

DEAF SPEECH ASSESSMENT USING DIGITAL PROCESSING TECHNIQUES

C.Jeyalakshmi.¹ and Dr.V.Krishnamurthi ² and Dr.A.Revathy ³

¹Department of ECE,Trichy Engineering College,Trichy.

lakshmi.jeya67@yahoo.com

²Department of ECE,Trichy Engineering College,Trichy.

profvkmurthi@yahoo.co.in

³Department of ECE,Saranathan College of Engineering,Trichy.

revathidhanabal@rediffmail.com

ABSTRACT

This paper mainly deals with analysis on acoustical characteristics of speeches of deaf people for the purpose of increasing the speech recognition rate. Since speech to text or sound system for a normal speaker is available, by designing a speech to text or sound system for deaf, they can make use of all computer aided devices and normal speakers can also communicate with them freely. Fundamental frequency or the pitch frequency of the vocal fold and resonant frequency of the vocal tract or formants are considered for analysis which are the foremost characteristics of speech. Compared to normal speech, there is a high variability in deaf speech and by hearing once we couldn't understand it. Deaf speech is taken from children in the age group of 5-10 years from Maharishi vidya mandir centre for hearing impaired. Another set of speech were taken from normal speakers for comparison. Initially the input is sampled, filtered, windowed and Pitch frequency is determined for each frame. Similarly first six formants are determined for each frame. The fundamental frequency contour of deaf children exhibit unusual characteristics, and the formants are also very closed. This shows that, pitch and formants cannot be used as features for deaf speech recognition. At the same time, variation in the pitch and formants for deaf is larger than normal speakers it can be used for speaker classification purpose.

KEYWORDS

Linear Prediction Coefficients(LPC), Pitch detection algorithm (PDA), Sub harmonic to harmonic ratio(SHR),Speech signal processing, deaf speech

1. INTRODUCTION

The Royal National Institute for Deaf People (RNID) is a charitable organization working on behalf of the UK's 9 million deaf and hard of hearing people, currently estimates that about 8.7 million people in the UK have some form of hearing loss, with about 673,000 people being severely or profoundly deaf. More than 400,000 people cannot use a voice telephone even with a hearing aid or other amplifier. The effect of hearing loss on an individual is largely depending upon the degree of loss and age at onset. If profound or total deafness be present at birth, or occur within the first few years of life, then that individual will probably develop communication skills using sign language. Most of the deaf people in the UK are British Sign Language (BSL) users. People who become hard-of-hearing or deafened later in life, through old age or illness, generally will continue to use spoken English. Depending on the degree of hearing loss, people in this group have several options: use additional amplification or a hearing aid, consider a cochlear implant, or learn to lip-read. In fact, lip (or speech) reading is an extremely difficult skill which requires the deaf person to study the lip movements and facial expressions of the speaker, together with numerous other factors (such as accompanying physical gestures) to determine what is being said. There are many potential obstacles to lip reading. Hearing aids and lip-reading are most effective in face-to-face communication between a small numbers of people. Unfortunately, there are many events such as public meetings and lectures where the speaker may be poorly lit or

too far away to be seen or heard clearly, or where high levels of background noise prevent the successful use of a hearing aid. It is in these circumstances when a simultaneous visual transcript of speech may be helpful. [1]

One of the problems associated with deafness is that it often results in poor-quality speech. This is most marked in those born deaf, since their inability to hear their utterances prevents the acquisition of speech in a normal way, and, indeed, severely affects many of the learning processes. With people whose hearing becomes severely impaired in later life, deterioration in speech quality may also take place because of the loss of acoustic feedback, even though they have well established skills in speech production. Various training methods are used for this problem of deaf speech. Most rely heavily on a trained teacher who basically demonstrates the correct production of an utterance that the pupil learns by feeling vibrations of the teacher's and then his own throat, nose etc. by hand, and by observing positioning of lips, tongue etc. by eye. In addition, electronic aids, such as pitch indicators, are sometimes employed. There is now a growing interest in the use of computer-based aids.[2]Two major obstacles have hindered progress in the development of speech processing aids for the deaf. The first is a lack of basic knowledge of how speech is acquired, produced, and perceived. Thus, even with the sophisticated electronic instrumentation of today we still do not have a perceptual aid that is substantially superior to a good quality conventional hearing aid. The second major obstacle is one of our own making in that, until quite recently, there have been very few attempts at objective evaluation of potentially useful aids. Without a body of objective data on which to build, it is virtually impossible to make progress in any systematic or reliable way.[3]

In another scenario when we want to recognize a deaf and dumb speaker's speech, in order to operate all computer aided devices and for effective communication with others, analysis of deaf speech is important.

