Automatic Speech Therapy For Kids Using Frequency Comparison

Joswin Valerian Concessao*[1], Deeksha Kundabarandady*[2], J Arpana*[3], Nidhi Prakash*[4], Sarika Hegde[#]
*UG Student, NMAM Institute of Technology
Nitte-576201, India

[#] Assistant professor, Computer Science Engineering
NMAM Institute of Technology,
Nitte -576201, India

***Abstract*—Automatic speech therapy for children assist patients to improve their voice disorder problems. The interactive feedback through games makes the process of speech therapy more entertaining for kids. In this regard, this paper discusses the design and implementation of an automatic speech therapy. Pitch of the given voice is computed using DSP techniques. Based on the pitch of the kids, the system generates a corresponding score by comparing histograms of the spoken words with that of the stored normal voice. The game is opened based on the threshold value of score. The implementation and experimental results are encouraging giving upto 90% accuracy.**

***Keyword*-Speech therapy, Interactive feedback, Game, Children, Voice disorder**

I. INTRODUCTION

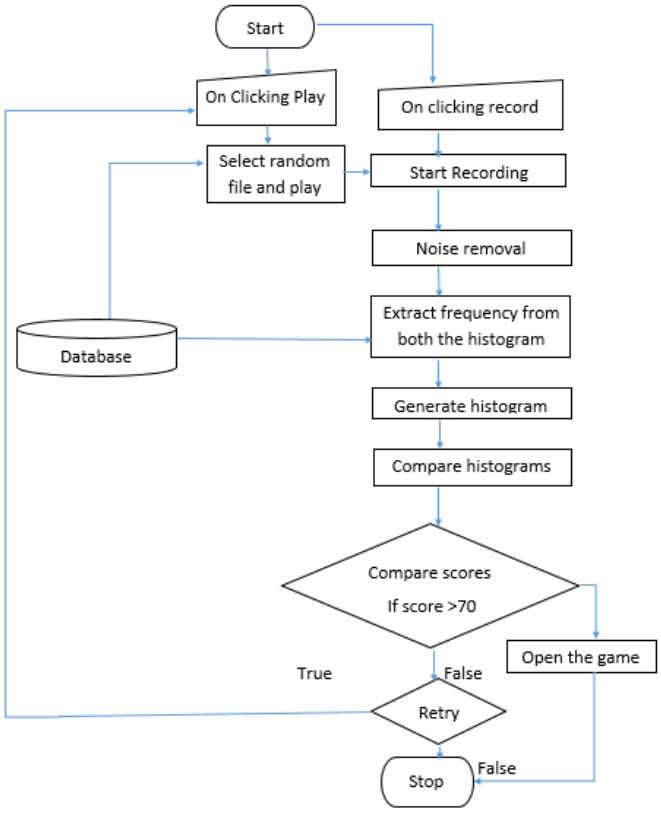
Language training is a mediation benefit that spotlights on enhancing a tyke's discourse and capacities to comprehend and express dialect, including nonverbal dialect. Language teachers, or discourse and dialect pathologists (SLPs), are the experts who give these administrations. Language instruction incorporates two parts: 1) organizing the mouth to create sounds to shape words and sentences (to address explanation, familiarity, and voice volume control); and 2) understanding and communicating dialect (to address the utilization of dialect through composed, pictorial, body, and sign structures, and the utilization of dialect through elective correspondence frameworks, for example, web-based life, PCs, and iPads).
Discourse issue is extensively delegated takes after:
Various auxiliary issues can cause voice issue. Contact ulcers, growths, knobs, polyps, and so forth are a portion of the notable basic sores causing voice issue. These sores additionally cause torment in the throat district. The situation of the sores influences the nature of voice delivered. This is called as structural voice disorder.
Functional Voice Disorders is caused when muscles of the larynx may not work legitimately, with no specific auxiliary or natural reason. The flawed utilization of the muscles may result because of unreasonable pressure. Along these lines, utilitarian voice issue is comprehensively named as muscle pressure dysphonia, or MTD. The term likewise incorporates a condition in which the voice sounds typical, yet aims agony, distress or exhaustion to the voice.
Neurogenic voice disorders are voice issue caused by issues in the focal or fringe sensory systems fall under this class. Loss of motion of the entire body, facial loss of motion, hemiplegia and comparable neurological conditions influences the nerves of the larynx, which thus bothers voice creation. Loss of motion of the vocal folds causes feeble, raspy or unpleasant voice or only a whisper.
Conversion dysphonia is a voice issue coming about because of physical appearance of a mental injury or strife. The individual may have encountered\ or seen a stunning occasion, for example, mishap or passing of a dear one, or sexual manhandle. Practical treatment for voice control will be effective just if the hidden mental condition is investigated. This is known as psychogenic voice disorder.
Strategies for language training is extensively delegated demonstrated as follows:
Sensory feedback causes your kid to wind up more mindful of the sounds that he is creating and how he is delivering them. For instance, the SLP may utilize sound-

