Journal of Otology 20 (2025) 12–19

Journal of Otology

journal homepage: www.sciopen.com/journal/1672-2930

Developing a novel assistive technology, empowering the deaf to speak, using visual feedback.

Singh S[1] ✉, Nagarajan K[2], Chachan R[3], Pandey P[3], Vijayalakshmi P[4], Verma N[5] and Singh A[6]

[1] *Amrita Institute of Medical Sciences and Research Centre Mata Amritanandamayi Marg, Sector 88 Faridabad, Haryana 121002, INDIA*

[2] *Couth IT, Hyderabad, India*

[3] *4S Medical Research P Ltd, New Delhi, India*

[4] *SSN College of Engineering, Chennai, India*

[5] *Manipal University College Malaysia, Melaka, Malaysia*

[6] *Henrico, Parham & Retreat Hospitals, Richmond, Virginia, United States*

ABSTRACT

Background: Profoundly deaf children beyond 5 years of age and deaf adults with no prior access to hearing have limited options to learn oral speech. They rely mainly on sign language for communication. This is because, in the absence of hearing, the brain of a deaf person is not able to process any feedback of its own speech efforts. Method: We have developed a novel assistive technology that provides visual feedback of speech efforts to the brain of a deaf person learning to speak on a smart phone or computer, in the absence of auditory feedback. We propose that a deaf person with compensatory heightened visual processing skills, should be able to use this visual feedback to develop speech. Results: The first prototype of our novel assistive technology was used on 72 deaf persons (children and adults) attending deaf schools in Delhi. They progressed from speaking no sounds to about 18 sounds in 6 months. Conclusion: Assistive technology using sensory substitution is a promising new approach to help deaf children and adults develop oral speech using visual feedback. More research needs to be carried out in developing this technology further.

KEYWORDS

deafness, sign language, cochlear implants, speech to text, deaf rehabilitation, deaf speech.

1 Introduction

Deafness is a significant disability worldwide (World Health Organization 2018). The main treatment options for any child born profoundly deaf are Cochlear Implants (Deep et al. 2019; Carlson et al. 2018), Sign language (Rosen 2010) and Total communication skills. (Thomas & Zwolan 2019).

Cochlear implantation, ideally at the age of 1 year (up to 3 years), followed by rehabilitation, enables a deaf child to develop hearing, speech and language using electrical stimulation of the inner ear. (Naples & Ruckenstein 2020; Aimoni et al. 2018). Late cochlear implantation, especially after the age of 7 years, helps children identify environmental sounds but results in very limited and variable hearing and speech development. (Deep et al. 2019; Monteiro et al. 2016; Singh et al. 2015; Singh et al. 2004; Chu et al. 2020; Sharma et al. 2020; Buchman et al. 2020) Sign language is very successful and has been the intervention of choice for the deaf in many countries who have developed advanced sign language of their own such as BSL (British Sign Language) (Deuchar 2013) or ASL (American Sign Language) (Reagan 1995). Total communication skills (Swanwick 2020; Wood 1992) involving a combination of sign language, gestures, lip reading, finger spelling and partial speech skills with the help of hearing aids or implants is a very popular choice with the deaf.

Sign language and total communication skills do not allow a profoundly deaf person to speak like normal hearing people despite the deaf having a normal voice box, lungs and oral cavity. Inability to speak is a direct result of hearing loss. A normal hearing child hears an external sound and tries instinctively to produce it in speech. What really happens is that all the elements of sound (spectral and temporal features) are decoded by the inner ear (Ferraro 2002; Møller &Moore 2001). This decoded dataset travels to the brain via auditory nerves and is processed by the brain's auditory cortex and other related areas thus enabling the listener to understand its meaning, provided the sounds are in a familiar language (prior training and exposure). When the person tries to speak, the brain uses the same information that it has received by hearing and directs the muscles of the voice box, lungs and oral cavity to recreate the same sound that it has heard. In the deaf, when exposed to a sound, the nonfunctioning inner ear fails to

Peer review under responsibility of PLA General Hospital Department of Otolaryngology Head and Neck Surgery.

* Corresponding author Singh S, e-mail: Shomeshwar40@gmail.com

https://doi.org/10.26599/JOTO.2025.9540003

Received 2 February 2024; Revised 9 December 2024; Accepted 20 December 2024

decode the spectral and temporal information of the sound, resulting in the brain not receiving any sound-related information to process, resulting in a deaf person's inability to speak. Over many years, speech therapists have worked hard with deaf pupils using tactile and visual sensations (Keate et al. 1994; Nickerson et al. 1976) but in the absence of auditory feedback, attained very limited speech outcomes. (Seal 2021; Ling 1978; Dunn & Newton 1986)

Hypothesis

We proposed the hypothesis that in a deaf person in the absence of hearing, if the brain could receive the spectral and temporal features of spoken sound through an alternate visual route, then the brain could process this information and use it to develop speech.

