Protocol Number:

Video-Assisted Speech Technology-003

Protocol Date and Version #:

2/28/23 - version #3

Study Title

Video Assisted Speech Technology to enhance functional language abilities in individuals with Autism Spectrum Disorder

Investigators (including sub-investigators)

Lois Jean Brady, M.A. CCC-SLP/CAS

Matthew Guggemos, M.S. CCC-SLP/CAS

Address / Phone - Contact details

955 Walnut Ave

Vallejo, CA 94592

Source of external funding

National Institutes of Health (NIH)

Hypothesis

Users of the VR integrated VAST application will show improved functional language abilities and oral motor coordination skills.

Background

Autism Spectrum Disorder (ASD) is a neurodevelopmental communication disorder resulting in functional language and behavioral delays affecting over 3.5 million Americans. These delays vary with the severity of symptoms that present in ASD. Individuals may be verbal and possess stronger functional language abilities whereas 25% of people in the ASD population are minimally verbal or non-verbal, having increased communication challenges. ASD limits many aspects of speech such as vocabulary acquisition where the individual retains a limited lexicon. This restriction on word choice leads to further limitations in the development of sentence structure. Conversation is kept to simple requests and commands and only basic topics are maintained for discussion, preventing the expression of abstract ideas. Alongside linguistic acquisition, oral-motor coordination is a crucial part of speech production. Students with ASD struggle with acquiring the necessary motor skills needed for speech articulation and activities such as chewing.

Current clinical techniques have shown varying degrees of efficacy in improving functional language proficiency. Most techniques follow a drill-like procedure, wherein the child is made to repeat various sounds and phrases until they are retained. However, such a process requires potentially over twenty therapy sessions to show improvement which may then only be focused on one aspect of speech. This significantly limits the linguistic and social skills a student will acquire. To improve the efficacy of these therapy sessions, new technology must be developed to provide the most effective educational experience. Video-assisted speech technology (VAST) is a method of using a video of a close-up model of the mouth and speaking simultaneously with it. Rather than present the individual with a static photograph of the initial phoneme, the entire sequence of oral movements can be presented sequentially via video-recorded segments of the orofacial area producing connected speech, combining best practices, video modeling, and literacy with auditory cues to provide unprecedented support the development of vocabulary, word combinations and communication.

In this SBIR Phase I proposal, iTherapy will develop a personalized educational experience for students with ASD by creating a virtual reality (VR) based VAST program to stimulate engagement and speech production practice. VR offers several benefits as a therapy technique: overcoming sensory difficulties, more effectively generalizing information, employing visual learning, and providing individualized treatment. As a user moves through the stages of the program, they will be immersed in a proactive environment where they will engross themselves with continuous content.

Study Goals and Objectives

Aim 1: Develop a beta product VR-based VAST platform. Currently, the alpha product VAST program utilizes filmed mouths enunciating words and phrases in a 2D setting. Developers will code the VR mouths and the graphic design for all linguistic elements. Coding will comprise of *C++ & Unreal Engine 4* where unreal engine utilizes a node-based language called *Blueprints Visual Scripting.* Further attributes of Unreal Engine include a skeletal animation system combined with higher polygon models. In addition to VR programming, the *Dolch Vocabulary list* will be integrated into the system as a base word list consisting of frequently used English words. Within the list are 220 easily recognizable service words including nouns and non-nouns which target the kindergarten-through-third grade age group. *Deliverable: The alpha product 2D software will be developed into a 3D VR environment to ensure maximum optimization of the product for client use.*

Aim 2: Develop hardware prototype. We will compare the utility of combining off-the-shelf VR goggles & Bone Conduction (BC) headphones with a 3D printing of a single cohesive unit. The strength of the BC headphones will be tested alongside standard headphones to determine if they provide optimal functionality. After testing each option, the product with the highest rate of efficacy will be used to create the integrated prototype*. Deliverable: Functional hardware prototype.*

4.2.c. Interventions

Type	Name	Description
Behavioral (e.g., Psychotherapy, Lifestyle Counseling)	VR-based VAST platform	The VR-incorporated prototype, 3D video-modeled VAST application
Behavioral (e.g., Psychotherapy, Lifestyle Counseling)	2D video modeling	VAST application using 2D video recordings and no hardware prototype

Aim 3: Demonstrate feasibility of VR-based VAST platform. We will pilot test the VR VAST app on 6 children with ASD. The recruited population will consist of 6 ASD children between the ages of 4 and 8 who will be recruited to participate in a 14-sessions-long study that will utilize the VR-integrated VAST application. The subjects will each receive a different piece of equipment: one group of five will receive a 3D VR-integrated VAST prototype, while the other group of five will receive a 2D VAST on a tablet. Sessions will be held twice a week with each lasting approximately 30 minutes (i.e. 30 minutes +/- 5 minutes).

