TELEPHONE SPEECH ENHANCEMENT FOR THE HEARING IMPAIRED

BONY TASNIM
ID: 03101050
C. Z. MURSHED
ID: 05310053

Department of Computer Science & Engineering

January 2009

BRAC UNIVERSITY, Dhaka, Bangladesh

TELEPHONE SPEECH ENHANCEMENT FOR THE HEARING IMPAIRED

THESIS

Submitted to the Department of CSE

BRAC University

By

BONY TASNIM
ID: 03101050
C. Z. MURSHED
ID: 05310053

Supervised by
Dr. AKM Abdul Malek Azad
Assistant Professor
Dept. of CSE, BRAC University

In Partial Fulfillment of the

Requirements for the Degree

Of

Bachelor of Science in Computer Science & Engineering

BRAC University

January 2009

DECLARATION

We hereby declare that this thesis is based on our explorations and investigations and the results found by ourselves. Materials of work found by other researcher are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Signature of	Signature of
Supervisor	Authors

Acknowledgement

First of all we would like to specially thank our supervisor **Dr. AKM Abdul Malek Azad** for his precious guidance and support. His helpful suggestions for our thesis work and various useful links to collect the required resources for our work has helped us to make our thesis paper a great success.

We would also like to specially thank **Engr. Nihad Sultana** (Head of Sales, Medical Solution Division, Siemens Bangladesh), **Dr. Satya Prasad Majumder** (Professor, Dept. of EEE, BUET) and **Dr. Md. Aynal Haque** (Professor, Biomedical Engineering, BUET), **F. M. Abdul Hafiz** (Manager, Service Centre, Siemens Mobiles, Siemens Bangladesh), **Md. Abul Hossein Khan** (Chief Audiologist and Hearing Aid Specialist) for their suggestions, support and links leading us to the collection of our required resources for our thesis.

Thanks also go to our Lab Technical officers, **Ahmed Al Amin** and **Asaduzzaman All Faruk**, in our Electronics and Communication Labs, where we could work with the DSP trainer kit, PLD trainer kit and other devices to carry on with the different experiments of our thesis work.

We would also like to take this opportunity to thank **School for Assisting Hearing Impaired Children (SAHIC)** and **Siemens Bangladesh** for their resource support and **Bangladesh University of Engineering and Technology (BUET)** for providing us their academic and library facilities.

Telephone Speech Enhancement for the Hearing Impaired

Bony Tasnim & C. Z. Murshed

CSE Dept., BRAC University, 2009

Supervisor: Dr. AKM Abdul Malek Azad

This thesis explores the problems hearing impaired people encounters using telephone and hence puts efforts to enhance the telephone speech quality. There are two aspects of this work: extension of the low band signal for the deaf and improving the classification and matching method in the speech coding process. Study of telephone bandwidth extension further provides the opportunity to design extension module for the low band signals for natural hearing. Together with the high band extension module this will retrieve lost quality and information. The speech frame classification technique and matching method based on fuzzy logic is proposed to overcome the drawbacks of conventional speech coding process. Fuzzified classification has the potential to increase the performance and the accuracy rate of the estimation from the narrow band telephone speech. Aspects presented here widen future implementation and tests of the proposed models.

Table of Content

List of Figures

Abbreviations

ALD – Assistive Listening Devices

BWE – Bandwidth Extension

CI – Cochlear Implant

CMM – Codebook Mapping Method

EM – Electromagnetic

FM – Frequency Modulator

FVD – Fuzzy Voicing Detector

FVQ – Fuzzy Vector Quantization

GMM – Gaussian Matrix Matching

HMM – Hidden Markov Model

LPC – Linear Prediction Coefficient

LSF – Line Spectral Frequency

PLP – Perceptual Linear Prediction

VAD – Voice Activity Detector

VDA – Voice Detection Algorithm

Chapter one

Introduction

About 40 million mild to severe deaf people have telephone conversations essentially. Often they have difficulty understanding the telephone speech. The intelligibility of telephone speech is much lower than that of person–to–person speech. This degradation results from the following major factors:

1. Loss of quality and information. Telephone speech is band-limited to 300Hz - 3400Hz. The spectrum above 3.4kHz, present primarily in fricative consonants such as 's' 'sh' 'ts' etc., is lost. This results in the muffling effect of telephone sound, which does not affect normal-hearing people, but greatly affects the hearing impaired. The spectrum below 300Hz contains the speech naturalness and quality. Loss of this band results discomfort hearing the telephone speech.

