ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Stuttered –Non-Stuttered: A Way of Classification Using Low Power ARM-Core.

N Raju* and P Neelamegam.

SASTRA University, Thanjavur, 613401, India.

ABSTRACT

An embedded system based approach for classification of Stuttered and non-stuttered speech is proposed in this paper using LPC2148 NXP microcontroller. The conventional classification is done based on the extraction of the features of the uttered word and then finding the GMM model for the same and scoring is done based on the Maximum likelihood. The proposed method utilizes less computation and uses only the speech rate timing information of the uttered word and the energy associated with the word is taken as the features for decision making. The uttered word beginning and the end is identified by using the interrupt associated with the key which is used for the recording of the uttered word. The calculation of the timing and the energy calculation for the uttered speech are done by using LPC2148 NXP microcontroller. This method gives 99.8% accuracy.

Keywords: Speech rate, GMM, Maximum likelihood, features, LPC2148.

**Corresponding author*

INTRODUCTION

Speech processing being a wide scope for research, with the cutting edge technology is finding huge application in all the fields. Speech recognition (SR) is been explored to greater extend by using various techniques like Hidden Markov Models combined with GMM, K-mean clustering, SVM.

Modern SR systems use various combinations of a number of standard techniques for improvisation of the recognition rate. Mostly for the SR system to work with very good results requires large-vocabulary system with context dependency for the phonemes. The modeling approach for the speech recognition is done using statistical way in which GMM , HMM are mostly used[1] . The SR systems used for the IVR(Interactive Voice Response) response or the human machine interface using voice response does well when the environment with respect to the recoding is more mute. The features extracted from the recorded speech signal like MFCC, Pitch ,spectral energy ,delta and delta-delta coefficients to capture speech dynamics works sufficiently well when the recording environment is good with respect to the SR system[2-3]. Many additional features might use heteroscedastic linear discriminant analysis (HLDA) for increasing the scoring of the SR system like minimum classification error (MCE) and minimum phone error (MPE)[4].

Existing conventional techniques [5-6] in the recent past ,does well when it comes to normal speaking person but when comes to the people with disability with respect to Speech disfluency ,which is considered as obstacle to normal speech flow, the whole SR system fails. The cause of disfluency may be due to various reasons like stressed, nervous, excited, or tired.

There are several normal speech disfluencies, including Repetition or correction, False start, Filled pause, Interjection, Exiting term, Discourse marker. These normal speech disfluencies does not cause any effect to the SR system. But the stuttered speech can hit the SR system to the greater extend if not taken care properly along with the noise model. This paper considers the stuttering speech. Researchers are still studying the underlying causes of persistent stuttering. A possible cause of persistent stuttering includes abnormalities in speech motor control, Genetics, Medical conditions, mental health problems.

People having the speech disfluency cannot perform well in SR system which gives a low scoring when working with. The consideration of the SR system totally depends on the speaking rate which in turn depends on the articulation rate, pause duration, length of articulatory pharse and the amount of pause time. The varying speech rate[7] brings done the performance of the SR system. If the existing SR system has to be used for all kind of people with or without speech disfluency then the speech rate has to be considered and rectified for improving the scoring of the SR system. The proposed method helps in finding the speech rate which can provide a front end for the Speech-to-Speech systems increasing the speech intelligibility and can make the job of SR system easier. This proposed system can be used for classification of the stuttered and non-stuttered speech using the time domain methods utilizing the time taken to utter the word and the energy in the speech signal. The calculation for the abnormal pauses (no sound), repetitions (st-st-stuttering), or prolongations (ssss stuttering) of sounds and syllables timing and the energy calculation for the stuttered speech is done by using LPC2148 NXP microcontroller and then compared with the normal speech.

PROPOSED METHODOLOGY

In this method a front end for the SR system for finding the speech rate[8] which is calculated by counting the number of syllables for the entire recorded time per person is proposed which can be used for the classification of the stuttered and non stuttered speech is developed by using LPC2148 NXP microcontroller. The sampling frequency is considered at 8Khz for the speech signal. The frequency for the microcontroller is set to 60 MHz by using the internal PLL0 module and external crystal 12MHz. Figure 1 block diagram of the experimental setup, shows the key attached to the external interrupt INT0 to the microcontroller is used to trigger the recording of the uttered word. The Figure 2 shows the flow of the designed module. In this the user before uttering the word presses the key for recording the uttered word. Once completed the word the user releases the key so as to confirm the completion of the uttered word.