One of the problems encountered in analyzing the speech of the deaf is the large variability between speakers. Differences between deaf speakers are substantially greater than differences between normal speakers and thus correspondingly more data are needed to separate out differences between talkers from characteristic differences between deaf and normal speech[4].The language skills of these children are, on the average, severely retarded; their speech production and their speech reception are, at best, of limited use; their vocabulary, grammar, and reading show great deficiencies relative to normal children. consequently, their education is restricted even when the most intense efforts are made to keep pace with normal education[5].The fundamental frequency (Fo) of speech i.e. pitch conveys prosodic information regarding normal communication patterns. Hence, it is essential that the Fo be measured accurately in assessing and in rehabilitating deaf speech[6].

Several investigators have reported the problems of profoundly deaf speakers with pitch control. The characteristic difficulties include abnormally high average pitch and unnatural intonation patterns. These anomalies are sufficient in themselves to make deaf speech sound unnatural and even unintelligible. So poor pitch control decreases the intelligibility of deaf speech. Small tactile pitch displays have the potential for supplying continuous corrective feedback for the improvement of the intonation patterns of deaf speakers. [7] By studying the individual subjects, we didn't find evidence for a clear distinction between the hearing impaired and normal hearing subjects by means of Fo. It can be concluded that the hearing impaired subjects showed more variation in their phonation than their hearing peers did. [8]

Pitch detector is an essential component in a variety of speech processing systems and the pitch contour of an utterance is useful for recognizing speakers. Accurate and reliable measurement of the pitch period of a speech signal from the acoustic pressure waveform alone is often exceedingly difficult for several reasons. One reason is that the glottal excitation waveform is not

a perfect train of periodic pulses. Although finding the period of a perfectly periodic waveform is straightforward, measuring the period of a speech waveform, which varies both in period and the detailed structure of the waveform within a period, can be quite difficult. A second difficulty in measuring pitch period is the interaction between the vocal tract and the glottal excitation. In some instances the formants of the vocal tract can alter significantly the structure of the glottal waveform so that the actual pitch period is difficult to detect. [9]

Improvement of the previously proposed pitch determination algorithm (PDA) is now developed. i.e. particularly aiming at handling alternate cycles in speech signal, the algorithm estimates pitch through spectrum shifting on logarithmic frequency scale and calculating the Sub harmonic-to-Harmonic Ratio (SHR). This algorithm performs considerably better than other PDAs compared. SHR can also be applied to voice quality analysis [10].

This paper is organized as follows. Pitch detection algorithm is explained in section 2. In section 3 formant extraction using LPC is given. The results are compared with deaf and normal speaker and discussed in section 4. Conclusion, references are given in section 5 and 6.

2. PITCH DETECTION ALGORITHM

Normal vowel production results from a quasi-periodic vibration of the vocal folds acting upon the air-stream escaping from the lungs. All sounds produced with vocal fold vibration are known as voiced sounds and the mechanism of speech production is shown in figure 1.

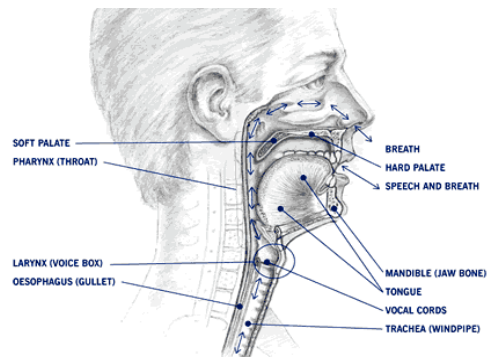


Figure 1: Mechanism of speech production system

While great progress has been made in understanding the physiological and psychological aspects of speech processing, much work remains to be done. An important contribution that auditory science can make to speech processing is to identify what features of the speech stimuli are relevant, and what underlying time frequency analysis strategies should be undertaken in order to extract them. Such features would then form the front end of a speech recognition system, or determine the structure of a speech coder. [11]

The fundamental frequency (F_0) of voiced sounds is determined physiologically by the vocal fold vibration rate. Control of F_0 is used to communicate prosodic features of speech such as stressing and intonation. Production of prosodic features is an essential part of the normal human communication process.

Previous reports on speech indicate that deaf individuals have a significantly higher F_0 than normal hearing individuals therefore, an accurate and valid measurement of F_0 is a critical element in the assessment and treatment of deaf speech. There are at least two methods for determining the F_0 of speech The deaf subject was 5-10 years old and had a deaf, deafened, hard of hearing loss. The normal hearing subject was 7-12 years old with

no significant history of hearing impairment or speech impediment. Each subject was instructed to prolong the isolated digits ten times. Before recording, the subjects were asked to practice for some time to familiarize them with the glottograph.

