SPEECH THERAPY SYSTEM FOR CHILDREN WITH CLEFT LIP AND/OR PALATE

Aneeta Sabu, Deepthi G.A.*, Sowmiya V., V.S. Devika, Lani Rachel Mathew, Amritha B.J.
Department of Electronics and Communication,
Mar Baselios College of Engineering and Technology,
Trivandrum, India
deepthianilkumar26@gmail.com

ABSTRACT

In India, Cleft Lip/Palate is a common congenital craniofacial deformity. These are openings in the lip and (or) palate. Various charities, such as Smile Train and Transforming Faces provide free cleft lip and palate surgeries. The child requires constant monitoring and therapy by speech therapists, after surgery. For very small children, therapy can be made more effective with the use of technology developed in close association with speech therapists. Often, multiple surgeries are required according to the age and growth of the child, as determined by the doctors. The therapy offered after each stage of the cleft treatment is different. The proposed speech therapy system is an affordable therapeutic system to provide training for correct articulation of words. The technique uses Support Vector Machine (SVM) to recognize speech based on features using Mel-Frequency Cepstral Coefficients (MFCC). This project aims at enhancing the learning experience of speech production for patients who have undergone cleft lip or palatal surgeries by developing a cleft suite which adopts the Supervised Learning algorithm. A GUI has been developed for selected syllables which aids the therapists/parents to train the affected patient to articulate words in the correct manner. The system will help the young patient to correctly articulate the words through practice. This will greatly increase the effectiveness of the surgery done for the patients, by easing the burden on the therapist and capturing the interest of the young patients (4 - 7 years of age).

KEYWORDS

Speech Therapy, Cleft Lip/Palate, Supervised Learning, cleft suite, Mel-Frequency Cepstral Coefficients (MFCC), Support Vector Machine (SVM)

1. INTRODUCTION

Cleft Lip/Palate (CLP) is a common craniofacial abnormality that is present at birth. It is a small gap or slit in the upper lip that may or may not reach till the roof of the mouth (palate). This happens as a result of lack of sufficient tissues in the lip or(and) mouth area which is required for complete development of the face. Cleft lip with or without palate (CL/P) and cleft palate alone (CPO) are two types of this anomaly. Clefts are thought to be caused by a mixture of genetic (inherited) and external causes like consumption of certain drugs by the mother during her pregnancy, deficiency of folic acid or usage/exposure to tobacco or alcohol. If a sibling, parent, or any other family member has had a cleft, the chances of a newborn developing one seems to be higher [1]. After birth, these children face various challenges, which depend on the kind and extremity of their cleft, including difficulties in feeding, hearing, dental issues and speech related or communication issues. Due to a number of medical issues related with cleft lip/palate, a team of experts are required for their treatment which includes a plastic surgeon, dentist, audiologist, paediatrician, ear-nose-throat (ENT) surgeon and speech therapist.

2. RELATED WORK

Changes in velopharyngeal function using orthodontic or dentofacial orthopedic maxillary mask used in face therapy for patients having cleft palate has been investigated in [2]. People with Pittsburgh weighted speech scores (PWSS) less than zero were stable after face mask therapy. Non-parametric and parametric methods to study the relation between velopharyngeal insufficiency and face mask therapy have been investigated. The use of a nasal articulometer to gather speech data or information from cleft lip/palate patients and the normal orator is described in [3]. The naso-articulo meter contains a nasal microphone, oral microphone, tablet for patients, tablet for speech language therapists (SLP) and a server. The nasal microphone is connected with the tablet for SLP and the oral microphone in this system is connected to a tablet dedicated to the patient, which is connected by bluetooth with tablet for SLP. A speech recognition module using Raspberry Pi3 is described in [4]. The implementation of 3D speech mentor and ultrasound in the medication of speech sound disorders is described in [5]. Ultrasound tongue imaging technology is utilized for retrieving and dealing with cleft lip and palate. Studies in South Africa in which patient records were recorded and data from the note down is transferred to Microsoft Access are described in [6]. Ethnicity, gender, cleft lip site, and palate, the past of cleft lip in the family were recorded.

3. SYSTEM DESIGN

The proposed system consists of a speech recognition module and a visual feedback system. The speech recognition module allows the recognition of sound and translation of the input language into text, with the help of machine learning. It is followed by a visual feedback system for the benefit of the children.

