

Atherapy System for Articulation Disorder Correction

Supriya K S
Student, Dept. of ECE
GSSSIETW
Mysore, India

Sushma Bhagavathi B S
Student, Dept. of ECE
GSSSIETW
Mysore, India

Supriya Pal
Student, Dept. of ECE
GSSSIETW
Mysore, India

Vidya P G
Student, Dept. of ECE
GSSSIETW
Mysore, India

Mahantesh U
Assistant Professor, Dept. of ECE
GSSSIETW
Mysore, India

Abstract— There is a need for support system to provide better articulation instruction and practice in special education classes for children with articulation disorder. MATLAB, a multi-paradigm numerical computing environment which allows matrix manipulations, plotting of functions and data, implementation of algorithms etc.... is a tool used to achieve the aim and objectives of our project. The main aim of our project is to encourage students to achieve one of the primary milestone articulations of phonemes. To provide training in articulating phonemes by continual practice through MATLAB and to promote interactive learning method during therapy sessions are few prime objectives. This system would be very user friendly for therapist during the therapy session as they provide visual cues. These visual cues will encourage articulation disabled kid to learn in a fun way. Whole process was divided into three phases. The first phase was to obtain the reliable and efficient microphone based on characteristics. Where, the phonemes are recorded using different set of microphones; each microphone's frequency response characteristic was analyzed and appropriate microphone was selected for articulation therapy. Second, the data base (Reference) was created using the recordings from the suitable microphone that was articulated by the children (7-8 years) without any articulation disorder. Finally phonemes are recorded from articulation disordered children (7-9years) and compared the frequency response characteristics with that of the reference. On recording the speech sounds and comparing the characteristics with the help of MATLAB tool, suggestions will be given to correct the articulation through visual cues. Using those suggestions the therapist can guide the children with articulation disorder to correct their articulation.

Keywords—*Articulation disorder; Normatives; Visual cues;*

I. INTRODUCTION

Articulation (pronunciation and talking) is the ability to physically move the tongue, lips, teeth and jaw to produce sequences of speech sounds, which make up words and sentences that can be easily understood, which others could interpret in order to express basic needs. Articulation disorders involve difficulties in articulating specific types of sounds. It

often involves substitution of one sound for another, slurring of speech, or indistinct speech and it has become a major challenge in the 21st century. Articulation problems reduce speech intelligibility and communication. It also affects person's interpersonal communication, personality, social adaptive capability and learning ability. Hence Therapy sessions are necessary for articulation correction. In specific it is more important for the children in the age group of 7-9 years, as it is a best time for detection and correction of articulation disorder. The therapy sessions help children to become clear confident communicators so that they can become fully engaged in school and benefit from the curriculum, develop self help skills and independence for activities of daily living, actively participate in life experiences, and build healthy social relationships. In the sequence of therapy, language therapist subjectively utilizes clinical experience to individualized assessment, treatment, and training. But assessing and treating children with articulation disorder is a major challenge for therapists, as it becomes difficult to analyze and select the proper strategy for correction of articulation in certain situations in spite of clinical experience. Improper analysis of articulation disorder may lead to wrong therapy strategy usage for correction of articulation and unsuccessful results. Hence articulation assessment and training system that is therapist friendly (easy to handle and maintain) and which can support the therapists in analysis and correction of articulation (correction factors and suggestions given through visual cues which gives the variations in the appearance of a graphic display which are intended to assist the therapist in selecting the therapy strategy for articulation correction more efficiently) is the imperious demands of language therapist. Also there is a need for automatic speech processing techniques that can be used in a therapy system, which supports therapy sessions in the practice as well as tele-medical therapy sessions. Thus, using such a system articulation disorders can be treated easily.

II. RELATED WORK

Ambra Neri with her team propose a work called “Feedback in computer assisted pronunciation training: technology push or demand pull?” examined the type of feedback that currently available Computer Assisted Pronunciation Training (CAPT) systems provide, with a view to establishing whether this meets pedagogically sound requirements. It show that many commercial systems tend to prefer technological novelties that do not always comply with pedagogical criteria and that despite the limitations of today’s technology, it is possible to design CAPT systems that are more in line with learners’ needs [1].