related input. She may record an example of your youngster's vocalizations and play them back. She would then play a chronicle of similar sounds, articulated effectively, with the goal that he can hear the distinction. Your youngster can likewise profit by visual criticism or watching himself or others deliver sounds.
Articulation speech therapy activities not with standing material input, your youngster's language teacher will utilize other language instruction exercises for enunciation to enable him to articulate certain sounds. The SLP may obviously express a word and ask that your adolescent rehash it. This will be done different circumstances to give your kid a lot of training.

Oral motor therapy techniques are some discourse issues portrayed by poor control of the oral engine muscles. This can meddle with your kid's discourse, as well as his gulping and sustaining capacities also. Oral engine activities and facial back rub might be utilized to enhance muscle tone. The language instructor may likewise deal with presenting sustenance's of different surfaces and temperatures keeping in mind the end goal to raise oral mindfulness. Oral engine treatment procedures regularly accentuate language instruction exercises that match developments with sounds. For instance, your kid may be requested to consider the adjusting of his lips as he creates the "sh" sound.
Language Intervention are speech and dialect issue cannot just affect a kid's capacity to verbalize plainly, yet additionally his capacity to comprehend dialect and communicate properly. A few children may have trouble articulating their musings, following or giving headings, or portraying a story. Your youngster's SLP will energize dialect securing by utilizing books, pretend amusements, and comparative language training exercises. [1]
The software designed is a game to help children having speech disorders by asking them to repeat the pre-recorded words. The score is produced accordingly. Contingent on the score of the child, there will be an advance in the game.

II. METHODOLOGY

This speech therapy software is a game which focuses on the child's development in speech. It trains children by asking them to repeat prerecorded words by allowing the child to listen to the prerecorded words first and then repeat it individually. then, once the child is pretty good at each word, the score of the child in the game increases. bubble blaster is a game which is used in this project which is played by the child which is an interactive game that makes the learning process more interesting and fun for the child. The steps, include recording, noise removal, extracting frequency, generating histogram and comparison.



A. Recording and Noise removal:

The audio of the patient is recorded using pyaudio library in Python. The recording starts when the user clicks the record button and continues recording until the user stops speaking. The recording automatically stops when it detects silence. Noise removal is done on silent chunks of audio data. These silent chunks or wind noise frequencies are replaced by NaN(Not a number) values.

B. Frequency Extraction:

The number of occurrences of a repeating event for unit time is called as frequency also referred as temporal frequency. It emphasizes the contrast to spatial frequency and angular frequency. It is an important parameter used in the field of science and engineering to specify the rate of oscillatory and vibratory phenomena, such as mechanical vibrations, audio signals (sound), radio waves, and light.

