# Voice Aid Design Issues for Hearing Impaired

J. NirmalaDevi

***Abstract---*** *Speech is the primary mode of communication for any human being. Some people like the hearing impaired have various issues in normal communication. Based on their hearing ability these people will have variability in their speech production process. In general, the hearing loss occur from normal to profound. In profound hearing loss the hearing impaired person cannot be able to even produce voice. In order to help and support the hearing impaired persons for leading their normal life, several researches has been made to identify, resolve and reproduce their voice related problems. Some research shows that voice can be reproduced at a minimum range but not completely. Some devices like Alternative and Augmentative Communication (AAC) are also used widely. Many people use Sign language as their mode of communication. All these methods cannot help them to enhance their life style and to communicate without an intermediate person or device. This article represents a novel method to reproduce synthetic voice for the hearing impaired persons from their own voice using a Voice-aid and its design issues.*

***Keywords---*** *Formants, Hearing Impaired, Pitch, Synthetic Speech, Voice-Aid*

## I. Introduction and Background

THE speech signals are collection of sound sequences. These sound sequences provide a symbolic representation of information based on the rules of any language. In order to understand the speech signal it is important to extract the structure of signal where in the information is encoded. The speech is radiated from the sub-glottal system which serves as the source of energy for the speech production [1]. During normal breathing, air passes the larynx and vocal tract gets unobstructed, creating little or no sound. When the vocal tract is opened or totally closed, the airflow is interrupted to create turbulent pulses of air. When the vocal folds are adducted and air is expired, sub glottal air-pressure pushes the vocal folds apart, and air flows rapidly. This immediately creates a partial vacuum between the vocal folds by Bernoulli's principle, which pulls them once again towards each other. It stops the airflow, building pressure again so that the folds again open and thus, a vibratory motion is set in. Figure 1 gives the complete process of normal speech production mechanism.

Language is one of the human communication components which is a system of codes and symbols that are used in communication. Hearing is essential to the acquisition of language and production of speech. Persons with hearing impairments often experience communication disorders because of their hearing loss. Hearing allows any individual to be aware of sounds and thus individual with hearing impairment will have difficulty in speech production [2].

The speech characteristics of Deaf persons is reviewed and discussed by Raymond S.Nickerson in September 1975. The research of Nickerson says that no two individual whether hearing or not can produce exactly same speech [3].

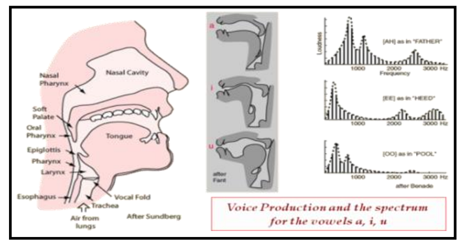


Figure 1: Voice Production Process

### A. Speech of Hearing Impaired

The speech of normal and hearing impaired persons differs in various aspects. Also in general there are major differences in speech between any two individuals.

#### a. Speech Characteristics

The speech characteristics of all the human beings are,

- Timing and Pitch
- Intelligibility and Voice quality
- Articulation and Glottal characteristics

The timing duration and rhythm of hearing impaired speakers are more deviation in the voiced and unvoiced syllables. The deaf speaker have slower speaking rate than normal hearing persons. The nasality affects the quality of the speech and the poor control of the velum brings articulatory problem [4, 5]. The fundamental frequency and formants are used to separate the speech information like stressed and unstressed vowels, consonants and nasals, so as to identify the meaning of the speech signal [6].

Deaf speakers fail to produce pitch variations and they tend to produce abnormally high pitch and lower the pitch, hence the average pitch level increases with the difficulty of the utterances [7, 8].

The voice quality of deaf speakers is determined based on the components like breathiness, nasality and hoarseness which are the dynamic features of deaf speaker. The deaf speakers have transition gestures that change one articulatory position to another. Moreover the voice of the deaf speaker is monotone voice [9].

*The terminology used in this article as Deaf-mute, Deaf speaker and the Hearing impaired person means one and the same.*

*b. Speech Production*

The vocalization development of normal and hearing impaired children was the same, but the number of vocalizations produced by them differs. The hearing impaired children produce minimum number of vocalizations and they stop producing them at their later stage of age development. The hearing impaired children could able to produce a low pitch vowel slower than normal children and they suffer in producing a high pitch vowels, consonants and fricatives. The vocal behaviour cannot be used to identify their sound pattern due to greater variation between any two normal and hearing impaired children [10, 11].

