Digitization Of Hindi Photo Articulation Test For Speech Sound Disorders

Chitralekha Bhat
TCS Innovation Labs
Mumbai, India
bhat.chitralekha@tcs.com

Apeksha Chaplot
Institute of Tech, Nirma Univ
Ahmedabad, India
12mict46@nirmauni.ac.in

Anjali Kant
AYJNIHH
Mumbai, India
speech1path@yahoo.com

Sunil Kumar Kopparapu
TCS Innovation Labs
Mumbai, India
sunilkumar.kopparapu@tcs.com

ABSTRACT

Interdisciplinary research on Speech Language Pathology is being pursued by several academic research groups, well spread over the globe. While the methodology for face-to-face speech therapy is well established, several mobile-device-based and web-based applications have emerged to assist either the speech language pathologist (SLP) or parents administering the therapy based on instructions of the SLP. Clearly technology is playing a major role in the form of assistive/augmentative technology. Although such tools and technology are widely in use in the West, there has been very little effort in this direction, for Indian languages. This paper describes the digitization of the well established Hindi Photo Articulation Test (H-PAT) , used for the assessment of articulation disorders, as a part of the pre-evaluation for Speech therapy. This paper will elaborate on H-PAT in its current form and show how digitization of H-PAT is envisioned to improve the quality of assessment from both the SLP's as well as the patients perspective. Digitized version affords flexibility to the user by means of personalization based on certain parameters. Also, this version accords SLPs better means of organization of work for future reference and monitoring progress.

Keywords

Hindi Photo Articulation Test (H-PAT), Digitization, User Centric Design

1. INTRODUCTION

The process of communication is a combination of the speech production process such as phonation, intonation, voice production, as well as language such as phonology, morphology, syntax, grammar, semantics, pragmatics [13]. A communication disorder manifests as inability in hearing, language and/or speech, ranging from mild to profound in severity, either developmental or acquired [1]. Any deviation from the normative during the process of communication, such as a speech disorder hampers the day-to-day activities and warrants Speech or Language Therapy. Speech Therapy is the process of identification and correction of the problems that manifest as unintelligible speech, carried out by trained Speech and Language Pathologists. Some of the common speech disorders are lisp, stuttering, speech sound disorders etc.[12]. Traditionally, speech therapy has been carried out face-to-face, however assessment of disordered speech is highly subjective. Maeir et. al. [8] highlight the impact of communication disorders on a country's economy and suggest that automatic speech processing techniques could reduce the cost of care for communication disorders. With this end object in mind, digitization of Hindi Photo Articulation Test (H-PAT) [4] was taken up. Objective of the digitization of H-PAT is multi-fold. (a) To reduce the efforts on part of the SLP while making the process more interesting and engaging to the patients through user centric thought process during the improvisation of the test. (b) To help the SLP better organize and maintain patient related information (c) To create a database of misarticulated speech in Hindi for future reference for both SLPs to monitor progress as well as for Speech research purposes. (d) As a stepping stone for automatic assessment of misarticulation during the pre-evaluation of speech therapy, hereby providing the SLP with cues for therapy. The rest of the paper is organized as follows: We describe the process of Speech Sound disorders and traditional approaches in Section 2 and describe the digitization of PAT in Section 3 and conclude in Section 4.

User Centered Design 2014, May 20-22

DOI 10.4108/icst.pervasivehealth.2014.255430

2. SPEECH DISORDERS AND TRADITIONAL APPROACHES

2.1 Speech Sound Disorders

Sounds in connected speech are produced by means of a sequence of articulator movements, requiring exact placement, sequencing, timing, direction and force of the articulators [2]. This sequence occurs simultaneously with precise airstream alteration, initiation or halting of phonation and velopharyngeal action in order to produce speech sounds. Assessment of speech sound production therefore requires considerable skill and knowledge. Speech sound disorders are generally classified as either motor-based or cognitive or

linguistic-based [6]. During the initial stages of the language acquisition process, children make mistakes such as omission or substitution of consonants, however a child is expected to produce sound correctly by a stipulated age range as shown in [3]. Articulation errors such as substitutions, omissions, distortions and/or additions are typically motor-based errors. An articulation disorder may be defined as difficulty in producing a single or a few sounds with no pattern or derivable rule [2]. The cause of misarticulation may be due to serious problems with muscle function such as cleft lip or palate known as organic speech disorder or the reason for this may be unknown, which is not related to anatomic malformations, called as a functional speech disorder. Example of an articulation disorder could be substitution of "w" sound for an "r" as in "wabbit" for "rabbit". Assessment of articulation identifies those sounds that the patient has difficulty producing. Therapy focus is on correcting individual error sounds, one by one. Phonological errors also considered Speech sound disorders, are more often cognitive/linguistic-based. The phonological process is concerned with organising the speech sounds into patterns of sound contrasts. Patients with phonological process deviations demonstrate difficulty in acquiring a phonological system, not necessarily in production of the sounds. Phonological errors can be grouped on some principle, forming patterns. These speech sound errors severely affect intelligibility.