2.1. Pitch extraction using SHR

The pitch extraction from a speech file is difficult because the glottis excitation is correlated with the vocal conduct. The PDA are based on three main methods :-

-frequency methods such as FFT, Cepstrum, STFT .

-temporal methods: based on the autocorrelation function such as, LPC, Parallel, PPA.

- time-frequency methods: spectrogram, wavelet.

Since the above methods exhibits some disadvantages, SHR is used in which pitch of alternate pulse cycles in speech is taken. This algorithm employs a logarithmic frequency scale and a spectrum shifting technique to obtain the amplitude summation of harmonics and sub harmonics, respectively. Through comparing the amplitude ratio of sub harmonics and harmonics with the pitch perception results, the pitch of normal speech as well as speech with alternate pulse cycles (APC) can be determined which is shown in figure 2. This algorithm is one of the most reliable PDAs. Furthermore, superior to most other algorithms, it handles sub harmonics reasonably well.

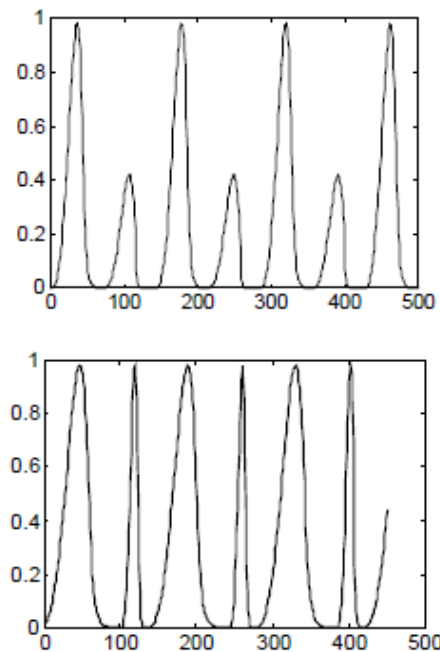


Figure 2: A schematic representation of glottal pulses with alternate pulse cycles (APC). (a) Amplitude alternation. (b) Period alternation.

Sub harmonic-to-Harmonic Ratio (SHR) is amplitude ratio between sub harmonics and harmonics. When the ratio is small, the perceived pitch remains the same. As the ratio increases above certain threshold, the sub harmonics become clearly visible on the spectrum, and the perceived pitch becomes one octave lower than the original pitch. These findings suggest that pitch may be optimally determined by computing SHR and comparing it with the pitch perception data.

3. FORMANTS EXTRACTION

Formants are defined as 'the spectral peaks of the sound spectrum $|P(f)|$ ' of the voice. Formant is also used to mean an acoustic resonance and in speech science, phonetics is a resonance of the human vocal tract. It is often measured as an amplitude peak in the frequency spectrum of the sound, using a spectrogram.

Formant values can vary widely from person to person, and all voiced phonemes have formants even if they are not as easy to recognize. Voiceless sounds are not usually have formants instead, the plosives should be visualized as a great burst.

Formant trackers typically have two steps: 1) computation of formant candidates for every frame, and 2) determination of the formant track, generally using continuity constraints. One way of obtaining formant candidates at a frame level is to compute the roots of a p th order LPC polynomial. There are standard algorithms to compute the complex roots of a polynomial with real coefficients. Each complex root z_i can be represented as

$$z_i = \exp(-\pi b_i + j2\pi f_i)$$

where f_i and b_i are the formant frequency and bandwidth respectively of the i_{th} root. Real roots are discarded and complex roots are sorted by increasing f , discarding negative values. The remaining pairs (f_i, b_i) are the formant candidates.[13]

In our experiments we have used $p=12$. We computed these LPC coefficients from 30-millisecond Hamming windows, with 20 milliseconds overlapping, using the autocorrelation method. Here we have calculated first six formants and only four formants are plotted for clarity.

4. RESULTS AND DISCUSSIONS

Two databases are used for evaluation. The first is the isolated words uttered by the normal speakers. The speech signal is sampled at 16KHz with 16-bit resolution. Here the frame length is taken as 40ms with 20ms overlap, 50Hz-200Hz for F_0 range and upper bound of the frequencies that are used for estimating pitch is taken as 1250Hz. with SHRP threshold is taken as 0.2. Then pitch values are estimated using SHR algorithm[10]. Another database is the isolated words from deaf and hard of hearing children in the age group of 5-10 years. Similarly pitch extraction is done using SHR.