The phoneme production of the hearing impaired children is much slower than the normal. The children tends to apply more number of other phoneme substitutions, then with a single sound and finally with a correct phoneme [12].

The hearing impaired children produce number of errors due to the articulation of consonants, vowels and diphthongs. Some of the typical articulatory errors are given below as investigated [13 – 16]. They are namely,

- Consonant Errors
- Voicing Errors
- Substitution Errors (Place of articulation)
- Substitution Errors (Manner of articulation)
- Omission Errors
- Consonant Cluster Errors

It is difficult to make the spectral measurements of hearing-impaired speech because of the mismatch between spectrograph filters and fundamental frequency [17]. The distinction between the voiced and voiceless segments is complex in hearing impaired children. The children who observe the voiced-voiceless distinction had high speech intelligibility while others do not who cannot observe [18].

*c. Formant Analysis*

The hearing-impaired speakers have difficulty in moving their articulators correctly in phoneme transition. Thus there is often distortion of formant frequency transitions. The changes in formant frequencies like direction, extent and duration of first formant to the second formant transition decides the place of articulation. As discussed above the hearing impaired speakers produce many errors involving the place of articulation. Generally the formants of hearing impaired are extent from their frequency range during transitions and they are short in duration or missing altogether [19-21]. Thus the overall speech production of hearing impaired starts like any normal speaker but the further development and progress is based on the hearing level of the speaker either through amplification or practice.

The pitch quality and intelligibility is poor and deviant at maximum pitch levels and their variance makes tougher to understand the voice of deaf speaker [22]. The investigation of Nico arends, clearly states that the glottal characteristics of the hearing impaired is insufficient for predicting the voice quality and it is extended to explain that the frequency-domain parameters will help to identify the same [23].

The purpose of this article is to demonstrate and understand how the voice characteristics of the normal and hearing impaired persons vary during their voice production process. It is understood at the outset that the synthetic speech can be generated by finding the equivalent text pattern for the stored voice samples of the hearing impaired speakers. It is evident and enough to study the fundamental frequency (F0) which itself prove the variations between the normal and hearing impaired speakers.

*d. Limitations of the Existing Technologies and Methods*

The already existing methods for the hearing impaired speaker communications are

- Lip Reading
- Sign Language
- Gesture Movements
- Writing & Reading

Existing technologies in order to help the hearing impaired speaker for the communications are

- Alternative and Augmentative Communication (AAC)
- Aavaz Device by Invention Labs, IIT Chennai
- Visual Aids

Most of these methods have various limitations and the time and cost factors are relatively high.

1. Lip-reading is difficult at best for many deaf people. Only about 30% of the English language is actually visible on the lips; a skilled lip-reader can bring her comprehension level to 70% with cues and knowledge of the subject matter.
2. When communicating via speech and lip-reading, people tend to repeat themselves if a deaf person doesn't understand what is said. RESTATE rather than repeat. Prior training is needed.
3. Needed additional tools.
4. Less effective and efficient.
5. Need for an alternative in case of failure.
6. Possibility for wrong interpretation while analyzing the inputs.
7. Not accepted and used broadly by the hearing impaired speaker community.

## II. Implementation

*A. Data Acquisition*

The data for the voice aid model is taken from the hearing impaired children. This model is focusing on speaker dependent and it can store any individual speaker inputs for further comparisons.

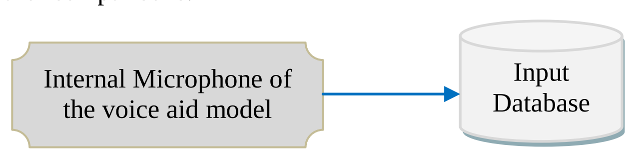


Figure 2: Input for the Voice Aid Model

*B. Various Steps for the Voice Aid Model*

The voice aid model works on various stages in order to produce synthetic speech for the hearing impaired. It focuses on the issues in each and every phase. There are four steps in processing the input voice signal of hearing impaired. They are namely,

a) Parameterization phase
b) Modeling phase
c) Design phase
d) Synthesization phase

*a. Parameterization Phase*

This is the first and foremost phase of the voice aid model. The voice is acquired from the hearing impaired children. Based on the need number of words and samples may be increased. Since this model is for speaker dependent the sample of an individual speaker is taken. The input word is analyzed in order to recognize the text pattern in the later stage. Each and every time the sample is added with the already recognized samples.