Three main metrics will be measured during the pilot study: *Percentage of Correctly Transcribed Words Using ASR, Articulation Accuracy,* and *Mean Length of Utterance (MLU)*. These metrics will show whether there has been improvement in the participant's functional language abilities, and oral motor coordination skills. They will be

analyzed by several trained staff members who will observe each recorded and monitored session.

Subject Selection

Recruiting: Subjects will be recruited by contacting (e.g. email or telephone) teachers and professionals (e.g. speech-language pathologists and behaviorists) who work with children with autism, aged 4 to 8. The research team will reach out to school administrators, teachers, and specialists through the iTherapy network of contacts to recruit eligible K-2 students diagnosed with ASD and their teacher or specialist. The VR-based VAST platform is designed for use with early elementary school-aged students (grades K-2). In order to conduct the intervention and gather data that would be useful for further development of the product, the research study must include early elementary school students.

In order to recruit enough children within the study's target population and to encourage retention, the study team will use online advertising via social media and other platforms (e.g. Craigslist) to reach participants. All participants will be given an Apple device (iPhone 6 for VR or refurbished iPad mini for 2D) for participating in the study and a free download of InnerVoice (valued at $99.99), an iOS AI-powered communication app that uses video self-modeling, designed and sold by iTherapy.

Number of participants: 6

Inclusion/Exclusion criteria: All participants will be recruited via campaigns posted on social media and/or other platforms: parents/guardians will be provided consent forms and all information regarding inclusion/exclusion criteria.

Children **under the age of 18** will be recruited and all protections mandated by NIH policy afforded to them.

In order to protect participants and their families from COVID-19 exposure, participants will undergo Aim 3's procedures within their homes: the subjects' data will be collected using a HIPAA-compliant telehealth video-streaming platform. The goal of this

procedure is to minimize any potential exposure to infection as a result of participating in this study. The study's team will follow CDC-recommended cleaning and disinfection guidelines to ship and retrieve the prototype VR headset (that is, the bone-conduction headphones along with the goggles sans the iPhones, which the participants can keep as part of participating in the study).

The proposed study will not discriminate based on gender, sexual orientation or race. Participants will be recruited from a pool of children at participating clinics and educational facilities independent of the above-mentioned aspects. There are no exclusion criteria based on gender, sexual orientation, or race.

Because data prior studies strongly support no significant differences of clinical or public health importance in intervention effect between subgroups, then sex/gender and/or race/ethnicity will not be recorded as subject selection criteria.

Children with a medically documented history of seizure disorder will be excluded from the study. According to Carnegie Mellon University Entertainment Technology Center (2019), individuals with a history of seizure disorder should avoid using virtual reality media.

Study procedures / research method

Due to health and safety concerns associated with the COVID-19 pandemic, the feasibility procedures will be conducted through a HIPAA-compliant telehealth platform. The subjects' caregivers or parents will help measure the impact of two pieces of technology: the 3D VR-integrated headset with bone-conduction and the 2D tablet-based application.

In order to prevent the spread of pathogens among large groups of subjects, the number of participants has been reduced from 30 to 10. Ten ASD students between the ages of 4 and 8 will be recruited to participate in a 14-sessions-long study that will utilize the integrated VAST application. The subjects will each receive a different piece of technology: one group of five will receive a 3D VR-integrated VAST prototype, while the other group of five will receive a 2D VAST version on a tablet. Sessions will be held twice a week with each lasting approximately 30 minutes (i.e. 30 minutes +/- 5 minutes).

Three primary metrics and three secondary metrics will be examined during the pilot study: **Change in Percentage of Correctly Transcribed Words Using Automatic Speech Recognition (ASR), Change in Articulation Accuracy, Change in Mean Length of Utterance (MLU)**, **Parent Perceptions of Communication Changes Resulting from Study Participation, Change in Type-Token Ratios,** and **Change in Response Rate to Treatment Stimuli**. The frequency and duration of sessions is

based on our extensive previous research and is intended to ensure a statistically relevant number of data points to measure the learning progress of the students.