4.1 PITCH COMPARISON OF DEAF AND NORMAL

The estimated values of F_0 is first taken for two deaf speakers, two normal speakers then deaf and normal speaker is compared for different isolated words and they are shown in figures 3 to 11 for three isolated words.

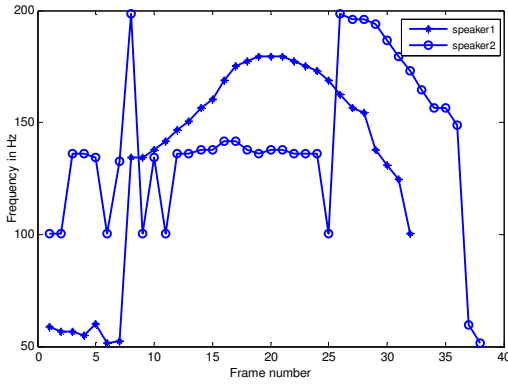


Figure 3: Pitch contour of two deaf speakers for word one

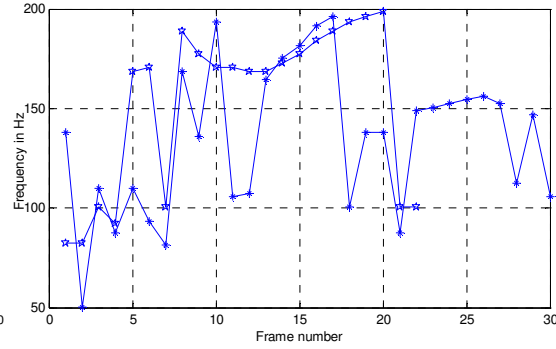


Figure 4: Pitch contour of two normal speakers for word one

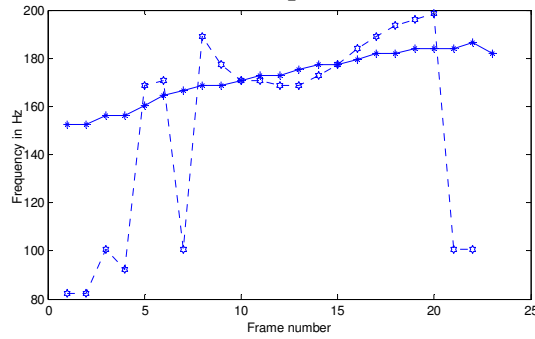


Figure 5: Pitch contour of deaf and normal Speaker for word one

From the above figures 3to 5 it is clear that variation in pitch contour between two normal speakers for the word one is less, compared to the deaf speakers. This variation is very large between a deaf and normal speaker since the speech production of the deaf is completely different.