All authors have to sign the copyright transfer form. The signed copyright form has to be scanned and uploaded by using the corresponding interface on the website. There are some restrictions regarding the length of the copyright file, so please read carefully the instruction provided there.

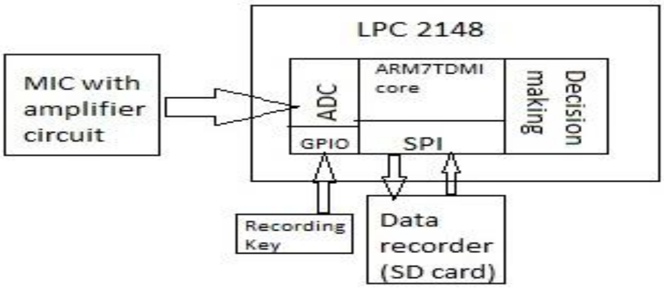

Figure 1: Block diagram of experimental setup

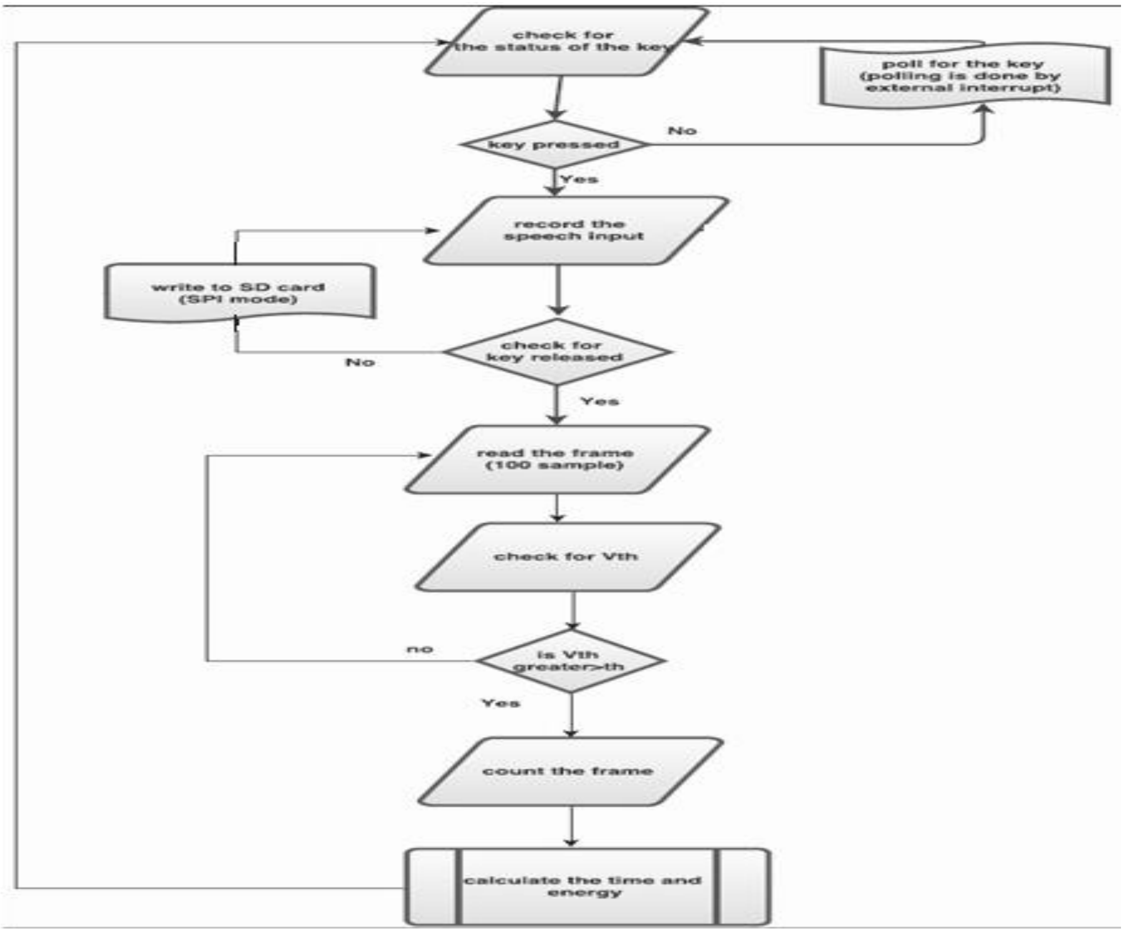

Figure 2: Design flow chart

Initially the recorded signal is buffered in the internally RAM of LPC 2148 which is of 32Kb can be extended to 40Kb if USB is not used. This memory capacity leads us to the recording of maximum 0.8192 seconds as calculated using equation 1. Each of the recorded samples is represented with 8 bit. Thus the memory inbuilt in the microcontroller is not sufficient enough for the stuttered speech as the internally memory cannot be utilized fully for doing only buffering. Thus SD card is been used as a data logger in which the uttered speech is recorded and written. The SPI port of LPC2148 is used for the communication between the SD card and the microcontroller. Since a simultaneous operation has to be done with respect to acquiring and recording the speech signal in one go, the conversion time of the ADC module which is having a 10 bit resolution is made to operate at the frequency of 4MHz which is the maximum frequency at which the said module can do its job. .

ISSN: 0975-8585

$$\text{Recording time} = \frac{Total\ number\ of\ samples}{F_s} \qquad (1)$$

where F_s is the sampling frequency

$$\text{MR} = \frac{Number\ of\ bits}{Sample} \times Total\ number\ of\ samples \qquad (2)$$

where MR is Memory Requirement.

The computation has to be done at much faster rate in order to obtain a reliable output. Thus the recording is done with the sample of 8 KHz and each sample is represented by 8 bit data.

The electrolyte MIC (sound detector) from the spark fun is been used for the recording. Sensitivity is taken care by the below given circuit figure 3 which is an audio amplifier using LMV324.

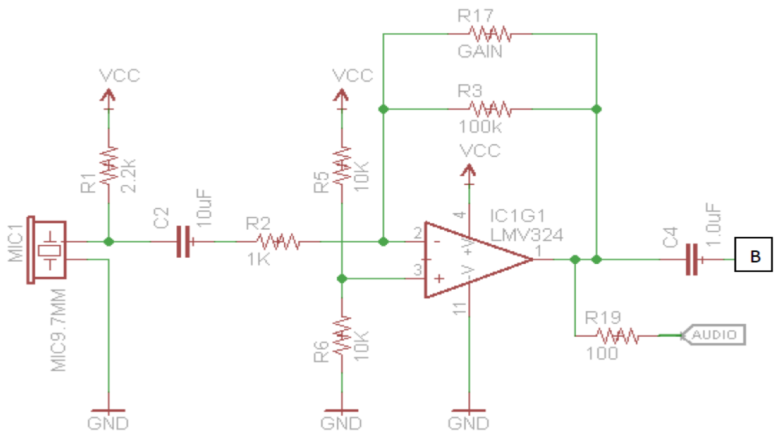

Figure 3: Audio amplifier circuit using LMV324

The level shift is done in such a way that the data lies between 0 Volts to 3.3 Volts. The said module does have the peak detector circuit. The gain of the circuit is easily adjustable by using R17.