Impact will be measured via the following:

Change in Percentage of Correctly Transcribed Words Using Automatic Speech Recognition will be used as a measure of speech intelligibility. Specifically, the closed-captioning ASR on the Google Meet teleconferencing platform will be used as pre- and post-test measures to assess whether an individual has changed their ability to accurately articulate the target words. A higher percentage of correctly transcribed words will indicate articulatory accuracy and will provide a calculation of speech development rate. Due to the COVID-19 pandemic, distance learning will likely be a supplementary or primary medium of service delivery for children on the autism spectrum. Therefore, rating intelligibility on teleconferencing platforms may serve as a functional measure that could reflect subjects' communicative aptitude within a new social environment (i.e. distance learning).

Change in Articulation Accuracy will be calculated by two speech-language pathologists (SLP) transcribing the subjects' spoken production of the target words. Both SLPs will calculate the change in % of correct phonemes in each attempted stimulus by determining the correct number of phonemes in each of the subject's production of the target words.

Change in Mean Length of Utterance (MLU) – participants (aged 4 to 8 years) will be given a pre- and post-test 15-minute language sample. MLU was calculated for tests and gain from pre-test to post-test was compared. This measure is calculated based on a change in the number of morphemes per utterance during pre-test and post-test language samples. During a five-minute period, two licensed speech-language pathologists (SLP) will observe a parent interacting and talking with their child. Both will SLPs transcribe the subjects' speech and calculate a mean length of utterance (MLU) for each subject. MLU will be calculated by determining how many bound and free morphemes are included within every spoken utterance produced by a subject. The total number of morphemes produced within the 5-minute period will be divided by total number of utterances, which will then produce the MLU for each subject. This procedure will be used for determining MLU in both the pre- and post-testing procedures.

Parent Perceptions of Communication Changes Resulting from Study Participation – this measure examined parent observations and perceptions of changes in their children's motor-speech, behavioral, and social communication skills after having participated in the study.

Change in Type-Token Ratios – type-token ratio measures the total number of unique words in a given segment of language. The goal of using this metric is to examine any change in lexical diversity following either the tablet- or VR-based treatment.

Change in Response Rate to Treatment Stimuli – Response rate gain will be measured to determine whether there are any significant changes in how often children responded to pre- and post-testing stimuli after having received treatment between the iPad Pro and VR goggles groups.

Each session will be monitored and recorded. Two members of the trained staff will then observe these videos and analyze the rates of the aforementioned metrics.

Outcome Measures

Type	Name	Time Frame	Brief Description
Primary	**Change in Percentage of Correctly Transcribed Words Using Automatic Speech Recognition**	Seven weeks--each subject participated in the study twice a week over a 7-week period for a total of 14 sessions. The first and last sessions (session #1 and session #14) were reserved for pre-test and post-test language sample collection and assessment.	15-minute pre- and post-testing was performed using speech recognition software and transcribed by a licensed speech pathologist. Differences pre- and post-intervention were compared across groups and within groups.
Primary	**Change in Articulation Accuracy**	Seven weeks--each subject participated in the study twice a week over a 7-week period for a total of 14 sessions. The first and last sessions	Change in % of correct phonemes in each attempted stimulus

		(session #1 and session #14) were reserved for pre-test and post-test language sample collection and assessment.	
Primary	**Change in Mean Length of Utterance (MLU)**	Seven weeks--each subject participated in the study twice a week over a 7-week period for a total of 14 sessions. The first and last sessions (session #1 and session #14) were reserved for pre-test and post-test language sample collection and assessment.	Participants (aged 4 to 8 years) were given a pre- and post-test 15-minute language sample. MLU was calculated for tests and gain from pre-test to post-test was compared.
Secondary	**Parent Perceptions of Communication Changes Resulting from Study Participation**	Seven weeks--each subject participated in the study twice a week over a 7-week period for a total of 14 sessions. The first and last sessions (session #1 and session #14) were reserved for pre-test and post-test language sample collection and assessment	Parent observations -- perceptions of changes in their children's motor-speech, behavioral, and social communication skills after having participated in the study

Secondary	**Change in Type-Token Ratios**	Seven weeks--each subject participated in the study twice a week over a 7-week period for a total of 14 sessions. The first and last sessions (session #1 and session #14) were reserved for pre-test and post-test language sample collection and assessment.	A type-token ratio measures the total number of unique words in a given segment of language.
Secondary	**Change in Response Rate to Treatment Stimuli**	Seven weeks--each subject participated in the study twice a week over a 7-week period for a total of 14 sessions. The first and last sessions (session #1 and session #14) were reserved for pre-test and post-test language sample collection and assessment.	Response rate gain measured whether there were any significant changes in how often children responded to pre- and post-testing stimuli after having received treatment between the iPad Pro and VR goggles groups.