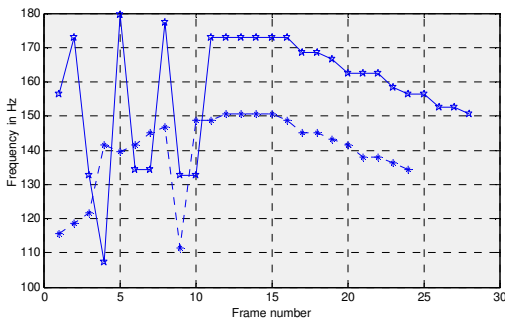


Figure 6: Pitch contour of two deaf speakers for word two

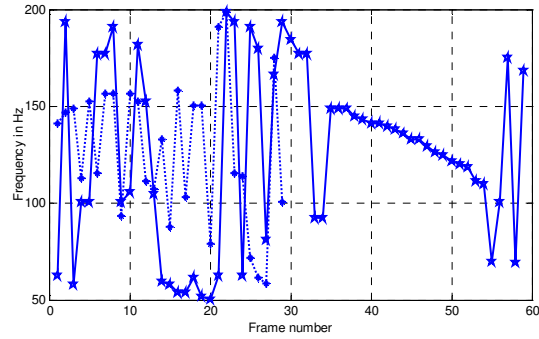


Figure 7: Pitch contour of two normal speakers for word two

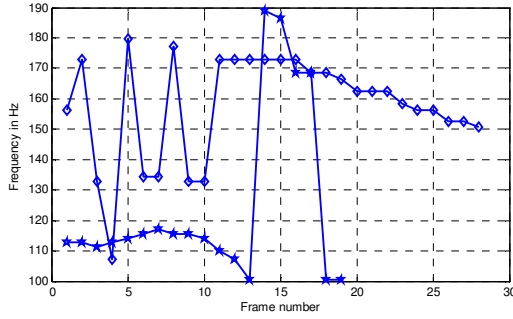


Figure 8: Pitch contour of deaf and normal speakers for word three

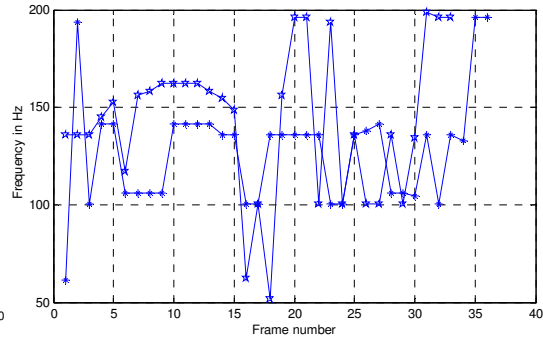


Figure 9: Pitch contour of two deaf speakers for word two

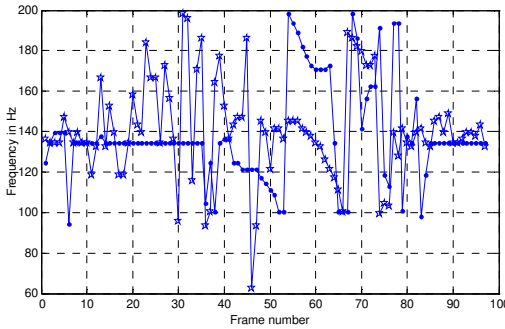


Figure 10: Pitch contour of two normal speakers for word three

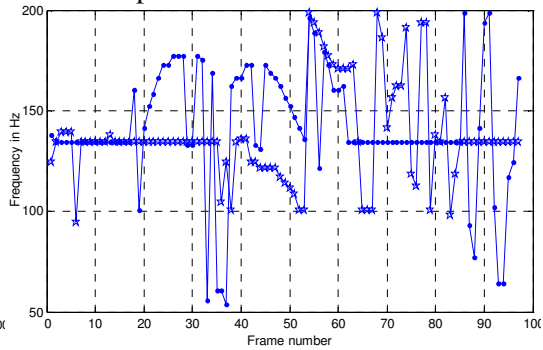


Figure 11: Pitch contour of deaf and normal speakers for word three

Like word one Pitch contour variation for word two and three is also shown in figures 6 to 11. Among these words only word two has too much variation. This shows that speech recognition rate will be somewhat reduced for the word two.

In general the pitch frequency for male will be 100Hz and for female is 200Hz for normal speaker. Here we have shown the mean value of the pitch frequency for both normal and deaf speakers for five isolated words in table 1 and 2.

From the table it is clear that female pitch frequencies are higher than male pitch frequencies. At the same time there is no much variations in the frequencies among the male speakers and among the female speakers. Since it is not common for all speakers we cannot use the pitch frequencies for speech recognition.

Table1. Pitch frequency f0 for normal speakers

Speakers	Input isolated words				
	One	Two	Three	Four	Five
N1-Male	153	192	162	171	162
N2-Male	135	213	160	164	158
N3-Male	155	169	163	174	159
N4-Female	245	275	255	217	200
N5-Female	259	247	196	219	181

Table 2. Pitch frequency f0 for deaf speakers

Speakers	Input isolated words				
	One	Two	Three	Four	Five
N1-Male	149	181	168	161	164
N2-Male	181	189	185	186	159
N3-Male	188	180	152	186	160
N4-Female	228	222	226	223	228
N5-Female	314	272	216	339	297

4.2 FORMANTS COMPARISON OF DEAF AND NORMAL

The speech waveform, spectrogram, first four formants of a deaf and normal speaker is shown in figure 8 to 11. From this figure it is evident that the bandwidth of the spectrogram is almost same for two normal speakers. At the same time the bandwidth and the formants are entirely different for two deaf speakers compared to normal speakers.

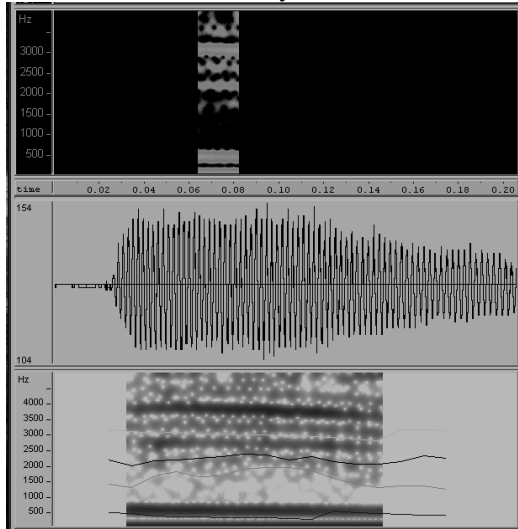


Figure 8: spectrogram, speech waveform, formant plot of normal female speaker for word two