There is ample evidence that one part of the brain can function in lieu of another, a phenomenon referred to as sensory substitution. (Bina 2020; Grady 1993; Yarkoni et al. 2010; Zeki 2004) This is clearly evident in blind people using tactile sensations to read and people without upper limbs using their lower limbs to perform tasks such as writing, holding a cup etc.(Striem-Amit et al. 2018). Sensory substitution has also been used in the deaf using a "vibratory vest" by which auditory information was captured, digitally processed, and delivered via skin using vibratory motors (Eagleman 2014). Like touch instead of vision is the obvious alternate sense for the blind, vision instead of hearing seems to be the obvious alternate sense for the deaf. The deaf naturally use visual feedback like gestures, lip reading, and hand & body signs for both receptive and expressive language resulting in a well-developed visual cortex at the cost of auditory cortex (Kral et al. 2001), that remains undeveloped (Seymour et al. 2017; Simon et al. 2020; Cardin et al. 2020; Bola et al. 2017).

One of the challenges in using a visual input instead of an auditory input is the nature of the signal and reaction times in the human brain (Shamma 2001). In comparison to auditory signals, visual signals have slower reaction times and longer processing pathways, resulting in much slower overall signal processing. This limitation in the speed of processing a visual signal limits the quantum of changing auditory signals that we can present to the eye in a visual format. Our own early experimentation representing sounds in a visual format revealed that the maximum duration of sound that we could comfortably present in a visual format was one second. Anything longer was difficult for visual processing. This meant that longer duration receptive speech was difficult to represent visually. But expressive speech in most languages like English could be depicted as a visual equivalent, since they consist of short phonetic elements combined to form a word (syllables and phonemes like /aa/, /ee/, /oo/, etc.). If the brain could learn how to speak a short phonetic element using visual feedback, it may be possible for the brain to learn to join them to make words.

One of the unique qualities of the brain is its ability to process complex signals and predict any missing blocks in a chain of signal. This phenomenon has been well established in the world of cochlear implants. In comparison to a normal inner ear with thousands of inner hair cells, a cochlear implant only has a maximum of 10 physiologically different channels to carry different frequency information from the inner ear to the brain (Clark 2012). Yet many cochlear-implantees can develop immaculate speech with training. Clear speech development in the cochlear-implantee confirms that the brain has been able to satisfactorily process the limited input signal. It is proposed that the brain does this by predicting the missing frequency information with experience and training (Lazard et al. 2012). We therefore suggest that even if the visual feedback is limited in its detail, the brain, with training, will be able to use it to develop speech.

Building the visual feedback sensory substitution solution

A normal hearing person learns to speak a new sound by matching the results of his own speech effort and the new sound he is learning to speak, until they both sound the same (the auditory feedback loop) (Rauschecker et al. 2008; Lee et al. 2007; Guerrette et al. 2018; Kujala & Näätänen 2010). Based on this, we set out to develop a sensory substitution solution that will provide both these feedbacks (one's own speech effort and a new sound) using computer hardware and software.

The algorithm was initially developed using Octave (an open-source software which is largely compatible with Matlab) and is used for scientific programming and numerical computations. The Octave implementation was then rewritten using C/C++ programming language and deployed on a laptop for initial testing and validation. This proof-of-concept implementation was then developed as a production quality Android smartphone application using a free software called Android Studio. The final application was implemented such that it works even on a basic low-cost smartphone.

We first set out to create the visual equivalent (VE) of any sound less than one second long that can be displayed on a smartphone screen. The main objective of the visual equivalent is for the brain to receive all the characteristics of the sound being practiced, namely the frequency content (spectral information) and its specific order (temporal information). Using computer software, we developed an algorithm that defined a sound by 13 Mel-cepstral coefficients (frequency-based information) (Lokesh & Devi 2019). A typical phoneme sound has a duration between 240 to 600 milliseconds. A two-dimensional array of cepstral coefficients and time was computed for a phoneme sound by using 20 millisecond of speech samples at a time. The x-axis shows the 13 cepstral coefficients, and the y-axis shows these values changing every 20 milliseconds. They were then normalized to have a value in the range of 0 to 63. A color map using 64 colors was then used for visualizing the cepstral coefficient array as a visual equivalent. (Figure 1) These color images show unique patterns for each phoneme sound. The VE contains all the spectral and temporal information that is meant to travel to the brain via the inner ears. Repeating the same sound again and again will generate the same VE. A different sound will have a different VE. A picture or script in contrast to our VE does not carry any spectral or temporal information needed for the brain to develop speech. As a result, a deaf person can use them to understand the meaning but cannot use the information to develop speech.

Next, we created a library of VEs (Reference VE) representing 30 phonetic elements from English spoken by normal men and women (example /aa/, /ee/, and /oo/) since English is a commonly used language in India (Table 1). Normal hearing children and adults were asked to repeat the 30 phonetic elements and the Visual Equivalent created was added to the library. The application uses this library of VEs to match the phoneme sound uttered by the user. We used 30 phonemes to start with since joining these 30 would allow a deaf person to start saying many useful words. In due course, we can develop many more Reference VEs as needed.