In communication, the usable voice frequency band ranges from around 300 Hz to 3400 Hz. It is thus that the ultra-low frequency band of the electromagnetic range somewhere in the range of 300 and 3000 Hz is additionally alluded to as voice frequency. The data transfer capacity apportioned for a solitude voice-frequency transmission channel is normally 4 kHz,

including guard bands, permitting a sampling rate of 8 kHz to be utilized as the premise of the pulse code modulation system utilized for the digital PSTN. Per the Nyquist– Shannon sampling theorem, the sampling frequency (8 kHz) must be at least double the highest component of the voice frequency through proper filtering prior to sampling at discrete times (4 kHz) for effective reconstruction of the voice signal.

Using the pyaudio library, frequency can be extracted from a 16-bit mono channel .wav file with a preset chunk size.
The frequency extraction code has a formula using which the frequency is calculated for each chunk (frame) of audio data.
The extracted frequency is appended to an array chunk by chunk.

Formula :

$$x1 = (y2 - y0) * .5 / (2 * y1 - y2 - y0)$$

The FFT of the data times a window (a blackman window), the FFT is squared, the bin that has
the highest value is found and a quadratic interpolation around the peak is used using the log
of the max value and its two neighbouring values to find the fundamental frequency.
The accuracy can be increased or decreased by increasing or reducing the chunk size respectively.
The chunk size should be a multiple of 2 to make full use of the FFT.

This process is repeated for all the prerecorded audio samples present in the software database. The average of these arrays is calculated and stored in a new array.
Now, the child's audio is used. The frequency from this file is extracted and stored in an array.

C. Histograms:

Histogram is used to find the number of frequencies that occur between some preset ranges. Matplotlib Library is used to get the histogram features. The frequency arrays are the inputs for the histogram code and the histogram is plotted. The 2-histogram data is further stored in 2 arrays.
Histogram is used to represent the numeric data accurately. It is different from a bar graph, while the bar graph relates two variables, histogram relates only one variable. In order to construct a histogram, the first step is to "bin" the range of values, which is to divide the entire range of values into a series of intervals which need not necessarily be of same size and then count how many values fall into each interval. If the bins are of equal size, we have a rectangle which is formed over the bin with height proportional to the frequency. If the bins are not be of equal width, we have a rectangle formed which is defined to have its area proportional to the frequency. The vertical axis is frequency density and not frequency, where frequency density is the number of cases per unit of the variable on the horizontal axis.

D. Comparison of two histograms:

The 2 histogram arrays are compared, and a score is generated depending on how accurate the values are. This score will be used to tell the child his/her performance.

III. RESULT AND DISCUSSION

The dataset is prepared by recording the selected words, listed as, Apple, Mango, Orange, Pineapple, Banana, Watermelon, Grapes, Strawberry, Cherry, Papaya. Five speakers of age-group 5 to 15 are uttered two samples of each word, totally constituting 14 samples of each word. The mean of the frequency for normal speaker and patient are, 296.42 and 241.72 respectively. After all the background calculations and computations are completed, the score gets displayed on the display as shown as in Fig.1.

The computed frequency is compares with the threshold frequency. If the frequency is above threshold, then the child gets to play the bubble blaster game. If the child wants to try the game again then he/she will have to redo the speech therapy and the game continues.