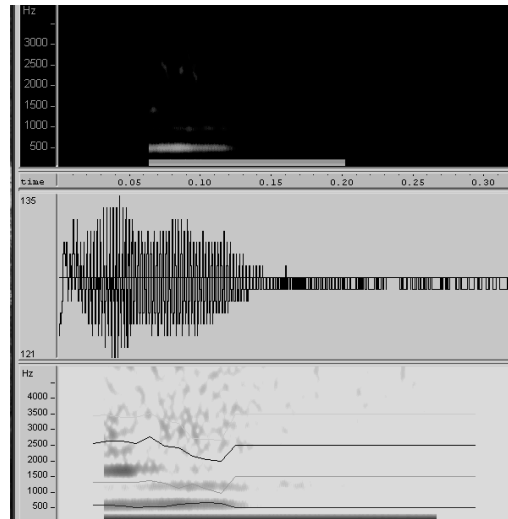


Figure 9: spectrogram, waveform, formant plot of normal male speaker for word two

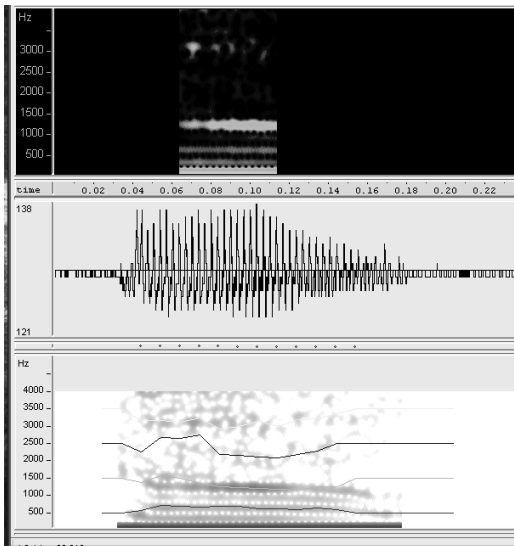


Figure 10: spectrogram, waveform, formant Plot of deaf male speaker for word two

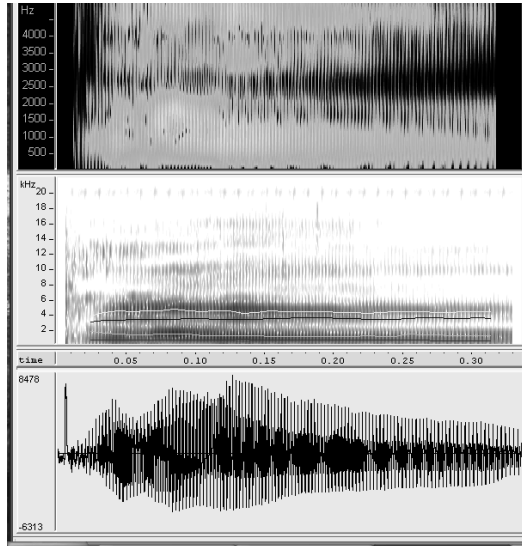


Figure 11: spectrogram, formant plot, speech waveform of deaf female speaker for word two

The following table 3 and 4 shows the first five formant frequencies for the normal and deaf speakers. The first formants for male deaf speakers are lesser than normal speakers. But for female speakers the formants are higher than normal speakers.

Table 3. Formant frequency of normal speakers for word two(I frame)

Speaker	F1	F2	F3	F4	F5
N1-Male	560.5	1650.8	2416	3209	4121.9
N2-Male	4404.	3258.0	1075.8	2308.3	2113.7
N3-Female	509.4	1479.4	2610.8	3360.6	4288.4
N4-Female	3902.	2791.4	1697.3	445.14	547.57

Table 4. Formant frequency of Deaf speakers for word two(I frame)

Speaker	F1	F2	F3	F4	F5
N1-Male	410.1	887.93	1890.8	3032.9	4337.0
N2-Male	4184	3288.2	1998.6	1038.1	401.71
N3-Female	1051	282.94	2227.2	3270.6	4188.8
N4-Female	4185	3197.1	2013.7	668.36	1364.4

The plot of variation in formant frequencies among two normal speakers, among two deaf

speakers are shown in figure 12, 13 for the word one. Similarly normal verses deaf for the word two is shown in fig.14. The data1 to data6 are the first six formants and for some speakers 6th formant is not present.