It should be possible to create the library of Reference VEs in many other languages in the future. Detailed studies will need to be carried out to explore as to which languages are amenable to creating Visual Equivalents. Languages in which spoken words can be broken down into syllables, should be more amenable in

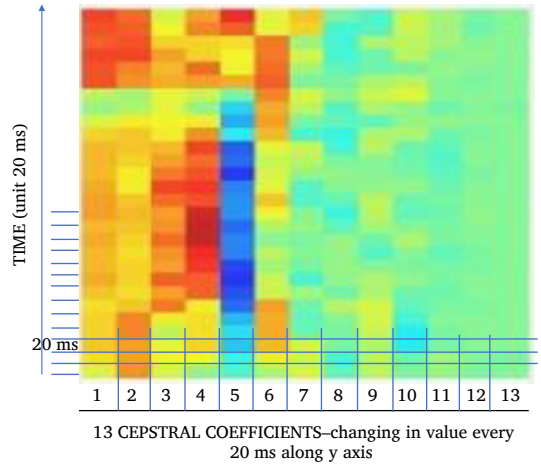

Figure 1 Visual Equivalent created from 13 Cepstral coefficient values of spoken sound changing every 20 ms.

Table 1 30 Phonetic elements in English (word in brackets defines the exact intended phonetic use of the phoneme).

a (again)	ar (farm)	ee (feet)	oo (boot)	o (pot)
p (pet)	m (my)	h (hop)	n (no)	b (bee)
k (key)	g (go)	d (do)	t (tea)	f (fun)
y (you)	r (red)	l (look)	s (soon)	ch (cheap)
sh (sheep)	z (zoo)	j (jump)	v (van)	th (thin)
dh (that)	ow (cow)	ay (payer)	oy (boy)	ie (pie)

comparison to others.

We chose to implement our solution on a smartphone due to its ubiquitous availability. The computational power available even in a basic smartphone was sufficient for this purpose. We created an Android based User interface that allowed a deaf person to see the VE of any sound spoken by them (Test VE) and compare it to the Reference VE. Figure 2 outlines the block diagram of the application. Note that the phoneme to be uttered by the user is already known to the application via the user interface. The microphone of the smartphone captures the input speech signal consisting of the phoneme sounds at a sampling frequency of 8000 Hz and converts it into 16-bit PCM (Pulse code modulation) samples. Each frame of 160 samples (20 milliseconds) is first processed to remove the background noise. The clean speech signal is then fed to a voice activity detector, which classifies the frame as active or silence. If the frame is marked as active, cepstral coefficients are computed. The time cepstrum is obtained by collating the cepstral coefficients of a burst of active frames. An objective score of the input phoneme sound is computed by comparing its time cepstrum with the reference phoneme time cepstrum database using the Dynamic Time Warping algorithm (Juang 1984). Finally, the user is presented with the following feedback: (a) VE of the phoneme sound uttered by the user, (b) VE of the reference phoneme sound for visual comparison, (c) Objective score in the form of green, orange, or red buttons with the color green being a "very good match," the color orange indicating a "reasonable match" and the color red indicating "needs more practice", and (d) Volume level of the uttered phoneme sound so that the user can calibrate the distance between their mouth and the microphone. Visual comparison allows the user to get strong feedback as to how well they spoke the sound (Figure 3).

We proposed that if a deaf child with practice could modulate his speech in such a way that the Test VE closely matches the respective Reference VE, then this exercise would empower the user to speak the sound depicted by the Reference VE.

2 Initial experience

Aim

To carry out a pilot observational study evaluating the possibility of profoundly deaf children and adults developing basic speech skills using visual feedback of their speech efforts.

Method

BIRAC (Biotechnology Industry Research Assistance Council, Government of India) funded and monitored the research project as well as its implementation in children in deaf schools including any ethical issues (BIRAC Expert committee approval 2017, 2019, BIRAC IIPME Grant 2017-19, BIRAC SBIRI Grant 2019-20). There were no ethical or regulatory issues reported since the solution was introduced as a smartphone speech and language learning application on standard Samsung phones, which are approved for use in public and by children. The research was carried out in accordance with the principles laid out by the World Medical Association Declaration of Helsinki. (WMA General Assembly 2008).

Subjects

The proposed solution was offered to profoundly deaf children and adults attending deaf schools who relied on gestures and sign language for communication. They did not attend any other classes on oral speech development. Children and adults with hearing aids or cochlear implants or with previous speech skills (ability to speak more than 5 sounds) were excluded from the study. Consent to use this application in school purely as an observational study, was taken from the students and their parents/legal guardians in the presence of schoolteachers for 12 months. Same was documented and filed in project records. Parents were made aware that this is an experimental observational study and only those who appreciated and liked our hypothesis,