2.2 Traditional Speech Assessment Process

The Speech therapy methodology currently being followed is outlined. During the first visit, a physical examination is conducted, post which the Hindi Photo Articulation Test (H-PAT) is administered, whenever misarticulations are heard during the informal assessment.H-PAT is a standardized test that was designed by a team of leading SLPs in 1988, funded by the UNICEF. The pictures were chosen through a country-wide survey in order to chose the most appropriate picture to represent a particular phone at a particular position. H-PAT in its current form is a book of photographs as shown in Figure 1 in black and white, accompanied by a set of sheets in which the patient details, the analysis of patient speech for each word is noted down by the SLP during the test.

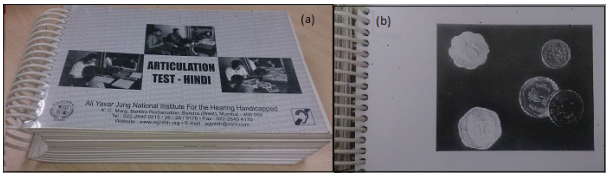

Figure 1: (a) The Book form of H-PAT (b)A picture in the H-PAT book

This test requires the patient to look at hundred and forty four pictures and speak the corresponding word. Figure 1 shows an example of a picture shown to the patient.

Each word which is part of the H-PAT word set, is picked with an aim to elicit a specific phone (sound) at a specific position such as word initial, middle and ending. For example the image in Figure 4 should elicit the response 'paise' in Hindi, with the phone sequence /p/, /E/, /s/, /e/, used to assess the consonant /p/ in the initial position. The assessment chart for the consonant /p/ is as shown in Figure 2.

Phoneme	Position	Word	C	S	O	D	A
/p/	Initial	पैसे					
	Middle	टोपी					
	Final	साँप					

Figure 2: Assessment Chart for phoneme /p/

Based on H-PAT, the kind of misarticulation is decided by the SLP and the observations are manually noted. A chart is prepared with CSODA - Correct, Substitution, Omission, Distortion, Addition for each phoneme. Percentage consonant correct (PCC) given as

$$PCC = \frac{\text{Number of consonants correctly produced}}{\text{Total Number of consonants}} \quad (1)$$

PCC is a metric used to gauge the severity of the articulation disorder. During the second visit a Pre-Therapy evaluation of the patient's speech is conducted and the SLP is involved informally. This assessment is directed towards investigating any perceptual disorders in the patient. Stimulability of articulation, through which a keyword wherein the patient can pick up the articulation is conducted. Therapy plan is made in which short term goals are planned for ten sessions and long term goals are set. Generally the 11th session is a progress report or a test. A bar chart for each goal is created for analysis at the end of the ten sessions. The entire process could involve administering a filtered version of the H-PAT as per the misarticulation of the patient and is currently completely manual.

3. DIGITIZATION OF HINDI PHOTO ARTICULATION TEST

Hindi Photo Articulation Test (H-PAT) is a work in progress, with the end objective of enabling users to administer the test at a time and place of their convenience by making it available on the internet. The sections to come elaborate on the features of the digitized H-PAT in comparison with the existing paper-based version.

3.1 User centric design

The two sets of users considered during the design of the digitized H-PAT, (a) the patients who are the subjects of H-PAT and (b) SLPs who administer H-PAT. Compilation of results, comparison with baseline, monitoring of progress becomes tedious using H-PAT in its current paper-based form, since this involves a huge amount of auditing and book-keeping.
The following are the key design considerations from the patient's perspective

- Based on the age group of the user, the images displayed are either animated versions of real objects or realistic pictures. This brings in the element of personalization based on age parameter as well as serves

to engage younger patients. Several other parameters such as education level, geographical location, mother tongue etc. are being considered.