2. Additional noise introduced by the interaction between the phone and the hearing aid. The electromagnetic coupling effect of the phone-handset circuit and the hearing aid coil results in the feedback and amplification of background noise. A cell phone's electromagnetic emission is often picked up by a hearing instrument as a buzzing noise. Also, different hearing devices are not compatible with certain kinds of cell phones.

3. Lack of visual cues. In a person-to-person conversation, a hearing aid or CI user often uses lip-reading or other visible cues to help

understanding the other person. When talking on the phone, the audio signal is the only information source that he can make use of.

To address the first problem, bandwidth extension of the narrowband telephone speech is essential. To do the required extension to a wider band speech several algorithms are available. The high band (3400-8000Hz) extension is already developed for the hearing impaired. In this thesis, we have evaluated existing bandwidth extension techniques and proposed a low band (50-300Hz) extension module for qualitative and natural hearing experience by the hearing disabled people.

For the second problem, different phone adapters in the form of assistive listening device are available in the markets. Different categories of ALDs and phone adapters are briefly evaluated in this thesis.

And finally, with recovered lost speech components, natural voice, reduced noise and eliminated EM coupling effects, the need for visual cues are greatly decreased.

This thesis is organized as follows: Chapter Two is a review on current telephone features and devices available and bandwidth extension methods that have been developed; Chapter Three provides the basic concepts of speech generation, artificial speech generation and hearing process and disabilities; Chapter Four proposes a extension module for the low band extension for the deaf; Chapter Five explains the need for appropriate classifier and mapping system and proposes a fuzzy logic based speech classification and matching method; Chapter Six concludes the thesis with future aspects.

Chapter Two

Literature Review

This chapter is written as a summary of the exploration done in the fields of telephone usability of the deaf, assistive devices, telephone adapter, bandwidth extension and speech coding.

Hearing aid and other hearing technologies are successfully being used by hearing impaired. Efforts have been taken to improve the telephone speech intelligibility for the deaf using both intellectual methods and assistive devices. Wireless telephone adaptors are designed to solve the electromagnetic coupling affects.

Review on existing bandwidth extension algorithms and speech coding methods is provided in later sections.

Sections 2.1, 2.2, 2.3, 2.4 below provide reviews on deaf users' telephone usability, assistive listening devices, telephone adapters, and bandwidth extension techniques respectively.

2.1 Telephone Usability

The ability to carry a conversation over telephone without any visual cues has been an important indicator of the life quality of the hearing impaired. As technology prevails, telephone usage has essentially become a part of life for most of them. Considering their needs, recently impressive development and extent of telephone options for the hearing impaired can be found.

As per some survey results [8], a very good number of deaf respond to telephone, and more than 51% cochlear implants initiate calls. But very few words they could hear correctly and the percentage doubled up when listening twice. Only a quarter of the patients in a deaf institute had a significant degree of telephone ability and more than a half of them were good indicator of telephone competence. The statistics differed from adult to prelingually deaf children and from male to female. Also they were found to be more telephone usable with familiar callers than unfamiliar callers.

Dealing conversations with unfamiliar callers about unfamiliar topics, they reported the telephone speech quality as weak, hollow, tiny, having echo, fuzzy or distorted in other ways. Background noise was a problem to 94% of them and 76% of them thought the speech was very soft. Lack of clarity was reported by 66% of the users, and this could not be solved by any amplification.

Compatibility between hearing instruments or CIs and the cellular phones is also found to be poor and confusing. The broad spectrum radio signal is like a noise to them. Only few expensive cellular phones are compatible with hearing devices having telecoils or adapters. Wireless adapters and cell phones with wireless connectivity are very sophisticated and expensive.