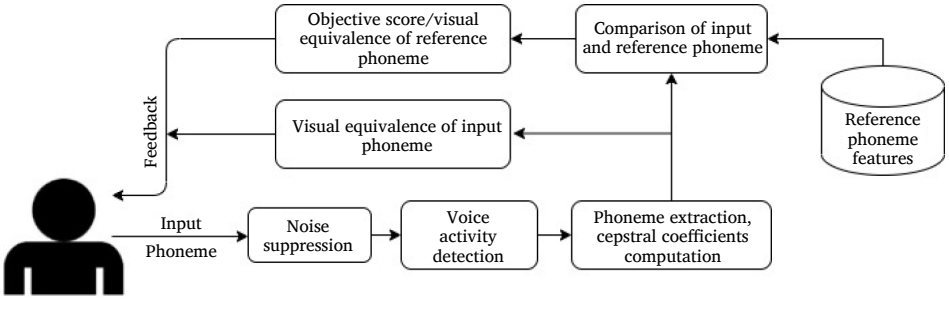

Figure 2 BLOCK DIAGRAM of the proposed assistive technology

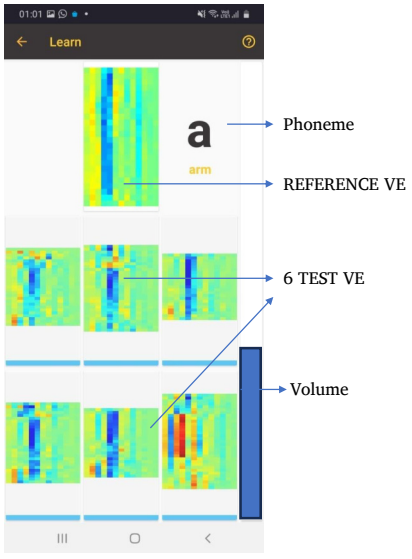

Figure 3 Reference and Test Visual Equivalents (VE).

agreed to participate. We did not exclude adults at this stage since they showed a lot of interest in the potential new technology, and we felt that their heightened visual processing skills will only be helpful in using visual feedback to develop speech.

Training

The trainer (teacher with experience of handling deaf children but not a professional speech therapist) visited deaf schools 3 to 4 times a week. During each visit, she conducted 1 to 3 training sessions with 4 to 5 children for about 45 minutes. No monetary incentive was offered given the deaf-school setting. At this point, it was not possible for us to enroll a control group, since the number of deaf students in these schools were limited and those students who were not interested in our project declined to participate in any follow-up activity, as they felt it to have no benefit for them. We accepted this weakness at this stage in our study, given the novelty of the technology. 30 phonetic sounds were introduced to the children one by one.

The training protocol, detailed below, was developed based on inputs from Teachers of the Deaf. The children's regular school schedule, interest and attention span was taken into consideration. The parents were also involved in finalizing the protocol making sure that the additional class was not too much burden on the children.

The students were encouraged to initially try and produce any sound they were comfortable with like /ma/ and /pa/ and see the Visual Equivalents on the smart phone screen. Once they seemed to understand that the Visual Equivalent was connected in some way to what they were speaking, they were then encouraged to change the sound and notice the change in the Visual Equivalent. The next step was to ask them to try and create a Visual Equivalent with their speech efforts that would match a Reference Visual Equivalent (representative of a specific phonetic element) shown to them on the smart phone. Their instruction was to attempt to speak the sounds in such a way that the Test Visual Equivalent of their spoken sound would visually match the Reference Visual Equivalent displayed on the smartphone screen as much as possible to their individual satisfaction. A phonetic element was considered learned if the child could satisfactorily demonstrate speaking it out to the trainer during the group session, subjectively evaluated by the Teacher for accuracy and clarity.

Every session comprised of time spent on revising what they had learnt in the last session and learning to speak new sounds. As such every session was training and follow up combined in one. The main mode of communication with all these children was a combination of gestures, informal signs and demonstrations. The Teacher was very familiar with this style of communication based on her experience. The order in which the phonetic elements were chosen to train was based on natural learning milestones of normal hearing children. /a/, /ar/, /ee/ in the beginning. /s/, /ch/, /sh/ later and /oy/, /ie/ towards the end. Given the school setup, assessment and monitoring were very subjective with the teacher qualitatively documenting when a child was able to speak a sound with reasonable clarity in her notes.

Data collection

We evaluated how many phonetic elements the child was comfortably reproducing on weekly visits. All the information was collected by the trainer in each session and collated in a database for analysis later. Sounds that were once declared learned were practiced in every session before moving on to more sounds.

Statistical Analysis

All statistical analyses were performed using R-software version 4.1.0 (R Core Team, 2022), a language and environment specifically designed for statistical computing. p-value less than 0.05, was considered statistically significant.

The study included a total of 72 participants, with a median age of 11.6 years. The interquartile range (IQR) of 9.5 to 17.2 years suggests that half of the participants were between these ages. The sample was predominantly male (69%, n = 50), with females constituting 31% (n = 22) of the participants (Table 2).

Table 2 Participant demographics (N = 72)

Characteristic		N	Median
Age		72	11.6* (9.5, 17.2)
Sex**	Female	72	22 (31%)
	Male		50 (69%)
Follow Up Duration (weeks)		72	7.0* (4.0, 12.3)

*Median (Inter Quartile Range = 75th percentile-25th percentile),

**n(%),

Follow up duration (weeks) = Total number of weeks that the children underwent training and follow up.

The follow-up duration (FUD) in weeks had a median value of 7.0 weeks, with an IQR from 4.0 to 12.3 weeks (Table 2).

3 Results

Improvement in the ability of the subjects to speak new phonetic elements after training with our solution was evaluated. Before training at the start of the study, the median score for phonetic elements spoken was 2.0, with an IQR (interquartile range) from 1.0 to 3.0. However, after training on the last day of training and follow up, the median score for spoken phonetic elements increased to 7.0, with an IQR from 6.0 to 11.0, indicating a substantial enhancement in the number of phonetic elements spoken by the participants after training. The Wilcoxon signed-rank test was performed on known phonetic elements before training and after training with our novel assistive technology. The p-value was extremely small (<0.0001), indicating that there is a statistically significant difference before and after training. (Table 3)

Table 3 Comparison of Phonemes spoken before and after training with our novel assistive technology based on sensory substitution.