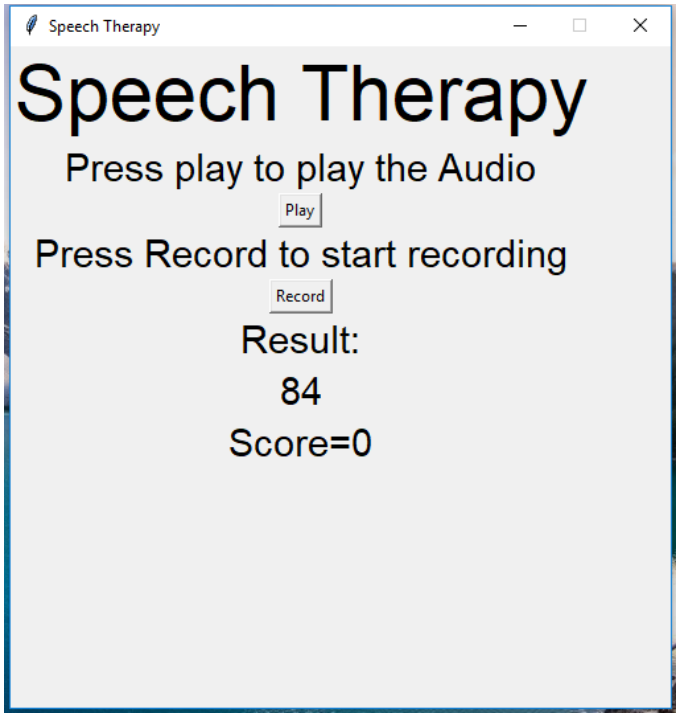


Fig.1 Frequency distribution of the patient after performing speech therapy

When the play button is clicked, a file (fruit name) is randomly selected and played. On clicking the record button, the recording begins and records until silence is detected. Before saving the recorded file, the noise is removed. Histograms are plotted for all the 14 sample files of the randomly selected fruit and the average histogram is plotted using the 14 histograms. Histogram is then plotted for the patient's voice (the recorded file) and both the histograms are compared with a variation of +/- 3. If the difference is less than 3 then it is considered as success for that range. The percentage is calculated considering the total number of success. If the score percentage is greater than 70 then the game opens. Once the patient losses the game, he gets a chance to retry and improve his score and this process repeats.

Let us consider a test run of this software. Following are the results we got from the test run.