2.2 Assistive Listening Devices

Assistive listening devices (ALDs) are types of hearing assistance technology, and aim at improving the quality of life of the deaf. ALDs are designed to pick up audio signals from the desired source and minimize the noise and interferences. Different environments have different needs and therefore different ALD applications have different designs. These are the major ALDs contributing to the daily life of the deaf and their telephone usage, as well [1], [8]:

- Hardwired Devices; free of electronic interferences but losses mobility.
    - Telecoil enabled adapters
    - TV listening systems
- Induction Loop Devices; easy installation, but vulnerable to EM interference.
- Telephone ALDs; effective to some extent.
    - Amplified telephones
    - Text telephone systems (TTYs)
- Vibrating Devices; alarm system attached to body for incoming calls
- Wireless Devices; mobile, covers large area
    - Infrared light devices- directional, has privacy, but complicated and expensive
    - FM receivers- good for conferences as no privacy
    - Bluetooth receivers- secure private use, medium range

The Bluetooth receiver meets the most effective features to solve the second problem in the telephone usability of hearing impaired (discussed in following section).

2.3 Phone Adapters

A very specific approach to address the electromagnetic coupling affect and buzzing noise problem encountered due to the close interaction of the phone handset and hearing instrument was discussed in [8]. A wireless phone adapter based on Bluetooth technology was proposed and the implementation was tested by people with hearing loss.

There are other types of phone adapters which can solve the problem. The telecoil enabled hearing devices, when switched to 'T' position and the audio jack is inserted, routes the signal to the device from the phone. This option confines the user to the wire length. Personal FM receivers are also usable with large area coverage, but as the signal is broadcasted anyone tuning the same frequency can hear the conversation. The use of Bluetooth adapter provides the best solution.

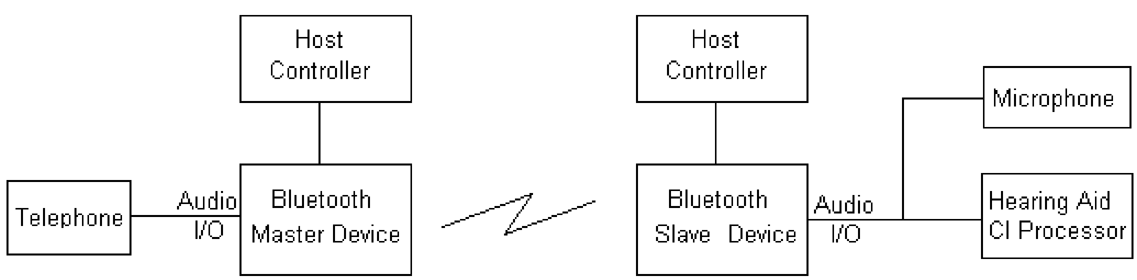


Figure 2.1: A Bluetooth based phone adapter.

A Bluetooth device that initiates a connection is called the master device and the other one is slave device. Both the devices are connected to host controllers where commands are set. After connecting as pairs both the devices communicate in a full duplex mode. The telephone speech will be wirelessly routed directly to the hearing instrument which the adapter is connected. The security and compatibility is defined in its protocol stacks and can be permitted or blocked by pairing choices by the master device.

Using Bluetooth adapter the user can talk about 100 meters around the base phone. This most popular technology covers a variety of end devices to be compatible with the adapter. This personal ALD can automatically find the local Bluetooth enabled devices to communicate through. These adapters are available in market now though little expensive yet.

2.4 Speech Enhancement by Bandwidth Extension

Human speech is band limited to 50-8000Hz, but in telephone systems this bandwidth is not provided. Today's telephone networks still provide poor audio quality due to historical limitations. The reason for the poor audio quality of the telephone network is the very limited bandwidth that is provided. Analog networks, for example, provide only a bandwidth of about 3.1 kHz (0.3-3.4kHz). This leads to reduced speech quality and even intelligibility.