Risk / Safety Information

Participants in this study will face little to no harm, risks, or discomfort. Participants will undergo the Impact Metrics within their own homes via a HIPAA-compliant telehealth video stream in order to minimize the risk of COVID-19 exposure. Additionally, all equipment affiliated with the study will be cleaned and disinfected, following CDC guidelines, prior to the subjects receiving the equipment.

Participants will face no greater than minimal risk during the implementation of Aim 3's Impact Metrics. Each assessment session will last 30 minutes (+/- 5 minutes) twice per week. During the course of each assessment session, the participants will interact with known adults to decrease the chances of them feeling uncomfortable. As the software is

100% virtual, no adverse effects, risks, or discomforts are anticipated, other than what the participant may experience looking at a typical tablet screen or wearing VR goggles. Additionally, the hardware used in this study is designed to be worn for long periods of time with minimal to no discomfort.

However, the research team recognizes that it is possible that some participants may experience feelings of frustration, embarrassment, or fatigue through participation due to the amount of sensory integration within the particular treatment. It is recognized that the issue of sensory and stimulus overload is common in children with ASD. Children on the spectrum sometimes experience inability to focus or distress when exposed to a surplus of stimuli. Avoiding this overload is not a matter of simplifying learning materials but rather managing the perceptual load children are exposed to.

sensory and stimulus overload is common in children with ASD. Children on the spectrum sometimes experience inability to focus or distress when exposed to a surplus of stimuli. Avoiding this overload is not a matter of simplifying learning materials but rather managing the perceptual load children are exposed to.

Should any Adverse Events (AEs), Serious Adverse Events (SAEs) — injuries, seizures, or emotional distress — and Unanticipated Problems Involving Risks to Subjects or Others occur, a session will be ceased immediately and a report submitted to the IRB within 48 hours by the on-site researcher. If an SAE were to occur — such as an injury or seizure — staff will contact emergency medical services to ensure the subject's safety. Staff will receive training in communication and behavioral de-escalation strategies, whose aim is to reduce subjects' exposure to emotional distress. The PI and Co-investigator, both certified autism specialists, will be responsible for implementing the training.

All stimuli and engagement topics will be thoroughly vetted to reduce any discomfort on the part of the participant. Research staff will be advised to immediately cease engagement upon any visual sign of participant discomfort. Participants will only participate in the study as a supplement, and not a replacement, of prescribed or existing therapies.

Monitoring and reporting of Adverse Events/Serious Adverse Events

The research team will monitor subjects' exposure to adverse events/SAE throughout the study. For example, we will ensure that the methods of data collection do not invade the privacy of students, parents, teachers, or schools. We will monitor participation and responses to oversee and reduce potential feelings of frustration or embarrassment. Additionally, all children using the VR headsets will be required to sit so as to avoid

falling or suffering other injuries that might occur from standing or walking while wearing the headsets.

There are no events that would preclude a participant from continuing the intervention beyond withdrawal of assent or consent. As detailed in the Protections Against Risk above, the study presented no greater than minimal risk to participants. Trials will be stopped upon any visual observance of discomfort of the participant by the educator or trained professional (as detailed in the Protections Against Risk above). Data will be reported from all sites upon session completion – compliance and monitoring will be ensured by an on-site researcher who will collect session data. Further data security measures are detailed above in Collaborating Sites.

Any IRB and/or ISM/DSMB actions will be reported to the NIH within 72 hours of occurrence by the project PI Lois Brady.

Study Oversight

Data and safety will be monitored per NIH guidelines by project PI and IRB. Individual responsible for monitoring are:

1. Lois Brady (project PI). Ms. Brady will ensure data security on a bi-weekly basis. Her position as CEO of iTherapy will include the monitoring of all data acquired from this study. She will ensure that all staff are adequately trained in the study protocol.

2. Matthew Guggemos (Co-Investigator). Mr. Guggemos will ensure safety on-site on a weekly basis. As the Chief Operations Officer at iTherapy, he is qualified to ensure data and safety
during data collection.

3. IRB will ensure protocol adherence to acceptable safety practices.

Data Management

[Who will be collecting, analyzing the data? Do you plan to have an independent data analysis? Data retention (i.e. how long so you plan to maintain records for this study). Include statistical analysis plan (if applicable)]

Data formats — data will be stored in commonly-used file formats (e.g., .csv for survey data, .mp3 for audio files, .mov for screen recordings). There are no applicable metadata standards that will be applied to these data.