Ikuyo Masuda-Katsuse in his work entitled “Support system for pronunciation instruction and practice in special education classes for language-disabled children” provides information about Support systems for pronunciation instruction and practice in special education classes for language-disabled children were developed. That encourages students to repeatedly practice the pronunciation they learned in their classes and promotes cooperation between teachers and outside experts who support the teachers. It helps the teachers making their work easy to improve their students’ pronunciation. It simplified not only articulation tests, but also the observation of the students’ pronunciation improvement processes [2].

R.Vijayalakshmi and S.Priya in the work called “An Interactive Speech Therapy Session using Linear Predictive Coding in Matlab and Arduino” proposed a system that aims at manipulating devices when the user input is correct and also indicates if the user input is incorrect. Speech recognition has been done using the concept of Linear predictive coding and Arduino Uno board is used for hardware interface [3].

Martin Russell with his team of researchers in work called “The STAR system: an interactive pronunciation tutor for young children” described the development and evaluation of a prototype interactive, computerized speech training aid which used phone-level HMM-based techniques from automatic speech recognition. The paper has also covered the development and evaluation of the underlying speech recognition technology and the prototype real-time system [4].

Heather Campbell and Tara McAllister Byun in their work entitled “Deriving individualized /r/ targets from the acoustics of children’s non-Rhotic vowels”, explains target selection by investigating the validity of individualized targets derived from children’s spelled vowels [5].

Khaled Tawfik in his work called “Towards The Development of Computer Aided Speech Therapy Tool in Arabic Language Using Artificial Intelligence” , explains the possible features suitable for the diagnosis phase and the corresponding classification accuracy. It describes the process of detecting Arabic speech articulation disorder using the dataset. Mel-frequency Cepstrum Coefficients using Matlab application [6].

III. METHODOLOGY

The proposed idea is to determine suitable microphone for recording different phonemes. Analyzing microphone characteristics and comparing the recorded phonemes from the

articulation disordered children (7-9 years) with that of the normal children (7-8 years). After acquisition and comparison of data, the necessary instructions are displayed with the help of visual cues, where in the therapists will analyze and guide the disordered children to correct their articulation.

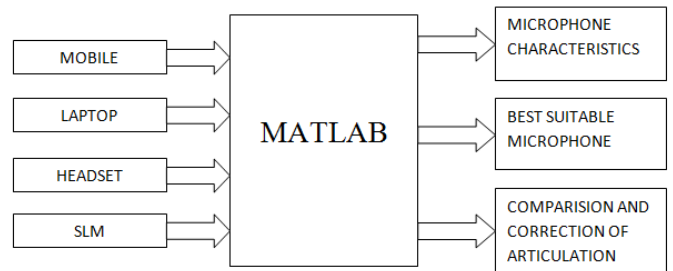


Fig. 1. Generalized block diagram.

Data acquisition is performed using various microphones (Ex: Mobile, Laptop, Headset, SLM: Sound Level Meter). PRAAT tool is used for the editing of speech samples and MATLAB is used for analysis and comparison of phonemes. The results are microphone characteristics (Frequency response), best suitable microphone for recording, comparison and correction of articulation.

