it draws a green colored box around the part of the image captured/uploaded and names it just above the box as shown in Figure 3.

Figure 2: *Main view of the IPID application*

Figure 3: *Object identification using COCO-SSD Package*

The other features of the application include Speech and Image Integration Module (SIIM), quiz and puzzle applications. SIIM is an advanced application which combines speech-related functionality and image processing capabilities. Two blocks are provided namely multiple images and single image through which the users can integrate image and the audio file/s. In multiple images, the user can upload multiple images on the left side with multiple audio on the right side. All these audio and image files will be merged to form a single video file which will have every image presented till the respective audio is played. As soon as the audio corresponding to the image ends the next image is shown and the next audio is played. The same procedure is followed with the single image where one image and audio file is integrated and a video is generated in which they can describe a scene. Quiz is essential to gather data and analyze the user's response time to evaluate system performance, user engagement, and communication effectiveness. The software provides data for data analysis and visualization, allowing researchers and clinicians to gain insights from the recorded response time data.

In puzzle application, a user can click/upload a picture and choose the level of toughness of the puzzle, the captured picture is divided into 2x2, 3x3, and 4x4 blocks depending on the level of the puzzle they have chosen.

3. RESULTS

Design and development of front end of the VSD system is done using ReactJS and Visual studio code, and then it is integrated with internet and webhosting. To test the application, 10 participants were selected. The reaction time of all the 10 participants was measured for 10 categories of objects before and after training the results clearly indicated that the reaction time scores were better for children after visual scene display-based AAC training compared to those who were not given VSD training. Naming accuracy was also measured 10 categories of objects before and after VSD training. The results clearly indicate that the naming accuracy scores were better for children who received visual scene display-based AAC training compared to those who were not given VSD training as shown in Table 1.

Table 1: *Naming accuracy of the participants before & after training with VSD application*

Categories	**Before training with VSD application (Accuracy in %)**	**After training with VSD application (Accuracy in %)**
Vegetables	50	85
Fruits	52.5	82.5
Elec. Appliances	52.5	85
Furniture	45	85
Colors	55	82.5
Vehicles	52.5	90
Animals	57.5	90
Flowers	62.5	82.5
Food	50	90
Household items	50	85

4. REFERENCES

[1] Binger, C., & Light, J. (2007). The effect of aided AAC modeling on the expression of multi-symbol messages by preschoolers who use AAC. Augmentative and Alternative Communication, 23, 30 – 43 https://doi.org/10.1080/07434610600807470

[2] 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, 112 – 125. https://doi.org/10.1044/1058-0360(2006/012).

[3] Romski, M. A., & Sevcik, R. A. (1996). Breaking the speech barrier: Language development through augmented means. Baltimore, MD: Paul H. Brookes Publishing Co.

[4] Caron, J., Light, J., Davidoff, B. E., & Drager, K. D. (2017). Comparison of the effects of mobile technology AAC apps on programming visual scene displays. Augmentative and Alternative Communication, 33(4), 239-248. https://doi.org/10.1080/07434618.2017.1388836

[5] Thistle, J. J., & Wilkinson, K. M. (2015). Building evidence-based practice in AAC display design for young children: Current practices and future directions. Augmentative and Alternative Communication, 31(2), 124-136. https://doi.org/10.3109/07434618.2015.1035798

[6] Babb, S., McNaughton, D., Light, J., Caron, J., Wydner, K., & Jung, S. (2020). Using AAC video visual scene displays to increase participation and communication within a volunteer activity for adolescents with complex communication needs. AAC: Augmentative and Alternative Communication, 36(1), 31–42. https://doi.org/10.1080/07434618.2020.1737966