- A counter that displays the number of pictures left gives the effect of gamification, providing incentive to the user for faster response.
- Audio corresponding to the word, can be played to provide auditory stimulation in case the user is not aware of the word to be spoken.
- The actual test can be shrunk to less than one hundred and forty four, by using the same picture for analysis of multiple phones, for example the word 'paise' in Hindi, with phone sequence /p/, /E/, /s/, /e/, can be used for word initial /p/, middle /E/ and /s/ and word final /e/. This will effectively bring down the number of pictures the patient needs to go through, in effect reducing the tediousness of the exercise.
- Since the original test was designed in 1988, some of pictures are difficult to identify for young patients, considering the current context. With improved usability in view, additional pictures were introduced to suit the current day context. Also, for the Indian scenario, wherein a majority of the urban population is multilingual and use both English and local language in daily conversation, sometimes the user is more familiar with the English word rather than the Hindi word corresponding to an image. Efforts are being made to identify such scenarios and enhance the test to handle them.

A screen shot of the baseline H-PAT is shown in Figure 3

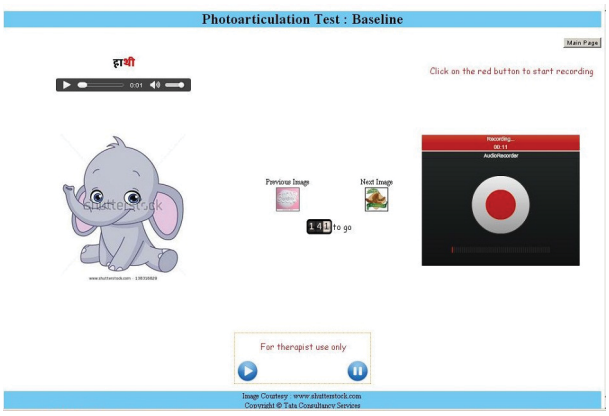

Figure 3: Baseline assessment for phone /I/ in the final position

Some of the above aspects such as (4) are applicable even from the SLPs perspective. Other key design considerations from SLPs perspective are enumerated below:

- Registration of patient with details such as age, mother tongue will help in analysis of patient speech post administration of H-PAT.

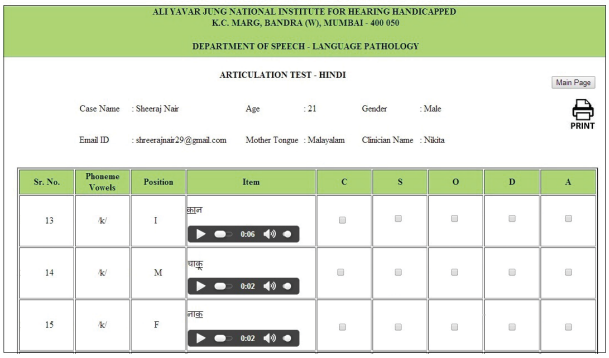

Figure 4: Screen shot of the report as seen by SLP

- A recording of the patient's speech will provide means to re-evaluate the speech based on patient's personal details if required.
- An already present auditory stimulation ensures consistency of pronunciation and will not be impacted by the subjectivity of SLPs speech.
- A digital report as shown in Figure 4 is generated in the format that the SLPs currently use. A facility is provided for the SLP to listen to the audio of patient speech for each word and evaluate in terms of CSODA. The report can be printed if required while a digital version is maintained for future reference during progress monitoring.
- Response time of the patient for each word is computed, this gives insights into the cognitive load on the patient and speech production post identification.
- Confidentiality of patient data is maintained since no external direct tagging of speech to patient is done, however the system internally does the book keeping and allows access to the SLP.
- A selective H-PAT can be conducted during the sessions that follow, wherein the SLP can select a few phones and positions of sounds, based on which a shortlisted H-PAT is created automatically. This will come in handy when the SLP wishes to evaluate the patient for corrections.

3.2 Speech data collection

For automatic assessment of disordered speech as discussed in [9], [7], [10], availability of speech data is extremely critical. The digitized Hindi Photo Articulation Test (H-PAT) serves as an excellent means by which misarticulation data can be collected to be used towards building systems for automatic assessment of speech as proposed in [11]. Such a database would also pave way for research into disordered speech in Indian languages. In addition to Hindi, with a little modification, the digitized H-PAT can be extended to other Indian languages as well.

3.3 Preliminary Evaluation

Enumerated below is the feedback received from the patients and SLPs as well as how concerns if any have been addressed.