The timing calculation is done once the speech data is recorded into the file utter.txt in the SD card. The timing consideration is taken from the time of the press of the button and to the release of the button and also the energy for the whole uttered word is calculated. The calculation is done as given in the equation 3. The noise is taken care by taking the recorded speech sample as frame of 100 samples and calculating the energy by using the equation 4. If the frame has the value above the threshold(V_{th}) then it is taken as speech signal or else the frame is dropped for the computation as it is a noise and the frame drooped is not taken for the purpose of the timing information and as well as the energy calculation.

$$Time = (total\ count\ of\ the\ frame \times 100 + R) \times \frac{1}{F_s} \qquad (3)$$

Where R residue frame is the last frame considered.

$$Energy = \sum_{n=0}^{N} x^2(n) \qquad (4)$$

In the experimental part all the uttered word is in tamil language ,since the subject considered for the recording have tamil as their mother tongue.Table 1 shows the measurement result done by using the LPC2148 microcontroller. The uttered words are all in tamil like Father – □□□□□(APPA), □□□□□(thandahi) . Mother – □□□□(TAY), □□□□(MATHA), □□□□□(AMMA), □□□□□(ANNAI).

TABLE 1: MEASURED DATA USING LPC2148

Uttered word	NORMAL		STUTTERED	
	Time(s)	Energy(dB)	Time(s)	Energy(dB)
அப்பா	0.6532	70.22	1.9295	69.12
தந்தை	0.6110	68.12	1.4878	59.33
தாய்	0.7835	66.02	1.7451	65.01
மாதா	0.8000	69.31	1.9178	63.20
அம்மா	0.6781	66.32	0.9009	62.36
அன்னை	0.7234	65.10	1.9256	53.16

Results of the table 1 gives the clear difference between the stuttered and non-stuttered speech with respect to the measured speaking rate in tamil the general rate of speech is considered as 180-200 words/min[8] which can be near about 250 syllables/min. This way of measurement is further given a comparison by doing the noise filtering and finding the start of the word and the end of the word after the key is pressed the second method is done using Praat tool where the word recorded is analyzed for the syllable marking and the articulation rate and the pausing rate[9] is calculated and the added sum is considered as the rate of speech the measured values are tabulated in the Table 2.