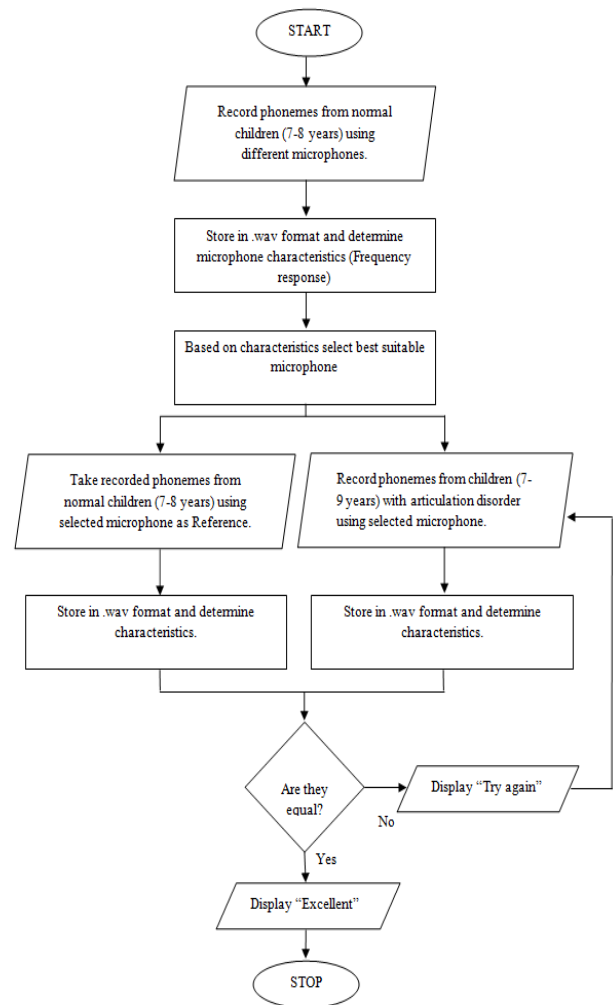


Fig. 2. Flowchart

Fig.2. Gives the flow diagram which describes the detailed procedure used to record analyze and interpret the comparative measures required for the articulation disorder therapy purpose and it is as follows:

A. Data Acquisition from normal children

Phonemes (26 Kannada phonemes) are recorded in Sound treated room with different microphones (Ex: Mobile, Laptop, Headset, SLM) for normal children of age group 7-8 years. These samples are stored in WAV format. The procedure followed for recording is as follows:

- Recording is performed in Sound Treated Room. This minimizes the effect of noise in the recording.
- Throughout the recording process we have used the same laptops, headset, SLM and mobile.
- The native Kannada speaking children were selected for recording and they are taken as normative which later will be considered as reference samples.
- They were instructed to articulate the phonemes for 10 times continuously, so that it will be helpful to pick one best sample.
- The distance between the microphone and the child is maintained same throughout the process of recording. This maintains the uniformity in the recording.
- Finally, the recordings were converted to WAV format.

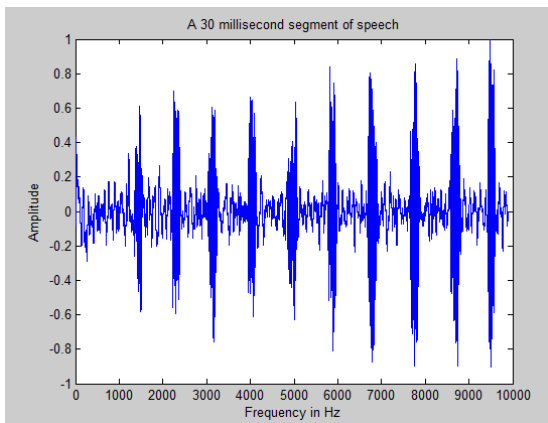


Fig. 3. Laptop (Phoneme “a”).

TABLE 1: List of Kannada phonemes

SLNo	PHONEMES	SLNo	PHONEMES	SLNo	PHONEMES	SLNo	PHONEMES
1.	A (ಉ)	8.	I (ಐ)	15.	O (ಓ)	22.	U (ಉ)
2.	Ba (ಬ)	9.	Ja (ಜ)	16.	Pa (ಪ)	23.	Va (ವ)
3.	Cha (ಚ)	10.	Ka (ಕ)	17.	Ra (ರ)	24.	Ya (ಯ)
4.	Da (ಡ)	11.	La (ಲ)	18.	Sha (ಶ)	25.	Au (ಔ)
5.	E (ಋ)	12.	Laa (ಳ)	19.	Sa (ಸ)	26.	Dha (ಢ)
6.	Ga (ಗ)	13.	Ma (ಮ)	20.	Ta (ಟ)		
7.	Ha (ಹ)	14.	Na (ನ)	21.	Tha (ಠ)		

Fig.3. Gives the waveform of the phoneme “a” recorded in laptop. Similar procedure is followed for rest of the phonemes

(26 Kannada phonemes) in other microphones. The phonemes are tabulated in TABLE 1.

B. Sample selection

One sample is selected out of 10 samples from the recorded phoneme using PRAAT tool. The procedure is as follows:

- Open PRAAT tool, select PRAAT objects, open, read from file, and browse the required audio file. This is shown in Fig. 4.
- Select view and edit option as shown in Fig. 5.
- Select the suitable sample, go to file option, save the selected sound as WAV file. This is shown in Fig. 6.
- Finally the sample will be as shown in the Fig. 7.