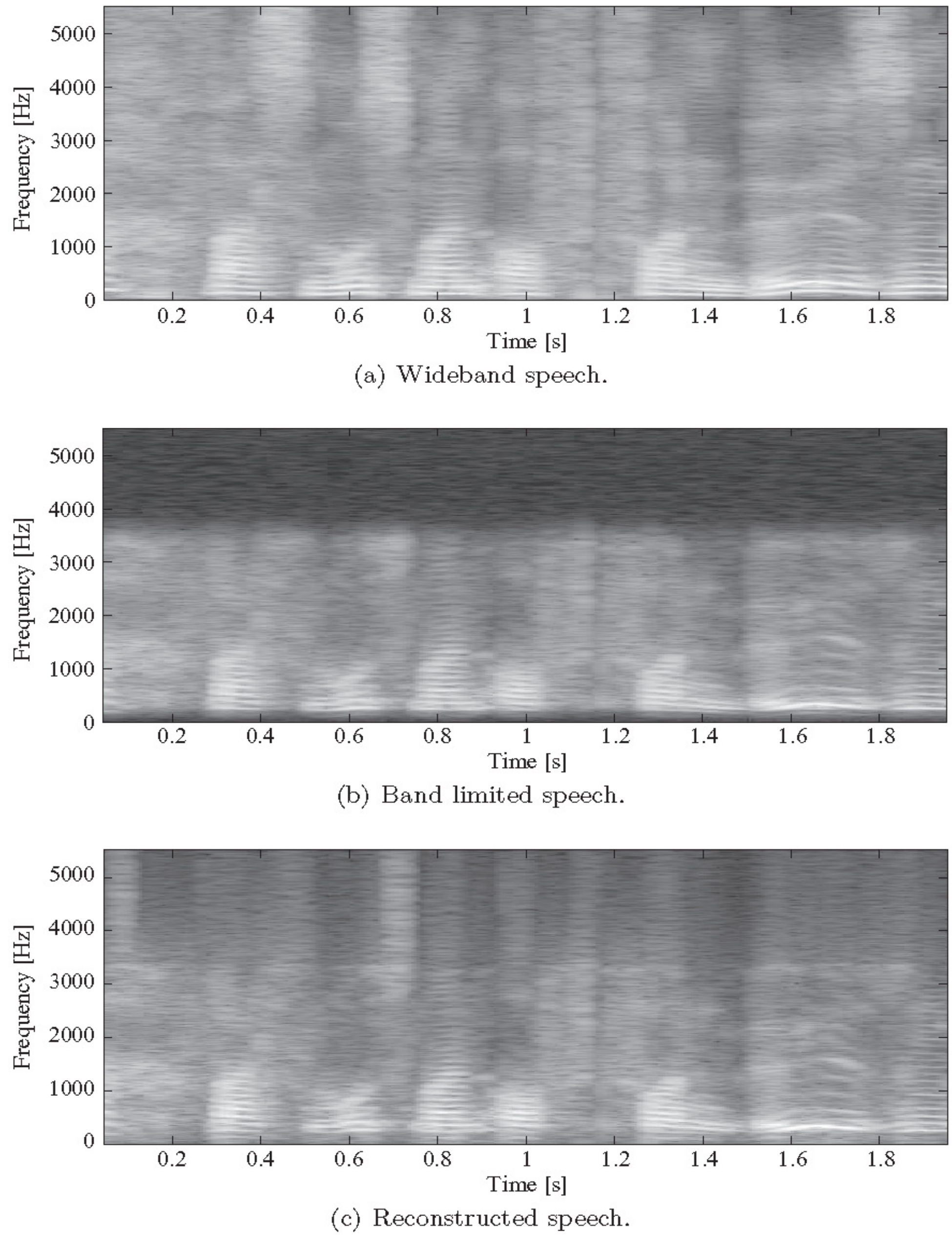
The loss of information [Fig.2.2] in 50-300Hz and 3400-8000Hz causes a muffled effect and degraded quality. The narrow band telephone speech is regular to normal people, for them the wideband speech is just more clear and natural. They survive for the redundancy nature of human speech. But for the hearing impaired, the loss of information due to narrowed telephone speech is the main reason for the difficulties in telephone conversations.

A typical property of analog networks is the difficulty of distinguishing between several fricatives like present in the words “feel” ([fi:l]) and “veal” ([vi:l]).

Another typical problem is that one is not able to distinguish similar voices (father–son problem) over the telephone.

To solve this problem telephone speech could be widen to larger band but it would increase crosstalk within the conventional 8 kHz sampling rate. So actually the sampling rate needed to be increased. But the telephone network is one of the most widespread networks all over the world. So rather than changing the network, the idea of end device modification is much more effective. The telephone speech enhancement is hence done by the bandwidth extension of the received narrow band speech.