	N	Before Training*	After Training*	p-Value
No. of phonemes known to the patients training	72	2.0 (1.0, 3.0)	7.0 (6.0, 11.0)	<0.0001

*Median (Inter Quartile Range) =75th percentile-25th percentile),
Wilcoxon signed-rank test used to compute p-value

The study utilized a plot to visualize the change in scores from the initial week (Week 0) for each subsequent visit, a method referred to as "Change from Baseline". This approach facilitated an understanding of the trend and an assessment of the effectiveness of the intervention. (Figure 4). The Wilcoxon signed-rank test with continuity correction was applied to the data collected over a span of 25 weeks. The test compared the median scores of each week to the baseline score (Week 0). The p-values were adjusted for multiple comparisons using the Bonferroni method (Bonferroni 1936). In the first week, there was no statistically significant difference from the baseline (p = 0.2067). However, starting from Week 2, a statistically significant increase in median scores was observed (p = 0.0030). This trend of statistically significant increases continued until Week 14. From Week 15 onwards, despite an increase in median scores, the adjusted p-values were greater than 0.05, indicating that these increases were not statistically significant when accounting for multiple comparisons. This trend continued until Week 20. From Week 21 to Week 25, although the median scores varied, all adjusted p-values were greater than 0.9999, indicating no statistically significant difference from the baseline. (Table 4)

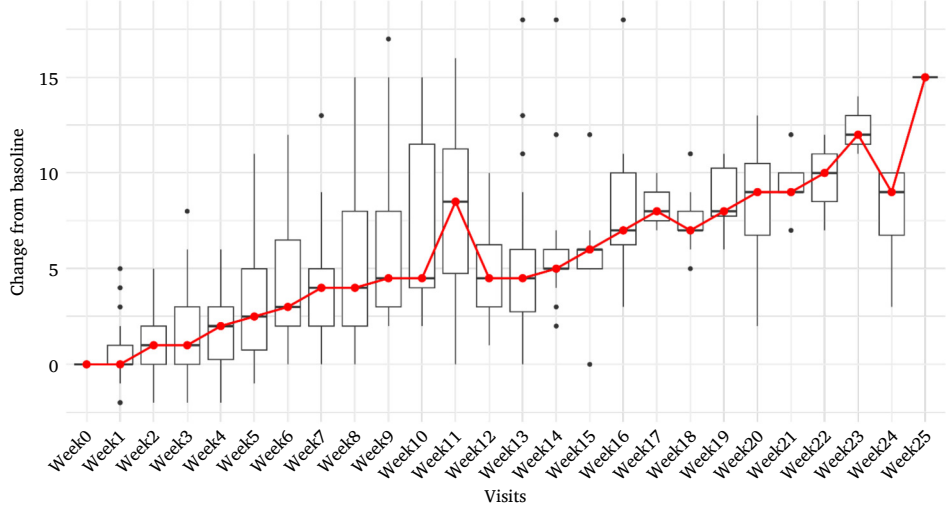

Figure 4 Boxplot of change in scores from baseline over time. The Box Plot indicates the change in the number of phonemes learnt every week. The Box represents the Inter Quartile Range (Q1 to Q3). Whiskers represent the minimum and maximum values. Black dots represent the Outliers. The horizontal black line inside the box indicates the median values. The connecting red line represents the median change from baseline for each week thus demonstrating the trend over time.

58 children stopped training and follow up by 14 weeks. 5 children continued to attend the sessions and stopped between 15 and 20 weeks. 9 children continued attending the sessions upto 21 to 25 weeks.

In conclusion, our results suggest that there were statistically significant increases in scores from Week 2 to Week 14. However, from Week 15 onwards, the increases were not statistically significant when accounting for multiple comparisons.

4 Discussion

The children and adults took great interest in using the application and in trying to speak. All children and adults in the school were approached and those who showed interest in trying a new technology participated in the study. There may be a slight bias in recruitment in favor of those who prefer modern technology but it's unlikely that will bias our results.

Surprisingly, there was no difficulty in getting them started with a new sound. This was achieved by encouraging them to speak any sound and then move on to speaking a specific sound guided by the Reference Visual Equivalent (VE). We encouraged them to focus on the VE and try to change it to match the Reference VE instinctively. Children very quickly connected their speech effort with the VE and demonstrated the ability to change the VE at will with remarkable ease. This display of interest and ability to modify the VE at will, supports our hypothesis that a visual equivalent can serve to replace auditory feedback in deaf children trying to speak. In other words, children can use vision instead of hearing to learn to speak. Simple sounds like /aa/ and simple meaningful ones like /ma/ and /pa/ were very easy for the child to reproduce guided by the real time visual feedback. Very early on, once the child started confidently speaking /pa/, the trainer got the child to repeat it twice in quick succession-/pa/ /pa/ creating the first instance of speaking a word-papa. Based on the results of this trial, we feel encouraged that visual feedback created using modern computing technology has promise in helping a profoundly deaf child or adult develop speech. This method also allows the child to practice speaking new sounds without assistance from a teacher for every practice session.