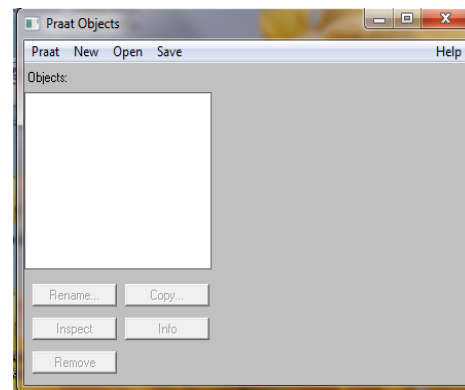


Fig. 4. Window 1.

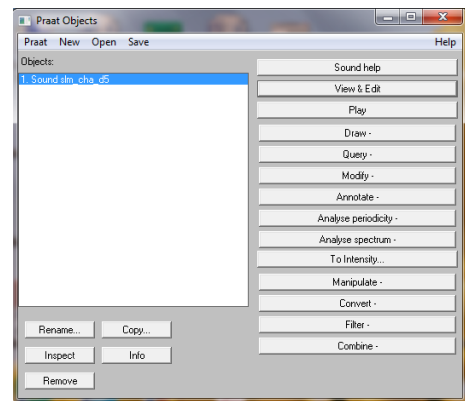


Fig. 5. Window 2.

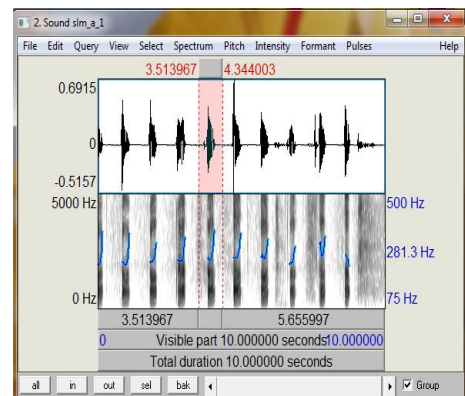


Fig. 6. Window 3.

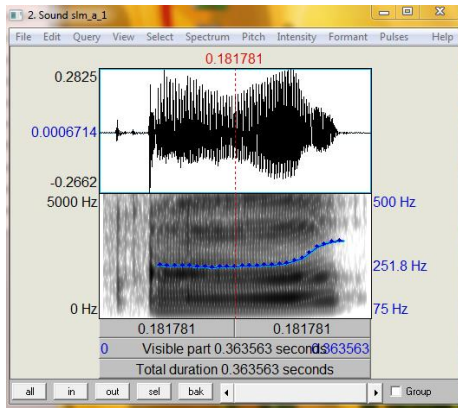


Fig. 7. Window 4.

C. Frequency response determination

The frequency response of the recordings from different microphones are analyzed and compared to obtain the best suitable microphone based on its characteristics (frequency response). The concept used in determining the frequency response is “Pitch detection via cepstral method.

In speech processing, pitch detection using the cepstral method is used to determine who is talking, for speaker separation, and for phase based speech reconstruction. Pitch detection is often done in the cepstral domain because the cepstral domain represents the frequency in the logarithmic magnitude spectrum of a signal. The Cepstrum is formed by taking the FFT (or IFFT) of log magnitude spectrum of a signal. The reason for using the FFT or IFFT interchangeably is because one will just give you a reversed version of the other, so each is equally valid for the processing we wish to do.

To detect pitch in frequency domain, Cepstrum method is a widely used algorithm. Cepstrum of a signal is obtained by taking the Inverse Fourier transform (IFT) of the logarithm of the spectrum of that signal [26]. Equation (1) gives the mathematical expression.

$$c[n] = F^{-1}\{\log(|F\{x[n]\}|)\} \tag{1}$$

Where, $x[n]$ is the sampled speech signal, F indicates its Fourier transform, and $c[n]$ are the Cepstrum coefficients. Cepstrum method reads a 40 ms window segment of the downsampled voice signal using 'audioread' function in MATLAB. Then the signal is multiplied by a hamming window. Fast Fourier Transformation (FFT) of this windowed frame gives the spectrum of the speech signal in frequency domain. Taking Inverse Fourier Transformation (IFT) of logarithm of the spectrum gives Cepstrum in quefrequency domain. Once in the quefrequency domain, the pitch can be estimated by determining the peak of the Cepstrum which represents pitch lag. The lag at which there is the most energy represents the dominant frequency in the log spectrum and thereby it gives the pitch frequency [23].