This is called parameterization phase since the various parameters of each word acquired is analyzed to differentiate it from the next word. The speech characteristics of the hearing impaired as discussed in the literature are varying from one another. The parameters taken as an first phase analysis here are

- Pitch or Fundamental frequency (F0)
- Formants

The database stored is used for offline analysis and during the regular use of the voice model the word is continuously streamed into the environment for online processing. The Figure 3 represents the model of parameterization phase.

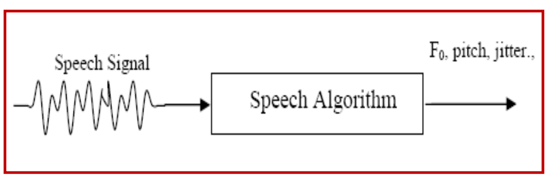


Figure 3: Parameterization Phase

*b. Modeling Phase*

The parameters of the first phase are taken every time for registration. Since the hearing impaired speaker cannot able to produce the word with the already generated utterances it is necessary to register the word each time into the database. The concept applied for registration and the database training is neural networks.

Every word of the hearing impaired is measured and registered using extracted feature value before storing in the database as shown in the Figure 4. The feature value is a numerical measure and that correlates approximately to word samples. Hence the database has collection of hearing impaired word samples along with the respective feature values.

Also when the input word is received from the parameterization phase the modeling phase compares the already stored feature value of the existing words in order to avoid the duplication. Once the word is analyzed its equivalent text content is reproduced. This will be done before the voice aid model is implemented in practice. Only trained model can produce efficient results.

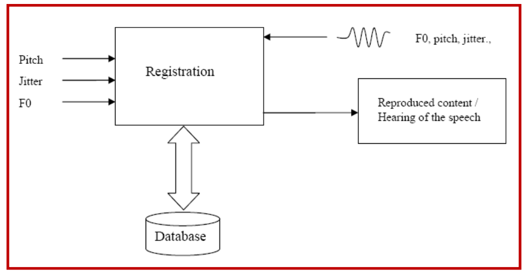


Figure 4: Modeling Phase

If a new word is produced by the hearing impaired speaker it is received as a new input and it will be automatically updated in the database along with its extracted feature value. Words produced by the hearing impaired are measured and registered using extracted feature value before storing in the database. The feature value is a numerical measure and that correlates approximately to stored reproduced text content samples. Hence the database has collection of hearing impaired word samples along with the respective feature values.

*c. Design Phase*

The reproduced text content is validated in the design phase. This phase concentrates to produce an equivalent text for the reproduced content from the modeling phase. The importance of this phase is it has to check whether the reproduced content is a valid text content of the database.

This means if a new word is used for this type of voice model it cannot reproduce the exact word instead it may create an equivalent word as matched.

If the feature value exactly matches the database by using reproduced speech (RS) it creates an equivalent text to text output (ETTO) as given in Figure 5. It is then used to reproduce the synthetic speech in the next phase. If any mismatch occurs then approximate or nearest feature value is taken into account.

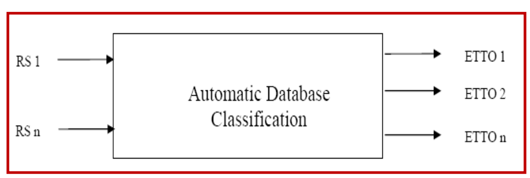


Figure 5: Design Phase

*d. Synthesization phase*

The words produced by the hearing impaired speaker differ from normally phonated speech in terms of voicing, pitch and formant structure. There is no perceived pitch period in hearing impaired and the voice is definitely noisy. This makes difficult to synthesize the new text pattern for the hearing impaired directly from their source signal.

Hence the previous stages are used to analyze the words for finding the text pattern which makes easier to generate the synthetic speech from other stored synthetic voices.

In order to produce the synthetic speech the ETTO is directly given to a speech synthesizer which in turn generates a synthetic speech output (SO) as mentioned in the Figure 6. The generated synthetic speech can be made use by any electronic gadgets for real time application.

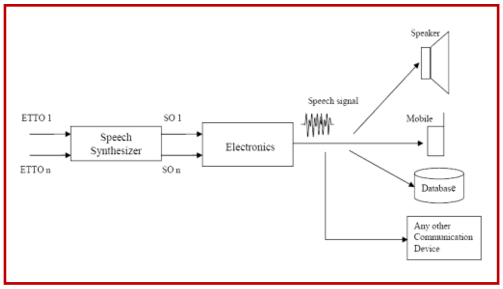


Figure 6: Synthesization Phase

## III. Result and Discussion

The results are discussed with few samples of both the hearing impaired speaker and the normal speaker. Also the results describes about pitch, formants and the spectrogram.