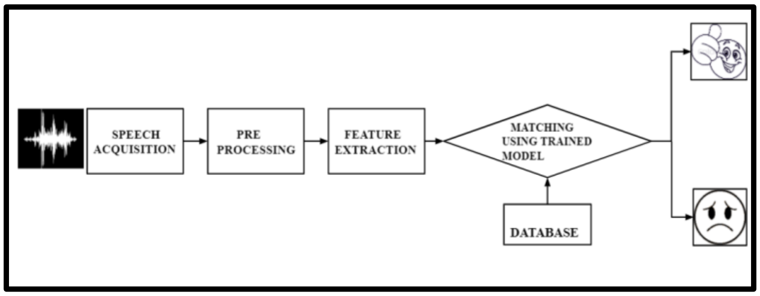

Fig. 1: Block diagram of speech recognition module

The speech recognition module consists of speech acquisition, pre-processing, feature extraction, database, matching with trained models and visual feedback system as shown in Fig. 1. Speech acquisition is performed by using a microphone for recording the sound. Pre-processing is the first phase in speech recognition, to distinguish the voiced and unvoiced signal and generate a vector numerically. Feature extraction will change the speech waveform to a form of parametric representation. MFCC extraction is the initial step in feature extraction, to identify the components of the audio signal that are important for the detection of linguistic information. The cepstral coefficient of mel frequency is an efficient technique used to collect and process signals as proposed by Davis and Mermelstein (1980). After extracting the MFCCs, the input speech signal is compared with the training model and similarity checking is performed. Training of data is done using the support vector machine (SVM) model. SVM is a popular Supervised Learning algorithm, that helps in classification and regression problems. Training is done with the help of a database consisting of normal speech samples created for selected phonemes.

4. METHODOLOGY

The methodology involved in the development of the system is described in this section.

4.1 Data Set Collection

The dataset used for training were collected from children of the age group between 4 years and 7 years. A few samples were collected from adults of age group 18-35 years. The dataset consists of samples of both boys and girls of the age group who did not have any cleft deformity. Informed assent was collected from the guardians of the children, and the full recording was conducted in their presence. Around 3000 samples were collected, and the samples included six syllables *pa, ba, tha, ka, dha* and *gha*. These are a few sounds produced from the anterior and posterior part of the vocal tract which need to be checked mandatorily. 500 samples of each syllable were collected. The syllables were recorded using a voice recorder app in a smartphone and the recorded voice samples were converted to a WAV file using Praat software.

4.2 Feature Extraction

MFCC (Mel frequency Cepstral Coefficient) based feature extraction is used for extracting the features of each sample collected. On successful running of the feature extraction code, feat.npy and label.npy files are created.

4.3 Model Training

Two models were created, one model was trained only with the voice samples collected from children and the second model was trained with voice samples of both adults and children.

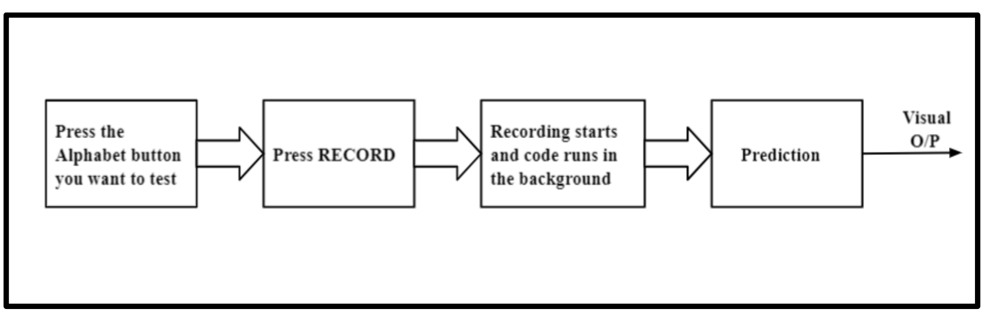

Fig. 3: Block diagram for the GUI development

A graphical user interface (GUI), developed using PyQt4, was used to train the patient to articulate words in the correct manner, depending on the results of similarity check. The block diagram is shown in Fig. 3. The developed GUI consisted of a screen which contained the buttons for each syllable used in the therapy. Syllables detected by the system, were displayed as the output in the command prompt. The button corresponding to the syllable which is to be recorded at a time is selected. The syllable is then recorded in real time by clicking the record button. Once the syllable is recorded, it is matched with the dataset and if the recorded sound matches with the syllable which was selected, a grinning face with a thumbs up emoji appears on the screen, as seen in Fig. 4(a). If not, a crying emoji appears as shown in Fig. 4(b).

Fig. 4 (a): Matched (thumbs up emoji)

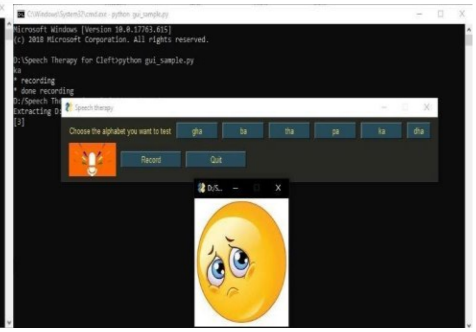

Fig.4(b) : Not matching (crying emoji)

5. RESULTS

The system was tested with different samples from children of age group 3 years to 7 years as well as with samples of adults of age group 18 years to 35 years. The number of times the syllable was correctly articulated by the system was observed. Participants were given IDs S1 to S8. IDs S1 to S3 are children in the age group 3 to 7 years. S4 to S8 are adults with normal voice.