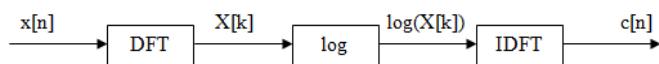


Fig. 8. Pitch detection using Cepstrum method.

A flow diagram of Cepstrum method for pitch detection is shown in Fig. 8.

Once in the cepstral domain, the pitch can be estimated by picking the peak of the resulting signal within a certain range. The Cepstrum is given in term of “quefrequency” which, besides being a terrible name, represents pitch lag. Therefore, the lag at which there is the most energy represents the dominant frequency in the log magnitude spectrum thereby giving you the pitch. There are of course some caveats to this approach. First of all, pitch and fundamental frequency are not actually the same thing, so depending on which peak your algorithm picks, you may be getting F0 (the fundamental) of F1 (one of the formants). Secondly, the Cepstrum is time shift variant. Therefore, you cannot just apply this method blindly. Instead, you need to precisely line up your time domain windows such that they start and stop exactly over a voiced speech segment. This is not a trivial task as most VADs often have errors and thus your Cepstrum will suffer from phase ambiguity. To get around this problem, we can use the differential Cepstrum and its variants such as the mean differential Cepstrum. This method is widely used and represents an important step in understanding the usefulness of this second Fourier domain.

D. Reference selection

Based on the frequency response best suitable microphone is selected. The microphone that is having the better frequency response and less noise is considered as the best suitable microphone for recording the phonemes. Then the phoneme recorded from the selected microphone is considered as reference. The reference is set separately for boy and girl.

E. Data Acquisition from articulation disordered children

From the selected microphone, phonemes (26 Kannada phonemes) are recorded from articulation disorder children (7-9 years). Due to early intervention delay the physical age of the normal kid usually don't match with that of hearing impaired and articulation disorder population. Hence the age range of the articulation disorder population for our project was considered to be 7-9 years. While recording the phonemes special care was taken to educate the care taker how the child need to cooperate before recording.

F. Comparison and correction of articulation

The frequency response of the phoneme articulated by the child with articulation disorder is compared with the reference. Based on the comparison results the instructions are provided using which the therapist can guide the child to correct the articulation if correction is necessary. The conditions and their respective output are: If the frequency response of phoneme articulated by child with articulation disorder is equal to the reference phoneme, then audio output: “Excellent” is given. If not equal, then one more condition is applied, which checks whether the frequency response of the articulation disordered child is within the range. If it is within the range then audio output: “Partially correct” is given else audio output: “Poor articulation, try again” is given.

IV. RESULTS

The proposed idea was to determining suitable microphone for recording different phonemes by analyzing microphone characteristics. Comparing the recorded speech

from children suffering from articulation disorder with the reference, then necessary instructions will be given to correct their articulation by therapist. As per the proposed idea the results that we obtained are as follows:

A. Microphone selection

Microphone Frequency response characteristic is determined for different microphones (Laptop, Headset, Mobile, SLM). Then results of each are compared with each other. The microphone that is having the better frequency response and less noise is considered as the best suitable microphone for recording the phonemes. The best microphone that we have considered is SLM. It is because of its better frequency response, good sound quality and less noise. One can also choose headset for their work as it is cost effective compared to SLM. This is depicted in the Fig. 9. Where, the plots in the first column give the input speech signal (phoneme) of different microphones. The second column gives the log magnitude spectrum of the corresponding plots in the first column. Then the third column gives their corresponding pitch frequency.