Here the analysis clearly explains about the issues in generating a synthetic speech from the source signal of the hearing impaired speaker.

### *i. Pitch*

The pitch of the normal and hearing impaired speaker is compared in the following Table 1 which shows that there is no relevance between the pitch periods among any two speakers and there is more deviation shown for the words produced by both the normal and hearing impaired speakers.

The sample subjects are asked to produce 5 words in Tamil as given in the result table by two speakers of both the groups.

Table 1: Comparison of Pitch between Normal and Hearing Impaired Speaker

| Tamil words | Normal Speaker 1 | Normal Speaker 2 | Hearing Impaired Speaker 1 | Hearing Impaired Speaker 2 |
|---|---|---|---|---|
| Amma | 285.816 | 261.703 | 393.14 | 296.88 |
| Appa | 271.542 | 272.362 | 341.37 | 281.13 |
| Athai | 271.930 | 267.577 | Undefined | 278.01 |
| Uppu | 278.883 | 260.244 | 390.00 | 374.34 |
| Iyyo | 289.547 | 269.781 | 239.71 | 485.16 |

A sample figure is shown between these two groups in order to demonstrate that the value does not correlate each other. The Figure 7 is for normal speaker and Figure 8 is for Hearing impaired speaker which shows deviation in the pitch periods. Finally it is also concluded that the fundamental frequency is the foremost parameter of any speech signal which helps to identify the phoneme classification and the similarity between any two samples.

Hence the pitch parameter cannot be used as an important parameter for the speech Synthesization for the hearing impaired speakers.

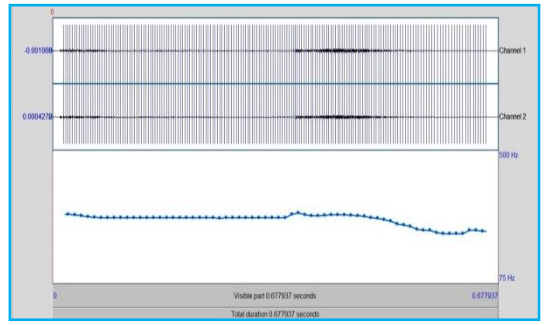


Figure 7: Pitch Display of Normal Speaker

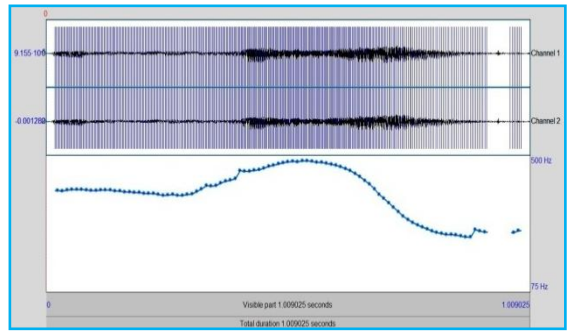


Figure 8: Pitch Display for Hearing Impaired Speaker

### *ii. Formants*

Formants are also an important parameter for the speech signal. In the results it is analyzed that the normal speaker formants have at least a relation between F1 and F2. But this have a basic deviation from normal to hearing impaired and also the location of the second formants F2 in hearing impaired speaker have more deviation and hence formants may be used to differentiate between any two words but not for reorganization purpose.

The following Table 2 shows a comparison between the formants of a normal and hearing impaired speaker. The sample subjects are asked to produce 5 words in Tamil as given in the result table for formants F1 by two speakers of both the groups.

Table 2: Comparison of Formants F1 between Normal and Hearing Impaired Speaker

| Tamil words | Normal Speaker 1 | Normal Speaker 2 | Hearing Impaired Speaker 1 | Hearing Impaired Speaker 2 |
|---|---|---|---|---|
| Amma | 937.08 | 370.64 | 706.54 | 430.56 |
| Appa | 798.5 | 807.45 | 776.78 | 738.849 |
| Athai | 806.13 | 771.54 | 1009.67 | 739.694 |
| Uppu | 530.953 | 278.81 | 425.69 | 454.04 |
| Iyyo | 634.450 | 541.21 | 545.65 | 732.83 |

A sample figure is shown between the groups to understand the dissimilarity between the formants position. In the Figure 9 (Normal Speaker) and Figure 10 (Hearing impaired Speaker) the formants are shown for F1, F2 and so on.