The duration of follow up in this study is relatively short since at the end of 24 weeks, the school shut down for holidays and we were unable to reconnect with the children and adults for many months. Also, the entire teaching exercise using our application was informal. As such many kids did not take our instructions

Table 4 Weekly analysis of Wilcoxon signed rank test results

Visit	Median (25th percentile-75th percentile)	** p-value	*Adjusted p-values
Week 0	2 (1 – 3)	Adjusted	Adjusted
Week 1	2 (1 – 3)	0.0083	0.2067
Week 2	3 (2 – 3)	0.0001	0.0030
Week 3	4 (2 – 5)	< 0.0001	0.0005
Week 4	4 (3 – 6)	< 0.0001	< 0.0001
Week 5	4 (3 – 6)	< 0.0001	0.0005
Week 6	6 (4.5 – 8)	< 0.0001	0.0010
Week 7	6 (5 – 7)	0.0005	0.0117
Week 8	7 (6 – 9)	< 0.0001	0.0005
Week 9	7 (6 – 9.25)	0.0005	0.0118
Week 10	7.5 (6 – 13.25)	0.0058	0.1440
Week 11	9.5 (6.5 – 12.25)	0.0039	0.0964
Week 12	7.5 (5.75 – 8.25)	0.0005	0.0118
Week 13	7.5 (4.75 – 8.5)	0.0007	0.0175
Week 14	9 (6 – 9)	0.0015	0.0369
Week 15	9 (8 – 9)	0.0133	0.3332
Week 16	10 (9.25 – 13.75)	0.0355	0.8881
Week 17	10 (9.5 – 11.5)	0.2500	> 0.9999
Week 18	10 (9 – 11)	0.0088	0.2212
Week 19	12 (9 – 13.25)	0.0137	0.3419
Week 20	12.5 (9.75 – 13.5)	0.0141	0.3537
Week 21	12 (10 – 13)	0.0579	> 0.9999
Week 22	12 (11.5 – 13.5)	0.2500	> 0.9999
Week 23	15 (15 – 16)	0.2500	> 0.9999
Week 24	11 (9.5 – 11.75)	0.0975	> 0.9999
Week 25	18 (18 – 18)	> 0.9999	> 0.9999

*Adjust p-values for multiple comparisons (Bonferroni method),

**Wilcoxon signed rank test with continuity correction p value

seriously and did not turn up for all the scheduled training classes as originally planned even though our trainer was present 3–4 times a week as originally planned. Interestingly though, in spite of this, many kids continued to show interest and progress in their ability to speak new phonetic elements. Overall, children learnt 1 to 2 new sounds every week. We did not perform any data imputation exercise for statistical analysis as that was not advisable with our incidence of missing data. Statistically significant progress as noted from 2nd to 14th week was in line with our expectations of progress made in speaking new sounds with time. Beyond 14 weeks, children probably started to take the exercise less seriously that caused the progress to not be statistically significant. Also high dropout rates after 14 weeks would have contributed to this trend. We were not able to get the children to practice speaking the sounds at home since they did not have access to mobile phones with our application at home and we did not want to burden the parents with this additional responsibility given the novelty of the proposal. The study strongly revealed that our current version needs significant improvements in the User Interface to retain interest of children and adults. Based on the current dataset, we will be working on this aspect of technology development. User interface also needs to be developed to help quantify learning. This would allow the deaf person to get a good understanding of how well he or she is doing and how much more time and effort may be needed to achieve a predefined level of speech.

The visual equivalent created for the current version was based on Cepstral coefficients. Speech can, however, be mathematically modelled using other characteristics such as spectral coefficients, continuous wavelet transformations, etc. More research is needed to establish the best method of creating Visual Equivalents. Development also needs to focus on using visual cues to emphasize where stress needs to be laid when pronouncing multiple phonetic elements making a word. This is possible by visually highlighting parts of some phonemes/syllables.

Late cochlear-implantees may potentially benefit from this technology. This is because these children have well developed visual cortices as needed in processing visual feedback. Combined with limited auditory inputs from cochlear implants, the brain should be able to use both visual and auditory inputs simultaneously to develop speech. Speech-to-text technology has emerged in the last few years as a robust option that the deaf can possibly use for receptive language understanding (Vijayakumar et al. 2020; Stolcke et al. 2006). Our proposed solution will add expressive language abilities. The same speech to text application can also potentially translate speech of the deaf people into text to give them live feedback of their own speech efforts. Text to speech applications are also developing for the deaf community. However, we feel that the deaf will feel more empowered if they develop their own speech, compared to speech produced by a phone or device.

Our technology cannot help children with dual sensory defects (deaf and blind). However, once the value in using an alternate sensory input is well established in developing speech, further research looking at tactile feedback may be helpful for the deaf blind.

5 Conclusion

The current pilot observational study suggests that sensory substitution has promise as a potential new technology that aims to improve speech outcomes in severe to profound deaf children and adults not timely treated by cochlear implants. Combined with speech-to-text transcription, once fully developed, this technology may become a very useful tool for the deaf and hard of hearing to scale new heights in speech clarity. Improvements in the algorithm creating a better Visual Equivalent of spoken sound combined with more effective User Interface and longer duration studies are needed to investigate the full potential of sensory substitution technology in speech development in the deaf.

6 Declaration

Singh S is the Founder & Promoter of 4S Medical Research P Ltd (New Delhi, India) the organisation that carried out the research.

Chachan R and Pandey P worked full time at 4S Medical Research P Ltd during the research project.