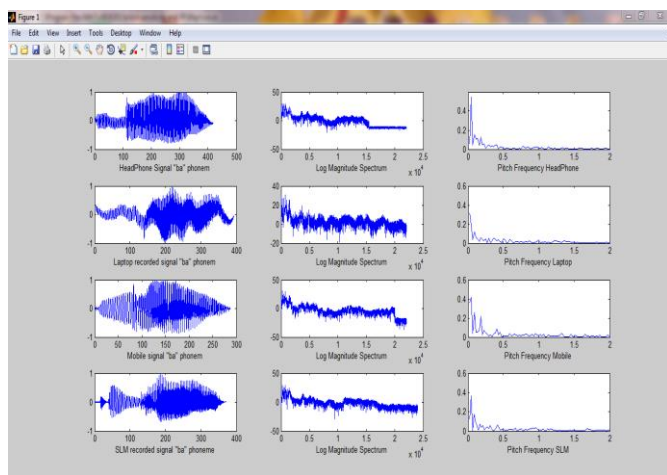


Fig. 9. Best suitable Microphone justification.

B. Comparison and correction of articulation

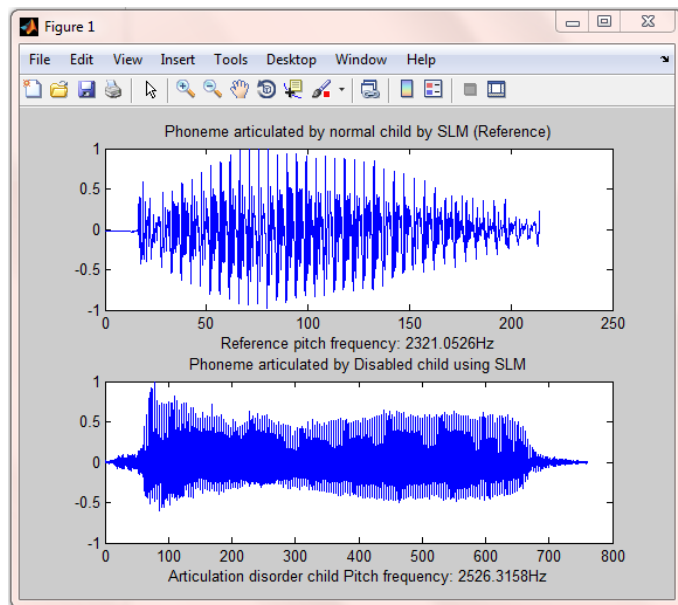


Fig.10. Comparison output

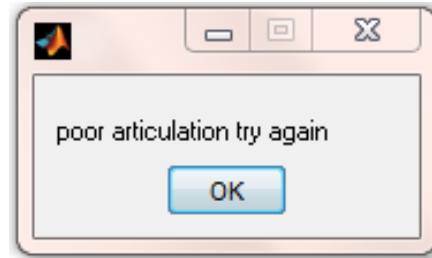


Fig.11. Instruction

Finally the comparison output will be provided that gives the spectrum plot and its corresponding pitch frequency. This is depicted in the Fig. 10. Also the instruction will be provided based on the comparison output. For example, if the pitch frequency of the phoneme articulated by the child with articulation disorder is less than the reference. Then the message box will be displayed which says “Poor articulation Try again” along with the speech output saying the same. This is as shown in the Fig. 11.