Bandwidth extension in this context means the estimation of the not transmitted frequency components out of the transmitted signal by exploiting the transmitted information included in speech signals and therewith increasing the speech quality. This approach yields the advantage that nothing has to be changed within the network – it is simply an optional feature on the terminal side. And this feature is an obvious for the hearing impaired to properly communicate over the telephone.



(a) Wideband speech.

(b) Band limited speech.

(c) Reconstructed speech.

Figure 2.2: The original wideband speech, narrowband telephone speech and reconstructed speech.

2.4.1 Basics of Bandwidth Extension

Based on the source-filter model for the speech production process as discussed earlier an intuitive approach for bandwidth extension algorithms can be established. The basic scheme for these algorithms is depicted in *Figure 2.3*. Starting from a band limited speech signal *s*nb(*n*) and following the replication of the source path the narrowband excitation (also known as residual signal) signal *e*nb(*n*) is estimated. Afterwards the broadband excitation signal ˆ*e*bb(*n*) is generated out of the narrowband excitation signal. In the next step the spectrally flat estimation of the broadband excitation signal is colored by the application of the estimated broadband spectral envelope resulting from the coefficients

$$\hat{\mathbf{a}}_{\mathrm{bb}}(n) = [\hat{a}_0^{\mathrm{bb}}(n), \hat{a}_1^{\mathrm{bb}}(n), \ldots, \hat{a}_{N_{\mathrm{bb}}-1}^{\mathrm{bb}}(n)]^{\mathrm{T}},$$

where *N*bb represents the order of the broadband all-pole filter. These coefficients are estimated in the replication of the filter path that also starts from the band limited speech signal *s*nb(*n*). Then the narrowband spectral envelope or the representing coefficients

$$\mathbf{a}_{\mathrm{nb}}(n) = [a_0^{\mathrm{nb}}(n), a_1^{\mathrm{nb}}(n), \ldots, a_{N_{\mathrm{nb}}-1}^{\mathrm{nb}}(n)]^{\mathrm{T}}$$

respectively are estimated. Analogous to above *N*nb denotes the narrowband all-pole filter order. These coefficients are useful for the above mentioned estimation of the narrowband excitation signal. Furthermore these coefficients are required for the estimation of the broadband spectral envelope ˆabb(*n*). After the coloration of the estimated broadband excitation signal ˆ*e*bb(*n*) with the estimated broadband spectral envelope we have a completely synthesized speech signal. Since the original signal within the telephone band is present, a band stop filter is used to get rid of the redundant frequency components. Finally, the complementary frequency components are combined within the summation unit resulting in an artificially supplemented and thereby quality improved telephone speech signal.

The usage of the source-filter model in an approach for bandwidth extension of speech signals is motivated by its extensive use and success in the field of speech coding.

The bandwidth extension process discussed so far is a very general view. Considering telephone speech (300-3400 Hz) and natural speech frequencies (50-8000 Hz), the process can be used for both low band (50-300 Hz) and high band (3400-8000 Hz) extension of band limited telephone speech, which is further discussed later.