Case 1: Model trained with samples of children only was tested with samples of children having defective speech as well as normal speech. It is observed from Table 1 that the system detected only 4 out of the 6 syllables correctly.

Case 2: Model trained with samples of both children and adults were tested with samples of children having defective speech as well as normal speech and from adults. It was observed that the system detected five out of the six syllables correctly and showed better performance compared to the model trained using samples of children only, as seen in Table 2.

The efficiency of the system when tested with the model trained only with the voice samples of children was found to be 66.6% only. The efficiency of the system when tested with the model trained with voice samples of both children and adults was 80.8%.

Table 1: Model trained with children only

Syllable	S 1	S 2	S 3	No: of Hits	Accuracy (%)
Pa	T	T	T	3	100
Ba	T	T	T	3	100
Tha	F	F	F	0	0
Ka	T	T	T	3	100
Gha	T	T	T	3	100
Dha	F	F	F	0	0
				Average	**66.60%**

Table 2: Model trained with children and adults

Syllable	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	No: of Hits	Accuracy (%)
Pa	T	T	T	T	T	T	T	T	8	100
Ba	T	T	T	T	T	T	T	T	8	100
Tha	T	T	F	T	T	T	T	F	6	75
Ka	T	T	T	T	T	T	T	T	8	100
Gha	T	T	T	T	T	T	T	T	8	100
Dha	F	F	F	T	F	F	F	F	1	10
									Average	**80.80%**

6. CONCLUSION

The work aimed at developing a tool that can be used in the speech therapy session of children with cleft lip and/or palate. The system used a machine learning algorithm to detect the syllables produced by the child and predict if he/she articulated correctly. MFCC based feature extraction was used for extracting the features of the samples collected. An advanced machine learning model, SVM (support vector machine) is used for training, to detect the syllable produced by the child and predict if it is correct or not with a good accuracy rate. A graphical user interface was also developed to make the system more user friendly. It enables a good interaction and a seamless user journey during the therapy session of children. Emojis are used to provide visual feedback. If the child articulates the word correctly, then a happy face with a thumbs up will appear and if the syllable articulated by the child is not correct then a crying face will appear. To develop a speech recognition system with a high accuracy rate, huge amount of sample data is required. As the number of data samples used for training increases, the efficiency of model training also increases and the system will detect the sounds more effectively.

7. ACKNOWLEDGMENT

The authors acknowledge the valuable suggestions from Manju S., Department of Audiology and Speech-Language Pathology, National Institute of Speech and Hearing, Thiruvananthapuram.

8. REFERENCES

1. Williams, C., Harding S., Wren, "An Exploratory study of speech and language therapy intervention for children born with cleft palate and lip" Cleft palate–Craniofacial Journal. doi.org/10.1177/1055665620954734, 2020
2. Narmin Helal, Matthew Ford, Osama Basri, Lindsay Schuster, Brian Martin, Joseph Losee."Relationship of velopharyngeal insufficiency with face mask therapy in patients with cleft lip and palate" -The Cleft Palate Craniofacial Journal 57(1),118-122, 2020
3. C.Wutiwiwatchai, P. Chootrakool, S. Kasuriya, K. Makarabhirom, N. Ooppanasak and B.Prathanee, "Naso-Articulometry Speech Database For Cleft-Palate Speech Assessment"-*Oriental COCOSDA- International Conference on Speech Database and Assessments,*Miyazaki, Japan, pp. 32-36, doi: 10.1109/ICSDA.2018.8693008, 2018
4. Hasan Gyulyustan, Svetoslav Enkov, **"**Experimental speech recognition system based on raspberry pi 3", *IOSR journal of computer engineering* 19(3):107-112, 2017
5. Z.Roxburgh, J.M. Scobbie & J. Cleland, "Articulation therapy for children with cleft palate using visual articulatory models and ultrasound biofeedback" , *Proceedings of the 18th ICPhS*, Glasgow , 2015
6. Prashantha Moodley**,** Mosimane Peter Silas Sethusa**,** Mohamed Imran Khan ,"The distribution of orofacial clefts at the medunsa oral health centre", *Department of Orthodontics, Sefako Makgatho Health Sciences University*, Pretoria, South Africa, 2018
7. https://www.mayoclinic.org/diseases-conditions/cleft-palate/symptoms-causes/syc-20370985#:~:text=A%20cleft%20lip%20is%20an,the%20mouth%20(cleft%20palate)
8. https://www.webmd.com/oral-health/cleft-lip-cleft-palat
9. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm
10. https://www.speechbuddy.com/blog/speech-therapy-techniques/top-4-cleft-palate-speech-therapy-techniques/
11. https://kidshealth.org/en/parents/speech therapy.html
12. https://www.assignmentpoint.com/science/medical/cleft-lip-and-cleft-palate-cause-symptoms-and-complications.html