The Research was carried out with the help of Grants awarded by the Govt of India (details below). The grant money was used to carry out the entire project including paying the salaries of Chachan R and Pandey P, paying Nagarajan K (Couth IT) to develop the prototype algorithm and Vijayalakshmi P (SSN College of Engineering) for her time devoted to the project.

All authors contributed significantly to this work. Singh S conceptualised the core concept, designed the entire study and imple-

mented it from start to finish. Nagarajan K designed and developed the algorithm that displayed spoken sound into a Visual Equivalent and created the scoring strategy in the application. Chachan R and Pandey P carried out and supervised the training of users using the new application. They also developed the detailed curriculum that was followed in the training and assisted Singh S in implementation of the project at all stages. Vijayalakshmi P contributed to technology development that helped complete creation of the application. Verma N and Singh A gave valuable advise during the development of the concept and the prototype and in analysis of results. All authors discussed the results and implications and commented on the manuscript at all stages.

7 Grants

This research was supported by grants from Biotechnology Industry Research Assistance Council (BIRAC), Department of Biotechnology (DBT), Government of India. The funding organisation had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the decision to submit the article for publication; or in the preparation, review, or approval of the article.

Grant support

Birac IIPME Grant 2017-19 (Bt/IIPME 0203/02/16)
Birac SBIRI Grant 2019-20 (Bt/SBIRI 1725/38/18)

Acknowledgement

The authors would like to thank Nakra R for help with initial trials, Akshay Patil for help with data analysis and Sabherwal A for scientific inputs during the study.

References

Aimoni, C., Corazzi, V., Mazza, N., et al., 2018. Binaural, sequential or simultaneous cochlear implants in children: A review. In: Cochlear Implants: Advances, Efficacy and Future Directions. Courtney, H.W., Eds. New York: Nova Science Publishers.

Swanwick, R., 2020. 4 'Total Communication' – current policy and practice. In Issues in Deaf Education. London: David Fulton Publishers.

Bina, A., 2020. The link between Brain and Hearing system. *J. Otolaryngol. ENT Res.* 12(4), 114–117.

Bola, Ł., Zimmermann, M., Mostowski, P., et al., 2017. Task-specific reorganization of the auditory cortex in deaf humans. *Proc. Natl. Acad. Sci. USA* 114(4), E600–E609.

Bonferroni, C.E., 1936. Teoria Statistica Delle Classi e Calcolo Delle Probabilità. Seeber. Available at: https://books.google.co.in/books/about/Teoria_statistica_delle_classi_e_calcolo.html?id=3CY-HQAACAAJ&redir_esc=y. Accessed November 18, 2023.

Buchman, C.A., Gifford, R.H., Haynes, D.S., et al., 2020. Unilateral cochlear implants for severe, profound, or moderate sloping to profound bilateral sensorineural hearing loss: A systematic review and consensus statements. *JAMA Otolaryngol. Head Neck Surg.* 146(10), 942–953.

Cardin, V., Grin, K., Vinogradova, V., et al., 2020. Crossmodal reorganisation in deafness: Mechanisms for functional preservation and functional change. *Neurosci. Biobehav. Rev.* 113, 227–237.

Carlson, M.L., Sladen, D.P., Gurgel, R.K., et al., 2018. Survey of the American Neurotology Society on cochlear implantation: Part 1, candidacy assessment and expanding indications. *Otol. Neurotol.* 39(1), e12–e19.

Chu, C., Dettman, S., Choo, D., 2020. Early intervention intensity and language outcomes for children using cochlear implants. *Deafness Educ. Int.* 22(2), 156–174.

Clark, G., 2012. The multi-channel cochlear implant and the relief of severe-to-profound deafness. *Cochlear Implants Int.* 13(2), 69–85.

Deep, N.L., Dowling, E.M., Jethanamest, D., et al., 2019. Cochlear implantation: An overview. *J. Neurol. Surg. B Skull Base* 80(2), 169–177.

Deuchar, M., 2013. British Sign Language. London: Routledge.

Dunn, C., Newton, L., 1986. A comprehensive model for speech development in hearing-impaired children. *Top. Lang. Disord.* 6(3), 25–46.

Eagleman, D., 2014. Plenary talks: A vibrotactile sensory substitution device for the deaf and profoundly hearing impaired. In: Proceedings of the IEEE Haptics Symposium (HAPTICS), Houston, USA.

Ferraro, J.A., 2002. Hearing: Its physiology and pathophysiology. *Ear Hear.* 23(2), 166.

Grady, D., 1993. The vision thing: Mainly in the brain. *Discover* 14(6), 56–66.

Guerrette, M.C., Saint-Aubin, J., Richard, M., et al., 2018. Overt language production plays a key role in the Hebb repetition effect. *Mem. Cognit.* 46(8), 1389–1397.

Juang, B.H., 1984. On the hidden Markov model and dynamic time warping for speech recognition—a unified view. *AT&T Bell Lab. Tech. J.* 63(7), 1213–1243.

Keate, E., Javkin, H., Antonanzas-Barroso, N., et al., 1994. A system for teaching speech to profoundly deaf children using synthesized acoustic and articulatory patterns. In: Proceedings of the 1st Annual ACM Conference on Assistive Technologies, Marina Del Rey, USA, pp. 17–22.