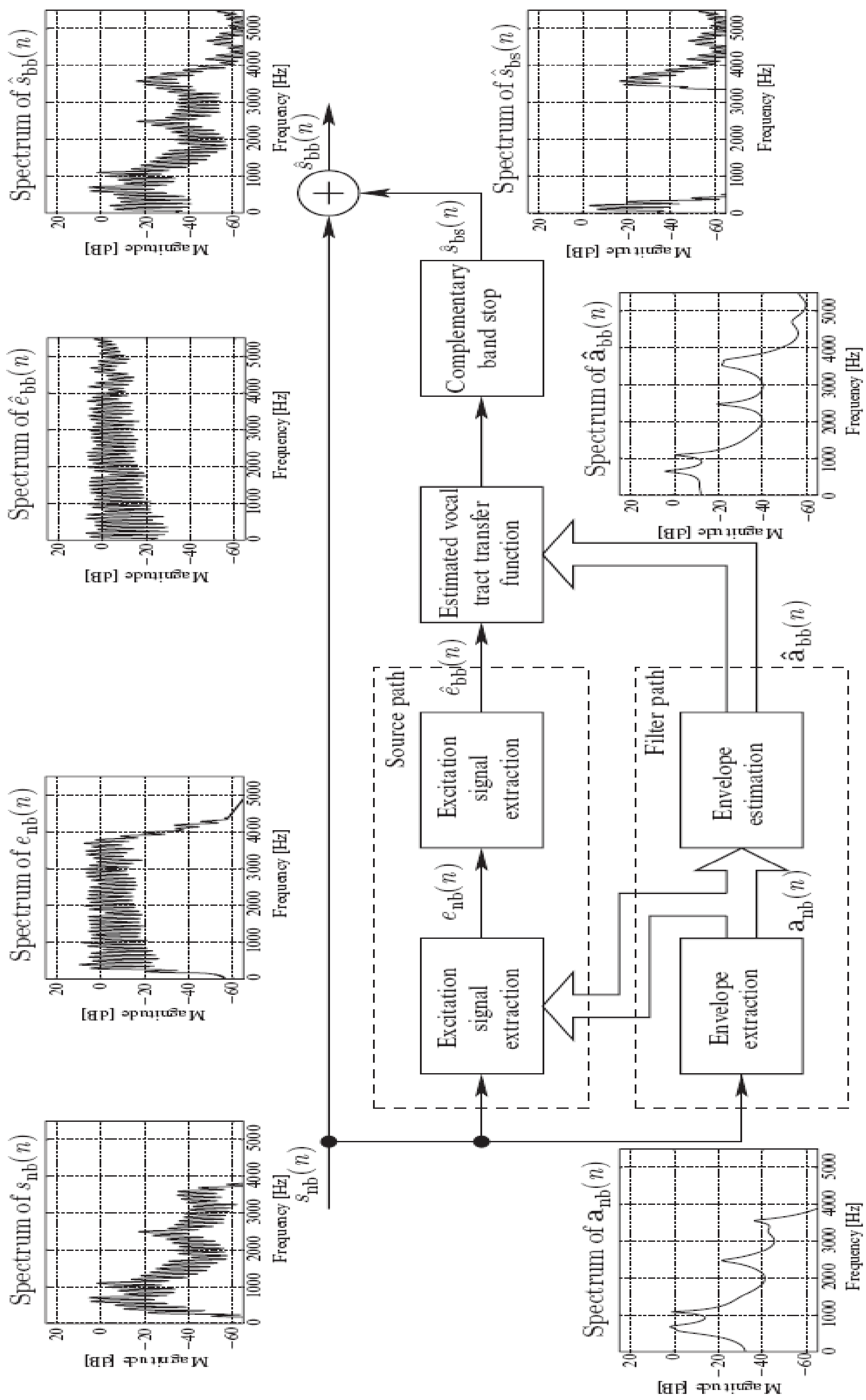


Figure 2.3: Basic concept of BWE

2.4.2 Residual Extension

The residual signal has a flat spectrum like white noise. In a voiced frame, such as vowels and semi-vowel consonants, the residual noise has periodicity. This appears as harmonic peaks in addition to the flat noise-like spectrum. These peaks occur in multiples of the pitch, the fundamental voice frequency of the speaker.

Therefore, the task of the residual extension module is to double the sampling rate, from 8 kHz to 16 kHz, while keeping the whole spectrum flat. If there are harmonics in the narrow-band residual, the wide-band residual should also have the harmonic structure. There are two methods in common use that accomplish that:

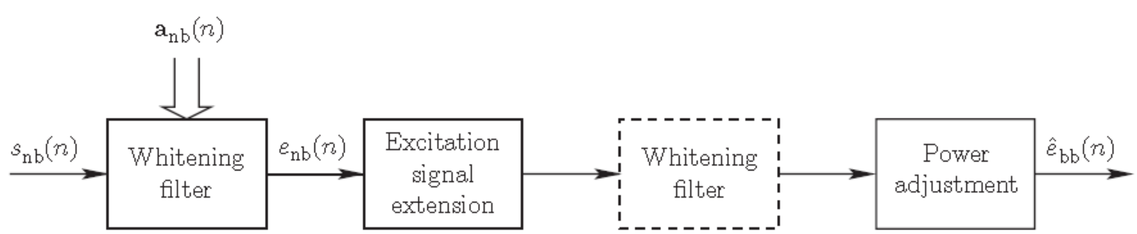


Figure 2.4: Overall system for residual extension

a. **Spectrum folding method:** This time-domain method proposed in [5] is easy to implement. The up sampling of the narrow-band residual is done by inserting zeros. This is equivalent to folding the spectrum of 0-4000Hz to 4000-8000Hz in the frequency domain. Since the low-frequency spectrum is flat and has harmonics, the resulting wide-band residual will also have a flat spectrum and harmonics in both the low-frequency part and the high-frequency part. One drawback of this method is that the harmonic structure is broken at 4 kHz. A possible solution is to change the sampling rate to a multiple of the pitch before performing the folding, but this requires accurate pitch detection. Another disadvantage is that harmonics in a real wide-band residual should have descending amplitudes, but in the folding method, the harmonics at highest frequencies are the reflection of the harmonics at lowest frequencies and therefore have the same amplitudes.

b. **Nonlinear distortion method.** As is shown in figure below, the narrow-band residual is first up sampled by interpolation and then fed into a nonlinear function. The distorted signal will have the desired bandwidth and harmonic structure over the whole spectrum. After the whitening filter, the spectrum is flattened and the wide-band residual is achieved. A popular nonlinear function is given below:

$$y(t) = [(1+\alpha)|x(t)| + (1-\alpha)x(t)]/2$$

where x(t) is the input signal, y(t) is the distorted output signal, and 'alpha' is a parameter between 0 and 1 [5]. When □=1, it becomes the absolute value function, which is used in [20] and achieves good results. Non-linear functions have been used in bandwidth extension algorithms mainly for the generation of low-frequency speech components to date [10].

2.4.3 Envelope Extension

The envelope extension method mainly consists of extracting the envelope from the narrowband speech signal and extending it. There are mainly two methods to do it.

a. **Linear Estimation Method:** The basic idea of linear estimation lies in the following linear equation:

$$\vec{y} = M \cdot \vec{x}$$

Vector *x_*, the set of parameters representing the narrow-band spectral envelope, is first extracted from an input signal frame. Then the corresponding vector *y_* representing the wide-band envelope is calculated by feeding *x_* into a group of linear filters. M is the matrix composed of filter parameters. Then the output spectra envelope is generated based on vector *y_*.

In general linear estimation method is used because it requires less memory and computation. One disadvantage is that the solution may yield invalid values representing a LPC filter with unstable impulse response. Therefore special adjustments have to be added to avoid invalid results, and this might introduce artifacts. Also, since we try to use a linear model to describe the nonlinear relation of narrow-band and wide-band parameters, a certain degree of distortion can be expected.

b. Codebook Method

Codebook mapping is a popular method to achieve spectral envelope extension [12][14][16]. For this application, the codebook consists of two columns as shown in the diagram below. The first column contains vectors composed of spectral parameters extracted from the narrowband signal, while the second column contains vectors extracted from the corresponding wide-band signal. When an input frame comes in, the parameters are extracted from it and compared with the vectors in the first column. By vector quantization, the vector closest to the input parameters is found, and the corresponding wide-band parameters are taken from the second column to generate the extended spectral envelope.

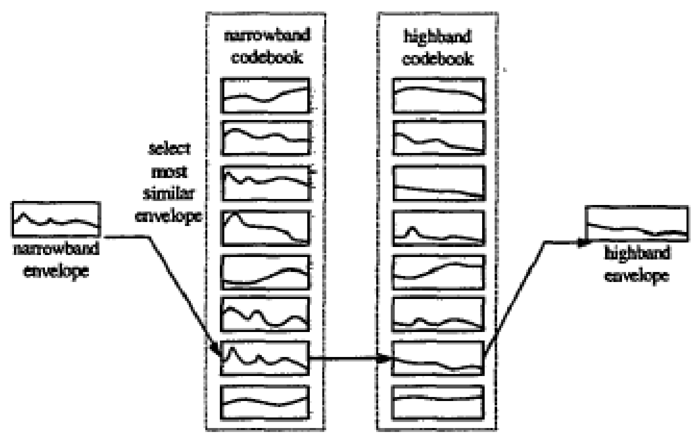


Figure 2.5: Codebook mapping method

The codebook is generated from a large training database of speech.

An obvious limitation of codebook mapping method is that the number of possible outputs is decided by the codebook size. Also, when the parameters of two narrow-band frames belong to the same group in vector quantization, the corresponding wide-band frames do not necessarily belong to the same group. The probability of such mismatches increases when the size of codebook increases.