- Patients found the digitized version of H-PAT engaging. The coloured images were clearer and the patients were able to identify the words better.
- The audio stimulus came in handy in some cases. Currently, audio stimulus is in the form of word spoken by a person with normal speech. However, the future versions could simulate the SLPs methodology of giving cues to the patient so that the patient can come up with the word on his own. This is envisioned as providing a game-like, more engaging experience to the user.
- SLPs felt that recording speech evidently would make the patients conscious about their speech and would hamper the natural process. This was addressed by recording in the background so that the patient is not conscious during the administration of H-PAT.
- The SLPs discouraged the written stimulus which was meant more for the SLP's reference rather the patients. However, it was found that older patients read the words instead of identifying the pictures, hampering the natural speech production of the patient. The written cues were kept hidden and would be enabled only if the SLP desires to do so.
- For children with Attention deficit disorder (ADD), the digitized version was more engaging whereas the child would not respond to the paper-based visual cues in a book form and the H-PAT had to be carried out only through auditory stimulation.

A video of a child responding to the digitized H-PAT as well as a canned demo that guides through the application, making the distinction between screens that would be used by the patient and the SLP, can be viewed at [5]. The digitized H-PAT in its current form is a work in progress. Several enhancements towards making H-PAT more comprehensive in terms of allowing the user to use it independently of the SLP at his convenience as well as automatically evaluating the patient speech are in the pipeline, as elaborated in the Conclusions and Future Work section.

4. CONCLUSIONS AND FUTURE WORK

Speech and hearing are critical to communication and socio-economic development of a human being. However, either due anatomical or functional/neurological reasons, there may be a communication breakdown, warranting speech therapy, that need to be administered by skilled SLPs. The focus of research is on automatically assessing disordered speech so as to increase the outreach of the SLPs, while not diluting the quality of service. Digitization of Hindi Photo Articulation Test (H-PAT) is envisioned as a step that would take us closer in achieving the above goal for Indian languages. Objective of digitization of H-PAT is multi-fold, providing a better user experience, shortening the time and reducing tediousness of taking the test, mechanism for book-keeping for the SLPs, recording of speech during H-PAT to serve as the critically needed data for automatic assessment of disordered speech. Personalization of the H-PAT based on certain parameters such as the age, education level, geographical location, mother tongue etc. are some of the few enhancements that would enhance the user experience. Also, automatic assessment of the patient's speech for CSODA and a detailed report based on parameters such as phone acquisition chart, keywords where the user can pick up certain articulation etc. would provide SLP with insights for planning the therapy sessions.

5. REFERENCES

[1] American speech-language-hearing association. `http://www.asha.org/policy/RP1993-00208/`. Viewed December 2013.

[2] Resource packet - assessment of speech: Sound production. `http://www.state.tn.us/education/speced/doc/71309SLIvoice.pdf`. Viewed December 2013.

[3] Speech & articulation development chart. `http://www.talkingchild.com/speechchart.html`. Viewed December 2013.

[4] Articulation test - hindi, Ali Yavar Jung National Institute for Hearing Handicapped, Mumbai, 1988.

[5] B. Chitralekha, A. Chaplot, and S. K. Kopparapu. `https://sites.google.com/site/awazyp/pat_hindi`. Viewed February 2014.

[6] P. F. John E. Bernthal, Nicholas W. Bankson. *Articulation and Phonological Disorders: Speech Sound Disorders in Children (7th Edition)*. Pearson, 2012.

[7] J. Kim, N. Kumar, A. Tsiartas, M. Li, and S. Narayanan. Intelligibility classification of pathological speech using fusion of multiple high level descriptors. In *INTERSPEECH*, 2012.

[8] A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth. {PEAKS} a system for the automatic evaluation of voice and speech disorders. *Speech Communication*, 51(5):425 – 437, 2009.

[9] A. Maier, T. Haderlein, F. Stelzle, E. Nöth, E. Nkenke, F. Rosanowski, A. Schützenberger, and M. Schuster. Automatic speech recognition systems for the evaluation of voice and speech disorders in head and neck cancer. *EURASIP J. Audio Speech Music Process.*, 2010:1:1–1:7, Jan. 2010.

[10] C. Middag, T. Bocklet, J.-P. Martens, and E. Nöth. Combining phonological and acoustic asr-free features for pathological speech intelligibility assessment. In *INTERSPEECH*, pages 3005–3008, 2011.

[11] A. Pande, S. K. Kopparapu, and V. Pandey. A mobile phone based speech therapist. In *Proceedings of SiMPE 2011, Joint Workshop with mobileHCI*, Stockholm, Sweden, Aug. 2011.

[12] Wikipedia. Speech disorder. `http://en.wikipedia.org/wiki/Speech_disorder`. Viewed December 2013.

[13] Wikipedia. Speech reading. `http://en.wikipedia.org/wiki/Speech_reading`. Viewed December 2013.