Kral, A., Hartmann, R., Tillein, J., et al., 2001. Delayed maturation and sensitive periods in the auditory cortex. *Audiol. Neurootol.* 6(6), 346–362.

Kujala, T., Näätänen, R., 2010. The adaptive brain: A neurophysiological perspective. *Prog. Neurobiol.* 91(1), 55–67.

Lazard, D.S., Giraud, A.L., Gnansia, D., et al., 2012. Understanding the deafened brain: Implications for cochlear implant rehabilitation. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 129(2), 98–103.

Lee, H.L., Devlin, J.T., Shakeshaft, C., et al., 2007. Anatomical traces of vocabulary acquisition in the adolescent brain. *J. Neurosci.* 27(5), 1184–1189.

Ling, D., 1978. Speech development in hearing-impaired children. *J. Commun. Disord.* 11(2/3), 119–124.

Lokesh, S., Devi, M.R., 2019. Speech recognition system using enhanced mel frequency cepstral coefficient with windowing and framing method. *Cluster Comput.* 22(S5), 11669–11679.

Møller, A.R., Moore, B.C.J., 2001. Hearing: Its physiology and pathophysiology. *J. Acoust. Soc. Am.* 109(3), 849–850.

Monteiro, C.G., de Andrade Cordeiro, A.A., da Silva, H.J., et al., 2016. Children's language development after cochlear implantation: A literature review. *Codas* 28(3), 319–325.

Naples, J.G., Ruckenstein, M.J., 2020. Cochlear implant. *Otolaryngol. Clin. North Am.* 53(1), 87–102.

Nickerson, R.S., Kalikow, D.N., Stevens, K.N., 1976. Computer-aided speech training for the deaf. *J. Speech Hear. Disord.* 41(1), 120–132.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. Vienna: R Core Team.

Rauschecker, A.M., Pringle, A., Watkins, K.E., 2008. Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Hum. Brain Mapp.* 29(11), 1231–1242.

Reagan, T., 1995. A sociocultural understanding of deafness: American sign language and the culture of deaf people. *Int. J. Intercult. Relat.* 19(2), 239–251.

Rosen, R.S., 2010. American sign language curricula: A review. *Sign Lang. Stud.* 10(3), 348–381.

Seal, B.C., 2021. Speech development for children with impaired hearing. In: Introduction to Aural Rehabilitation: Serving Children and Adults with Hearing Loss. 3rd ed. Hull, R.H., Ed. San Diego: Plural Publishing, Inc.

Seymour, J.L., Low, K.A., Maclin, E.L., et al., 2017. Reorganization of neural systems mediating peripheral visual selective attention in the deaf: An optical imaging study. *Hear. Res.* 343, 162–175.

Shamma, S., 2001. On the role of space and time in auditory processing. *Trends Cogn. Sci.* 5(8), 340–348.

Sharma, S.D., Cushing, S.L., Papsin, B.C., et al., 2020. Hearing and speech benefits of cochlear implantation in children: A review of the literature. *Int. J. Pediatr. Otorhinolaryngol.* 133, 109984.

Simon, M., Lazzouni, L., Campbell, E., et al., 2020. Enhancement of visual biological motion recognition in early-deaf adults: Functional and behavioral correlates. *PLoS One* 15(8), e0236800.

Singh, S., Liasis, A., Rajput, K., et al., 2004. Event-related potentials in pediatric cochlear implant patients. *Ear Hear.* 25(6), 598–610.

Singh, S., Vashist, S., Ariyaratne, T.V., 2015. One-year experience with the Cochlear™ Paediatric Implanted Recipient Observational Study (Cochlear P-IROS) in New Delhi, India. *J. Otol.* 10(2), 57–65.

Stolcke, A., Chen, B., Franco, H., et al., 2006. Recent innovations in speech-to-text transcription at SRI-ICSI-UW. *IEEE Trans. Audio Speech Lang. Process.* 14(5), 1729–1744.

Striem-Amit, E., Vannuscorps, G., Caramazza, A., 2018. Plasticity based on compensatory effector use in the association but not primary sensorimotor cortex of people born without hands. *Proc. Natl. Acad. Sci. USA* 115(30), 7801–7806.

Thomas, E.S., Zwolan, T.A., 2019. Communication mode and speech and language outcomes of young cochlear implant recipients: A comparison of auditory-verbal, oral communication, and total communication. *Otol. Neurotol.* 40(10), e975–e983.

Vijayakumar, K.P., Nair, A., Tomar, N, 2020. Hand gesture to speech and text conversion device. *Int. J. Innovative Technol. Explor. Eng.* 9(5), 1241–1246.

WMA General Assembly, 2008. World Medical Association, Inc. Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects (as amended by the 59th WMA General Assembly, Seoul, October 2008). World Medical Association.

Wood, D., 1992. Total communication in the education of deaf children. *Dev. Med. Child Neurol.* 34(3), 266–269.

World Health Organization, 2018. Deafness prevention: Estimates. Geneva: WHO.

Yarkoni, T., Poldrack, R.A., Van Essen, D.C., et al., 2010. Cognitive neuroscience 2.0: Building a cumulative science of human brain function. *Trends Cogn. Sci.* 14(11), 489–496.

Zeki, S., 2004. The neurology of ambiguity. *Conscious. Cogn.* 13(1), 173–196.