TABLE 2: MEASURED DATA USING PRAAT TOOL

Uttered word	NORMAL		STUTTERED	
	Time(s)	Energy(dB)	Time(s)	Energy(dB)
அப்பா	0.6232	73.56	1.9295	68.45
தந்தை	0.5110	66.53	1.5678	57.13
தாய்	0.6835	66.27	1.6465	63.29
மாதா	0.8015	66.33	1.6351	63.22
அம்மா	0.4866	67.63	0.5764	67.39
அன்னை	0.6297	60.21	1.6027	59.63

Figure 4 shows the plot of the uttered word □□□□□ by the stuttered subject in which lot of repeated syllable appears. The uttered word is recorded using the PRAAT tool with 8KHz sampling frequency.

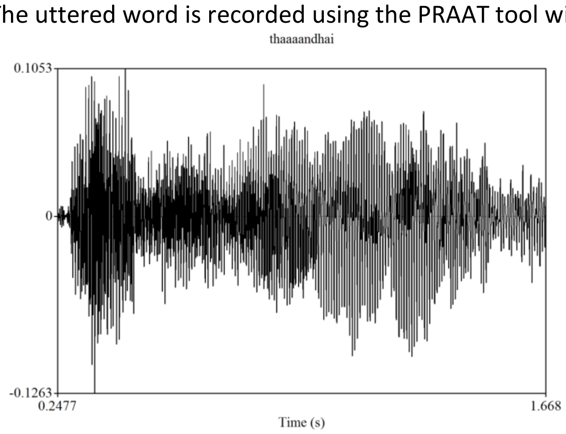

Figure 4: Plot for the word தந்தை stuttered one.

Figure 5 shows the plot of the uttered word □□□□□ by the stuttered[10] subject in which the repeated syllable is been removed manually by using the praat edit tool .The difference in the stuttered and

ISSN: 0975-8585

the edited stuttered speech can be visualized in the figure 4 and figure 5.

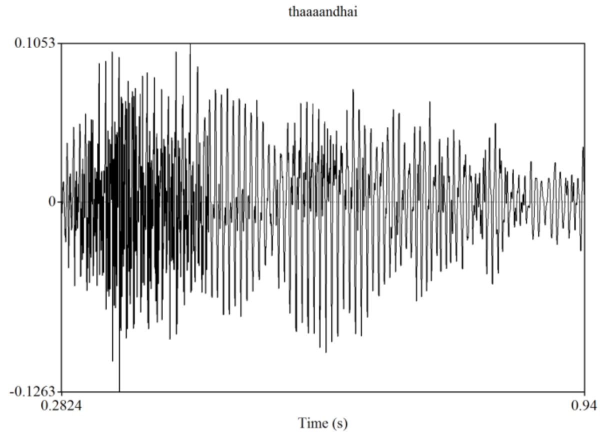

Figure 5: Plot of the stuttered word after removing the repeated syllable.

The Figure 6 which shows the plot of the uttered word □□□□ the normal subject and the Figure 7 shows the plot of the uttered word □□□□ the stuttered subject. figure 8 shows the edit version of the uttered word □□□□ the stuttered subject manually using the praat tool. We can notice that the edit plot shown in Figure 8 almost resembles the Figure 6. This gives a clear indication that the stuttered speech can be converted into the normal speech if we can detect the entire repeated syllable and remove the unwanted repetition. This gives a clear indication of improving the speech to speech intelligibility[11–12]

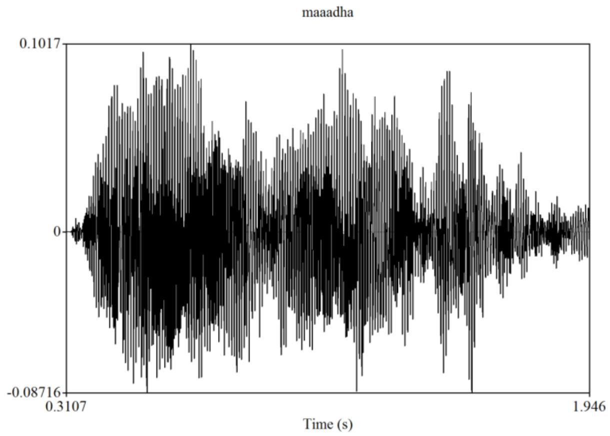

Figure 6: Plot of the normally uttered word மாதா.