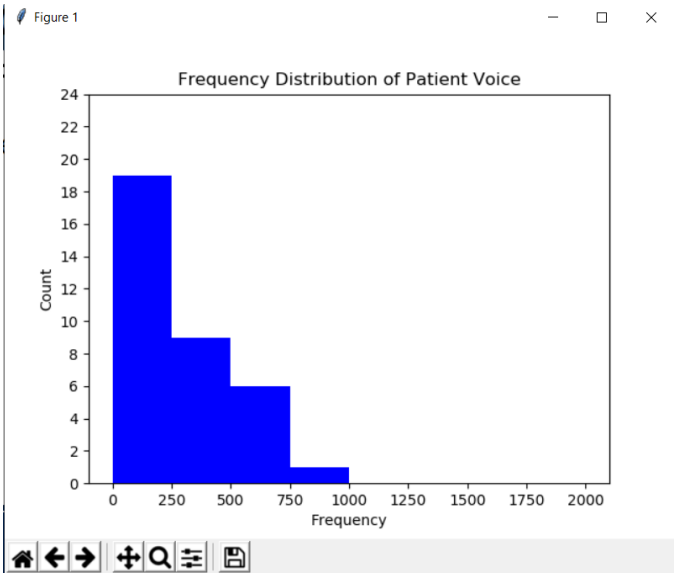


Fig.2 Frequency distribution of patient without disorder

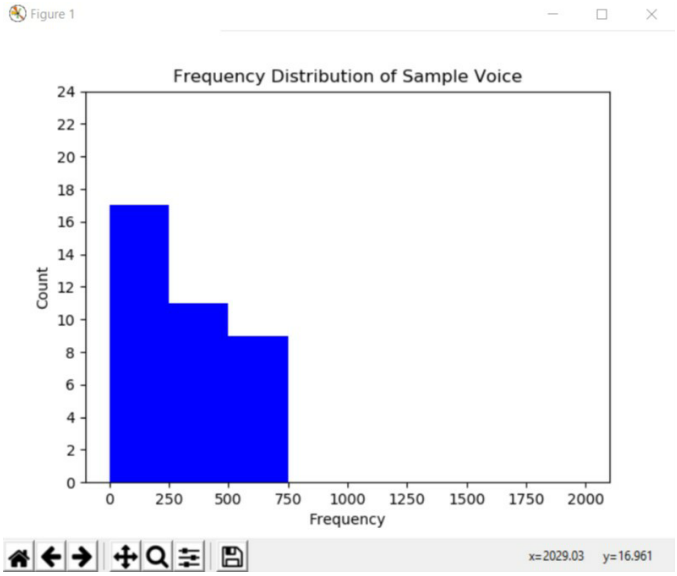


Fig.4 Average frequency distribution of pre recorded audio sample

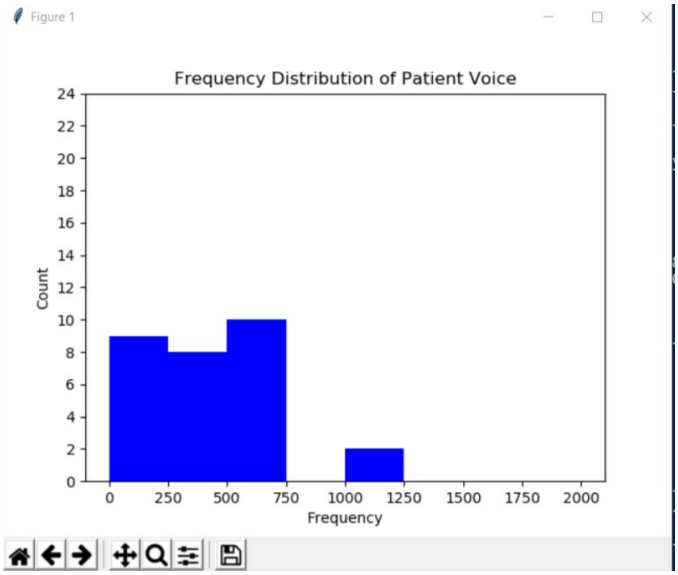


Fig.3 Frequency distribution of patient with disorder

Serial no.	Score	Trial Outcome
1	61	FAIL
2	92	PASS
3	61	FAIL
4	61	FAIL
5	61	FAIL
6	69	FAIL
7	61	FAIL
8	76	PASS
9	69	FAIL
10	61	FAIL
11	84	PASS
12	61	FAIL
13	69	FAIL
14	69	FAIL
15	76	PASS
16	76	PASS
17	76	PASS
18	53	FAIL
19	53	FAIL
20	53	FAIL
21	61	FAIL
22	84	PASS
23	61	FAIL
24	61	FAIL
25	69	FAIL
26	53	FAIL
27	69	FAIL
28	69	FAIL
29	69	FAIL
30	61	FAIL
31	76	PASS
32	69	FAIL
33	69	FAIL
34	69	FAIL
35	69	FAIL
36	69	FAIL
37	69	FAIL
38	69	FAIL
39	61	FAIL
40	69	FAIL
41	61	FAIL
42	69	FAIL
43	69	FAIL
44	53	FAIL
45	61	FAIL
46	53	FAIL
47	61	FAIL
48	61	FAIL
49	69	FAIL
50	53	FAIL

Fig 2. Trial scores and their respective outcomes for a patient with voice disorder.

IV. CONCLUSION

We can use voice bands and other voice features to train a child with speech disorder by with the help of another human.
At present, a patient must be personally given speech therapy and there is a constraint of time which leads to some patients not receiving speech therapy properly. Speech therapy software will help many patients to improve their speech (speaking skills). A patient can train themselves to improve their speech at a convenient time. Speech therapy can be done for free of cost.

V. REFERENCES

[1] The Parents website.[online]. Available : https://www.parents.com/kids/development/learning-disabilities/what-is-speech-therapy/
[2] American Speech-Language-Hearing Association(ASHA) website. [Online].Available:https://www.asha.org/practice-portal/clinical-topics/voice-disorders/