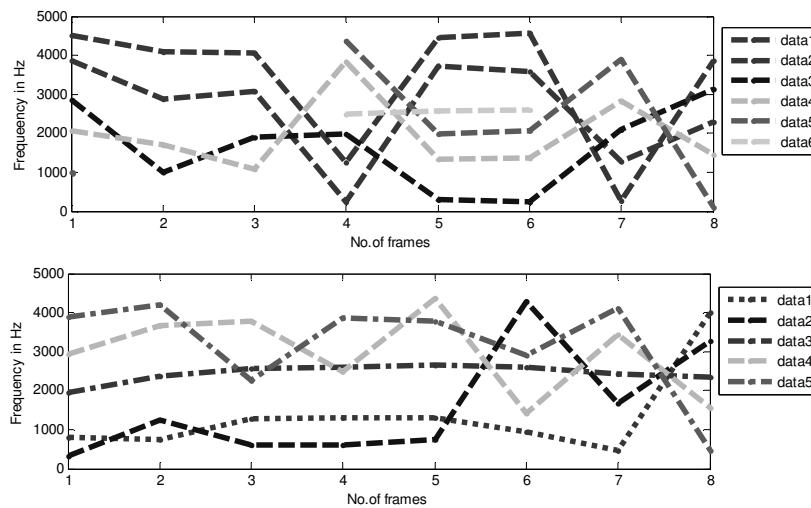


Figure 12: Formants of two deaf speakers for the word one.

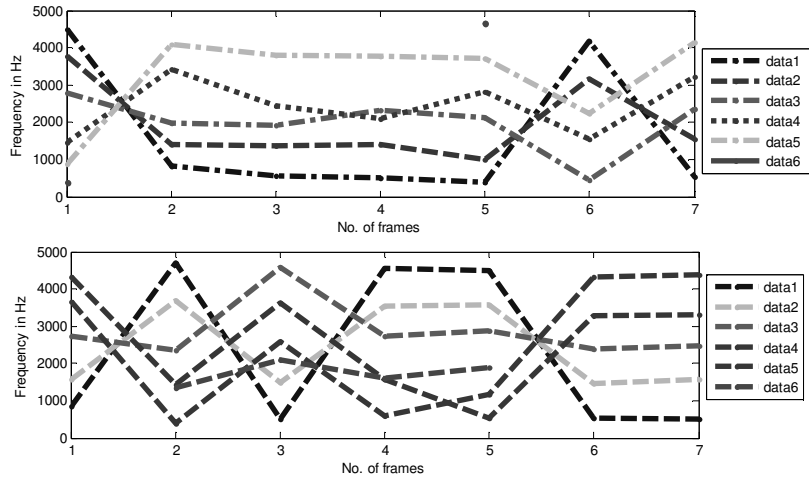


Figure 13: Formants of two normal speakers for the word one

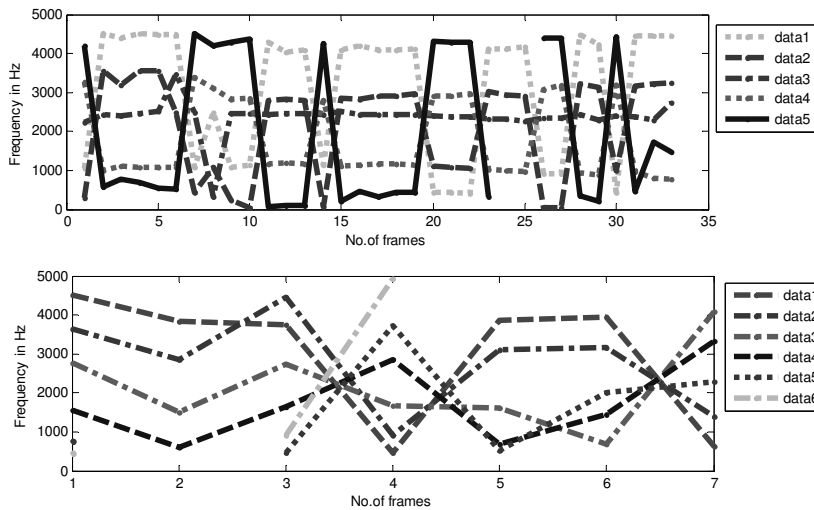


Figure.14 Formants of deaf and a normal speaker for the word two

From the figures it is understood that the formants are very closer for the deaf speakers. Due to this we couldn't easily find the formants for them. At the same time large variation exhibits in the formant plot among the deaf and normal speaker which is shown in fig.14. So that we can use the formants for classification of deaf and normal speaker.