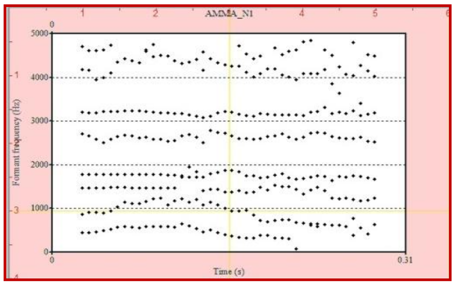


Figure 9: Formants for Normal Speaker

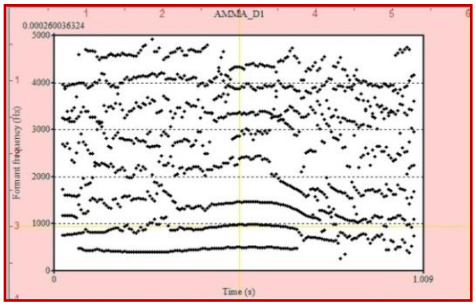


Figure 10: Formants for Hearing Impaired Speaker

*iii. Spectrogram Analysis*

A spectrogram is an image that shows how the spectral density of a signal varies with time. Spectrograms are used to identify phonetic sounds, to analyze the speech processing. From the spectrogram the pitch frequency of the normal and hearing impaired speech is clearly identified and the formant locations and bandwidths of hearing impaired speech differ from normally phonated speech is also seen.

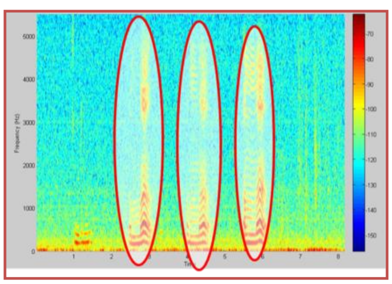

Figure 11: Spectrogram of Normal Speaker

The above Figure 11 is the result of a normal speaker where the energy distribution of the vocal cord is normal and in its proper position. Spectrogram graph is used to differentiate the vocal energy, stressed and unstressed vowels. Also from the figure it can easily visualized that how and where the deviation occurred. Hence based on the given results of spectrogram for normal and hearing impaired speaker it is understood that the energy and stress level of these two groups differ and exactly in hearing impaired speaker the energy and stress level is very high for the speech production.

The Figure 12 spectrogram clearly indicates that the energy distribution of the vocal card is heavy and it is difficult to point out the frequency of the voice produced in the hearing impaired speaker. This can be used to even differentiate the phoneme of each word which will help to identify the similar word patterns from one another.

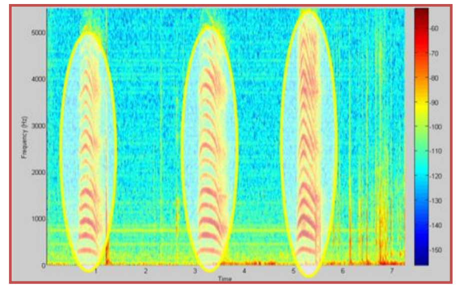

Figure 12: Spectrogram of Hearing Impaired Speaker

## IV. Conclusion and Future Enhancement

Even though the result of the synthetic speech reproduction is not accurate the signal can produce an equivalent voice for the hearing impaired speaker. This will be helpful for them to lead a normal life in the society. Also, this will increase the confidence level and support for them when dealing important or emergency situations without additional support. Since the above said parameters like pitch and formants cannot be used for identifying the difference between words of a single hearing impaired speaker some other parameters may be used to differentiate the speaker words for larger utterances.

The major issue of the voice aid model discussed is the comparison of word utterances and the parameter identification. The voice aid if supported with speaker recognition features it might reduce the burden in reproducing the synthetic speech. Also most of the existing research focuses on the Linear Predictive Coding (LPC) method for pitch detection and analysis.

There is also a possibility to synthesize the pitch periods and the formants location using LPC. If that is made possible then the Synthesization will be possible for the source speech signal of the hearing impaired.

Voice disorders are not life threatening. However daily issues encountered by an individual like hearing impaired should not be underestimated. This method provides an excellent and comprehensive framework for generating speech for individuals with voice disorders. The application of this proposed method is used for a normal communication by a hearing impaired individual for whom it will be essential to achieve the ultimate goal for enhancing the quality of life of the individual.