Dissemination — all data generated in this SBIR Phase I project is considered proprietary. A project report will be made available partial and/or aggregated form, upon request, to the intervention developers. A subset of the project report will be made available, upon request, to partner school districts or other organizations who assisted in recruiting participants.

While all efforts will be made to keep data confidential, there is a small chance that participation in research may cause a loss of privacy. There is also a small risk of data loss and misuse or breach of private information. No physical, social, cultural, financial, legal, or other risks to subjects are expected from this study.

The data will be kept for ten years beyond the life of the project before all the files are destroyed, to allow for future longitudinal studies. While the data are kept for these additional years, the same storage protocols described above will be followed.

For the different group, electronic images (e.g., pdf), spreadsheets, and database files

Statistical Design and Power

We have used the following formula to calculate power and size of an experiments where there are >2 means:

$$n = 2\left(\sigma\frac{z_{1-\alpha/(2\tau)} + z_{1-\beta}}{\mu_A - \mu_B}\right)^2$$

$$1-\beta = \Phi\left(z - z_{1-\alpha/(2\tau)}\right) + \Phi\left(-z - z_{1-\alpha/(2\tau)}\right) \quad , \quad z = \frac{\mu_A - \mu_B}{\sigma\sqrt{\frac{2}{n}}}$$

Assumptions:

Power, 1−β : 0.08

Type I error rate, α: 0.05

μA = 1

μB = 2

σ = 1

τ = 1

Source: Chow S, Shao J, Wang H. 2008. Sample Size Calculations in Clinical Research. 2nd Ed. Chapman & Hall/CRC Biostatistics Series. page 71.

MLU, SIS, and articulation rate between groups will be tested using an ANOVA with pairwise comparisons.

may be created from paper-based field notes as researchers code data into an electronic form and de-identify (anonymize) response records. Data will then be protected as mentioned above.

IRB Review / Ethics / Informed Consent

As participants are children, all additional protections for children involved as subjects in research will be observed. We do not expect the study to involve greater than minimal risk and will obtain IRB approval of adequate provisions for soliciting the assent of the children and the permission of their parents or guardians.

The study involves children with ASD. Informed written consent from parents/guardians of the child participants is required prior to beginning any part of the study. Parents/guardians will be present at all times during the testing and will be able to stop

the activities at any time for any reason. Children will be asked for verbal consent. Participants will be informed that they can stop the sessions at any time for any reason. Because of the varying stimulatory thresholds of those with ASD, participants will be closely monitored by their teachers or health professionals for any visual or verbal sign of discomfort. Should any sign be observed, the participant's session will immediately be ceased.

Confidentiality

The confidentiality of all participants will be maintained throughout the life of the project, and the research team will institute and monitor robust data security protocols. All members of the research team will complete an online course covering research with human subjects and data security prior to the study's commencement. Data will be collected and stored on a secure online server. A code key will be created to assign a unique number to each study participant (i.e. P001). The data collected in this study will in no way be affiliated with the participants' medical record number or personnel files. All participant data will be stored on a secure, encrypted data storage device and/or a password protected and encrypted internet- or cloud-based storage system, which meets HIPAA privacy standards. The files will also be password encrypted. No data will be accessible from any source that is not password encrypted. The password will only be known to IRB-approved research study members.

Intended Use of Data

We will use the data to examine whether the VAST VR system helps children with autism between the ages of 5 to 8 years improve their speech intelligibility, impact their articulation rate, and expand their MLUs. Further, the data obtained in this study will guide future design improvements for the VAST system. The overall intended use for the VAST system will be in special education classrooms for children who struggle with functional language acquisition and motor abilities. Students will be able to improve upon their oral coordination during their in-school therapy sessions which may last between one to three hours. The skills they learn will then be applied to their classwork and their daily interactions with peers and teachers. As their language abilities improve, teachers and students will be able to improve their cooperation and social interaction. By using the VAST system, students will be better able to understand and process the speech of their teachers and become more integrated in the classroom setting. Each student who participates in special education will require the use of a tablet or smartphone, VR goggles, and BC headphones.

The knowledge gained by this study has the potential to provide students with ASD with a comprehensive language education. The risks noted in the above section are minimal and reasonable in light of the benefits of the proposed study. Participation in this study will benefit all participants by supplementing their existing in-school therapy sessions while providing minimal to no risk.