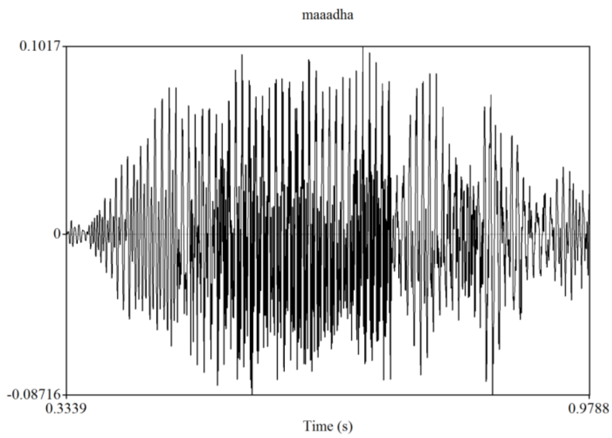

Figure 7: Plot of the uttered word மாதா (stuttering).

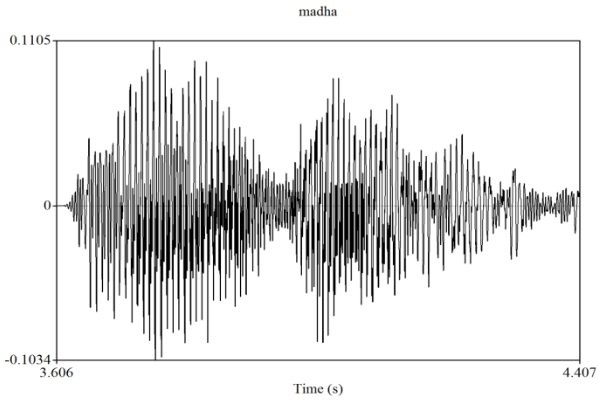

Figure 8: Plot of the uttered word மாதா (stuttering) after removal of repeated syllable.

The entire data is been collected from the SASTRA university staff and from the authors relatives who have developed the stuttering problem right from the birth and some of the data for the reference is taken from movie clipping.

CONCLUSION

The proposed method gives a good calculation for the speaking rate for tamil language which can be considered as the front end for the SR system to the people with stuttering disability . this can also be used for detecting the speech rate for doing the classification between Stuttered and Non-Stuttered speech, once the classification has been done with, then the repeated syllable can be removed so as to improvise the speech rate close to the normal ,which can be further enhanced after the classification of the speech by using various techniques like HMM model for improving the scoring of the SR system .This technique can be further used for finding the repeated syllable from the stuttered speech and by eliminating that from the recorded speech signal can give the concept of speech to speech which can increase the speech clearness.

REFERENCES

[1] Gibson.M , Hain T, IEEE Transactions on Audio, Speech, and Language Processing 2012; 20(10): 2648–2656.
[2] Li .P, Tang. h , IEEE Transactions on Circuits and Systems 2011;58(5): 961–970.
[3] Yu .K, Gales M. J. F., , IEEE Transactions on Audio, Speech, and Language Processing 2006; 14(5): 1694–1702.
[4] Yu .D , Deng .L, Gong .Y, Acero .A, , IEEE Transactions on Audio, Speech, and Language Processing 2009; 17(7):1348–1360.
[5] Yu .K, Gales .M, Woodland P.C, IEEE Transactions on Audio, Speech, and Language Processing 2009; 17(4): 714–723.
[6] Kim D. K , Gales M. J. F., IEEE Transactions on Audio, Speech, and Language Processing 2011; 19(2): 315–325.
[7] Rodero E,A Text&talk 2012; 32(3): 391–411.
[8] Schelten-Cornish .S Die Sprachheilarbeit 2007; 4(1): 136–145.
[9] Pellegrino J. L, François Farinas, Rouas J, Speech Prosody , International Conference 2004, 1(1): 1–4 .
[10] Yuan J, Liberman M, Cieri .C, Proceedings of ICPhS XVI 2007;16(1): 1337–1340.
[11] Arcuri C. F., Osborn .E, Schiefer A .M, Chiari B .M, Pro Fono 2009; 21(1): 45–50.
[12] Francis A.L, Nusbaum H.C, The Journal of the Acoustical Society of America 1996;99(4): 2493-2500.