REFERENCES

- [1] Ambra Neri, Catia Cucchiarini and Helmer Strik, “Feedback in computer assisted pronunciation training: technology push or demand pull?”, A2RT, Dept. of Language and Speech, University of Nijmegen, The Netherlands {A.Neri,C.Cucchiarini,H.Strik}.
- [2] Ikuyo Masuda-Katsuse ,“Support system for pronunciation instruction and practice in special education classes for language-disabled children”, Speech Communication: Paper 5pSCb37, 172nd Meeting of the Acoustical Society of America, Honolulu, Hawaii 28 November - 2 December 2016.
- [3] R.Vijayalakshmi and S.Priya, “An Interactive Speech Therapy Session using Linear Predictive Coding in Matlab and Arduino”, 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), ISBN No.978-1-4673-9545-8.
- [4] Martin Russell, Robert W. Series, Julie L.Wallace, Catherine Brown and Adrian Skilling, “the star system: an interactive pronunciation tutor for young children” computer speech and Language (2000) 14, 161–175 Article No. 10.1006/csla.2000.0139.
- [5] Heather Campbell & Tara McAllister Byun, “Deriving individualised /r/ targets from the acoustics of children’s non-rhotic vowels”, Clinical Linguistics & Phonetics, DOI:10.1080/02699206.2017.1330898.
- [6] Khaled Tawfik, “Towards The Development of Computer Aided Speech Therapy Tool in Arabic Language Using Artificial Intelligence”, Department of Computing & Information Systems, Cardiff School of Management, Cardiff Metropolitan University, April 2016.
- [7] Hung-Yu Su, Chun-Hsien Wu and Pei-Jen Tsai, “Automatic Assessment of articulation disorders using Confident Unit-Based Model Adaptation”,1-4244-1484-9/08/\$25.00©2008 IEEE, ICASSP 2008.
- [8] Yeou-Jiunn Chen, Jing- Wei Huang, Hui-Mei Yang, Yi-Hui Lin and Jiunn-Liang Wu, “Development of Articulation Assessment and training system with Speech Recognition and Articulation training strategies selection”, 1-4244-0728-1/07/\$20.00 ©2007 IEEE, ICASSP 2007.
- [9] Yeou-Jiunn Chen, Chung-Hsien Wu, Jiunn-Liang Wu, Hui-Mei Yang, Chih-Chang Chen and Shan-Shan Ju, “An Articulation Training System with Intelligent Interface and Multimode Feedbacks to Articulation Disorders”, 2009 International Conference on Asian Languages Processing, 978-0-7695-3904-1/09 \$26.00 © 2009 IEEE DOI 10.1109/IALP.2009.10.
- [10] Zhou X, Boyce SE, et al, “A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/ ”, J AcoustSoc Am. 2008;23:4466–4481. Y. E-WC.
- [11] Jerad Lewis, “Understanding Microphone Sensitivity”, Analog Dialogue 46-05 Back Burner, May (2012).

- [12] Lawrence R. Rabiner, Ronald W. Schafer, "MATLAB exercises in support of teaching digital speech processing", 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP).
- [13] Arnab Pramanik, Rajorshee Raha, "Automatic Speech Recognition using Correlation Analysis", 978-4673-4805-8/12/\$31.00©2012 IEEE.
- [14] Qiang HE, Youwei ZHANG, "A Speech Recognition and Speech Corpus System Based on Matlab", Proceedings of 2001 International Symposium on Intelligent Multimedia, video and Speech Processing, May 24 (2001), Hong Kong.
- [15] Herman Orgeron, "Method for Extracting the Frequency Response of an Audio System from a Recording", Journal for Undergraduate Research in Physics, 7 September 2011.
- [16] Manjula G and Shiva Kumar M, "Development of a Recording Protocol for the Assessment of Speech Disorders", 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICECCOT), 978-1-5090-4697-3/16/\$31.00 ©2016 IEEE.
- [17] Culton GL in his work called "Speech Disorders among College Freshmen: A 13-Year Survey", Journal of Speech and Hearing Disorders.
- [18] Ladefoged, Peter; Maddieson, Ian (1996). The Sounds of the World's Languages. Oxford: Blackwell. ISBN 0-631-19814-8.
- [19] Weston, A., & Irwin, J. (1971), "Use of paired-stimuli in modification of articulation", Perceptual and Motor Skills, 32, 947-957.
- [20] A.Czyzewski, B.Kostek and H.Skarzynski, "Diagnostic system for speech articulation and speech understanding", Institute of physiology and pathology, 2006.
- [21] "Microphone techniques for recording", A SHURE Educational Publication.
- [22] Clinton mcreery, "Studio Recording Techniques Clinton Mcreery".
- [23] Muhammad Navid Anjum Aadit, Sharadindu Gopal Kirtania and Mehnaz Tabassum Mahin, "Pitch and Formant Estimation of Bangla Speech Signal Using Autocorrelation, Cepstrum and LPC Algorithm", 19th International Conference on Computer and Information Technology, December 18-20, 2016, North South University, Dhaka, Bangladesh.
- [24] Bartek Plichta, "Best practices in the acquisition, processing, and analysis of acoustic speech signals", article 16, Volume 8, Issue 3 Selected Papers from NWAV 30, University of Pennsylvania Working Papers in Linguistics.
- [25] Dagmawi Mallie, "Voice processing using MATLAB as a tool", Technology and communication, Vaasan ammatikorkeakoulu VAMK, University of Applied Sciences, 2014.