As a result, Pitch and formant frequencies for deaf and normal speakers are taken for consideration and for each measurement, corresponding values were compared using two independent sample tests. The F_0 for deaf speech using SHR measures was significantly higher than the F_0 produced by the normal hearing subject (Table 1,2). In contrast, no significant difference was found between two normal hearing speakers and for two deaf speakers.

Likewise when the formants considered for normal speaker, the variations are less compared to deaf. At the same time there is a large variation between the formants of deaf and normal (fig.14).

5. CONCLUSIONS

The results of this study are based on two subjects, one deaf and one normal hearing. However, the differences observed in the two measurement are expected to occur in other deaf and normal individuals. The results of this study indicate that the differences in measurement of Fo in deaf speakers may be investigated further with a larger sample size. The measure of Fo provided by the SHR includes the fundamental frequency of the vibration of the vocal folds plus any other acoustical energy that is produced in the glottal area.

The pitch is sufficient for the identification of the deaf or normal speaker but has to be assisted by the first four (F1,F2,F3,F4) formants necessary for speech classification. But we cannot use the pitch and formants for deaf speech recognition since it is not common for all deaf speakers for the same word.

6. ACKNOWLEDGEMENTS

Our thanks to the Director Mrs.Geetha, the staff's and students of Maharisi vidya mandir centre for hearing impaired, who have co-operated towards recording of the speech.

REFERENCES

- [1] Dr.Colin Brooks,(2000).“Speech to text system for deaf, deafened and hard of hearing people”, *The Institution of Electrical Engineers IEE*, Savoy Place, London WC2R OBL, UK.
- [2] R.G.Crichton, M.A., and F. Fallside, B.Sc, M.A., Ph.D., C.Eng., M.I.E.E.(1974). “Linear prediction model of speech production with applications to deaf speech training” *Proceedings IEE*, Vol. 121, No. 8.
- [3] Harry Levitt, (1973) “Speech Processing Aids for the Deaf: an overview”,*IEEE, Transactions on audio and Electroacoustics*, Vol.Au-1,No.3.
- [4] Harry Levitt, Member, IEEE,(1971), “Acoustic Analysis of Deaf Speech Using Digital Processing Techniques” *IEEE Fall Electronics Conference*, Chicago, Ill.
- [5] J. M. Picjlett, (1969),“Some Applications of Speech Analysis to Communication Aids for the Deaf”, *IEEE Transactions on Audio Electroacoustics*, AU-17, NO. 4.
- [6] Prashant S. Dikshit', Edward L. Goshom², and Ronald L. Seaman'.(1993), “Differences in fundamental frequency of deaf speech using FFT and Electroglottograph”, *Biomedical Engineering Conference, Proceedings of the Twelfth Southern IEEE*, Page(s): 111 – 113.
- [7] Thomas R. Willemine Francis F. Lee,(1972),Fellow IEEE,“Tactile Pitch Displays for the Deaf”, *IEEE Transaction on Audio and Electroacoustics* VolL. AU-20, No.1.
- [8] Chris J. Clement, Florian J. Koopmans-van Beinum and Louis C. W. Pols,(1996),“Acoustical characteristics of sound production of deaf and normally hearing infants” *Fourth international conference on spoken language*, vol.3, 1549-1552.
- [9] Rabiner et al., (1976), “A Comparative Performance-Study of Several Pitch Detection Algorithms,” *IEEE Transactions on ASSP*, Vol. ASSP-24, No.5.
- [10] Xuejing Sun,(2002), “Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio”, *International conference on Acoustics, Speech and Signal Processing, IEEE. Proceedings. (ICASSP '02)*. Page(s): I-333 - I-336 vol.1

- [11] James W. Pitton, kusnawang, and Bing-Hwang Juang, ,(1996), Fellow IEEE, “Time frequency analysis and auditory modeling for automatic Recognition of speech”, *Proceedings of the IEEE*, Vol.84, no.9.
- [12]] Cherif Adnene,(2000),“Pitch and formants extraction algorithm for speech processing” *The 7th IEEE International Conference on Electronics, Circuits and Systems*, Volume: 1 Digital Object Identifier: Page(s): 595 - 598 vol.1 .
- [13] Alex Acero,(1999), “Formant analysis and synthesis using hidden markov models”.Related website is, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.137.9825>.

C.Jeyalakshmi

Received the B.E degree in Electronics and Communication Engineennng from Regional Engineering College in 2002 and M.E. degree in Communication systems from Saranathan College of Engineering in 2008. Currently she is working as Asst.Professor, in ECE dept., in Trichy Engineering college,Konalai,Trichy and doing Ph.D in the field of Speech recognition of Deaf people in Anna University of Technology,Tiruchirappalli.