## References

[1] L.R. Rabiner, R.W.Schafer, "Digital Processing of Speech Signals", Pearson Education, Third Edition, Pp. 55, 2009.
[2] Carol. A. Angell, "Language Development and Disorders, A Case Study Approach", Jones & Bartlett, First Edition, Pp. 1 & 89, 2010.
[3] Raymond S. Nickerson, "Characteristics of the Speech of Deaf Persons", The Volta Review, Pp. 342-362, 1975.
[4] Angelocci, A. A, "Some observations on the speech of the deaf", The Volta Review, Pp. 403-405, 1962.
[5] R.S. Nickerson, K.N. Stevens, A. Boothroyd, & A. Rollins, "Some observations on timing in the speech of deaf and hearing speakers", BBN Report No. 2905, September 1974.

[6] G. Fairbanks, "Recent experimental investigations of vocal pitch in speech", Journal of the Acoustical Society of America, Vol. 11, Pp. 457-466, 1940.
[7] J. Martony, "On the correction of the voice pitch level for the severely hard of hearing subjects", American Annals of the Deaf, Pp. 195-202, 1968.
[8] T.R. Williemain. & F.F. Lee "Tactile pitch feedback for deaf speakers", The Volta Review, Pp. 541-554, 1971.
[9] D.R. Calvert, "Speech sound duration and the surd-sonant error", The Volta Review, Pp. 401-403, 1962.
[10] R.E. Stark, "Vocalizations of the pre-school deaf child", Progress report, Baltimore, The Johns Hopkins School of Medicine, 1967.
[11] M.P. Mavilya, "Spontaneous vocalization and babbling in hearing impaired infants", International symposium on speech communication abilities and profound deafness", Washington, Bell Association, 1968.
[12] C. Stoel-Gammon, "The acquisition of segmental phonology by normal-hearing and hearing-impaired children", Speech of hearing impaired: Research, training, and personnel preparation", University park press.
[13] C.V. Hudgins & F.C. Numbers, "An investigation of the intelligibility of the speech of the deaf", Genetic Psychology Monographs", Pp-289-392, 1942.
[14] A. Markides, "The Speech of deaf and partially hearing children with special reference to factors affecting intelligibility", British Journal of Disorders of Communication", Pp-126-140, 1970
[15] C.R. Smith, "Residual hearing and Speech production in deaf children, "Journal of Speech and Hearing research, Pp-795-811, 1975.
[16] N.S. McGarr, "Evaluation of speech in intermediate school-aged deaf children", Special assessment and speech improvement for the hearing impaired, Alexander Graham Bell Association, Pp-45-66, 1980.
[17] A.W.F. Huggins, "Speech timing and Intelligibility", Aattention and performance VII, Lawrance Erlbaum Associates, Pp-279-298, 1978.
[18] R.B. Monsen, "The production of English stop consonants in the speech of deaf children", Journal of Phonetics, Pp-29-42, 1976.
[19] J. Martony, "Studeies on speech of the deaf", Quartely progress and status report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden, Pp-16-24, 1966.
[20] R.B. Monsen, "Second formant transitions of selected consonant-vowel combinations in the speech of deaf and normal-hearing children", Journal of Speech and Hearing Research, Pp-279-289, 1976.
[21] H. Roathman, "A Spectrographic investigation of consonant-vowel transitions in the speech of deaf adults", Journal of Phonetics, Pp-129-136, 1976.
[22] Nancy S. McGarr & Mary Joe Osberger, "Pitch deviancy and intelligibility of Deaf speech", Journal of Communication disoreders, Pp-237-247, 1978.
[23] Nico Arends (et al.,), "Predicting voice quality of deaf speakers on the basis of glottal characteristics", Journal of speech and hearing research, Pp-116-122, March 1990.

**Nirmala Devi. J** The author was born in Tiruchengode, Namakkal (Dt) on March 25th 1981. She got both her UG graduation in B.E Information technology and PG graduation in M.E Computer Science and Engineering from K.S.R College of Technology, Tiruchengode on 2003 and 2008 respectively. She secured 35th University Rank for her PG and secured 35th Rank from Anna University of Technology, Chennai. She is currently pursuing her Ph.D in the area of Speech Processing. She has a total of 6 years teaching experience in various colleges like Paavai Engineering College, Sengunthar Engineering College for Women and presently she is working as an Associate Professor and Head in the Department of Information Technology, Excel Engineering College, Namakkal. She is a Life member of ISTE. She has few publications in various National and International Conferences. Her areas of research interests include Speech Signal Processing, Telecommunication and Networking. (E-mail:jvandana25@gmail.com)