To address the above problems, improved versions of codebook mapping were proposed:

1. Codebook plus interpolation**.** When a set of input parameters comes in, a number of closest codebook items are found. The output is computed as the weighted sum of the corresponding wide-band parameter-vectors, based on a certain statistical model [12].

2. Multiple codebooks. Speech frames are classified into several groups, and one codebook is trained and used separately for each group. In [14], two codebooks were trained and used for voiced and unvoiced frames, and the performance was found to be superior to other codebook methods.

3. Statistical codebook searching. In order to reduce mismatching, when making the decision on an upcoming frame, the information from a number of previous frames is taken into consideration to find the codebook item with the highest probability. In [16], a codebook search method was proposed, based on hidden Markov models (HMM).

The problems of codebook method have been further addressed in our thesis, and we have tried to come up with better solutions using Fuzzy Logic techniques.

Chapter Three

Sound Generation and Hearing Process

The generation process of artificial speech has been modeled from the generation process of original natural speech. Therefore, here we explore through the biology of natural speech generation and investigate if the synthetic speech reconstruction is reliable or not. As this work put efforts to overcome hearing loss constrains, we further have a deep look into the auditory system and hearing devices' functionality.

Sections 3.1, 3.2, 3.3 explain the original human speech generation mechanism and its working model, and synthetic speech generation following the same, respectively. Section 3.4 shows the hearing process, while section 3.5 addresses the hearing disabilities and the functions of hearing instruments.

3.1 Human Speech Production

A simplified illustration of the human speech production system is in Figure 3.1. The process of human speech synthesis may therefore be summarized as follows:

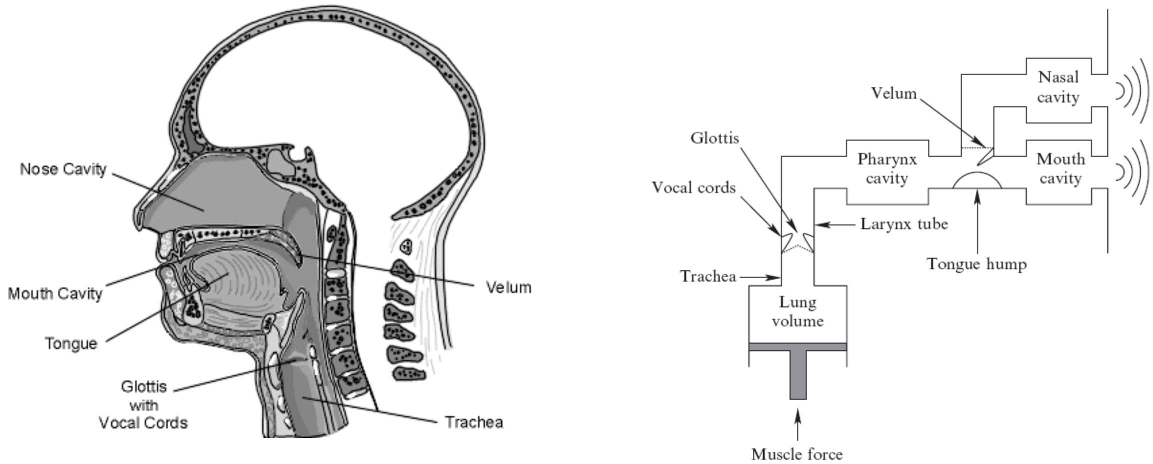


Figure 3.1: The human speech production system

Air pushed out from the lungs travels into the trachea, then up into the glottis, where it is periodically interrupted by the movement of the vocal cords. The tension of the vocal cords is adjusted by the larynx so that the chords vibrate in a oscillatory fashion, resulting in the production of *voiced* speech. During *unvoiced* speech, constrictions within the *vocal tract* (oral cavities – mouth, throat, etc.) forces air through the constriction to produce turbulence. An example is the /s/ sound (fricative) in the word "six".

There are many other anatomical components that contribute to the production of speech, such as the velum, teeth, lips and tongue. These are referred to as *articulators* and move to different positions in order to produce various speech sounds.

3.2 Modeling the system

The speech production system can be further mathematically modeled as follows:

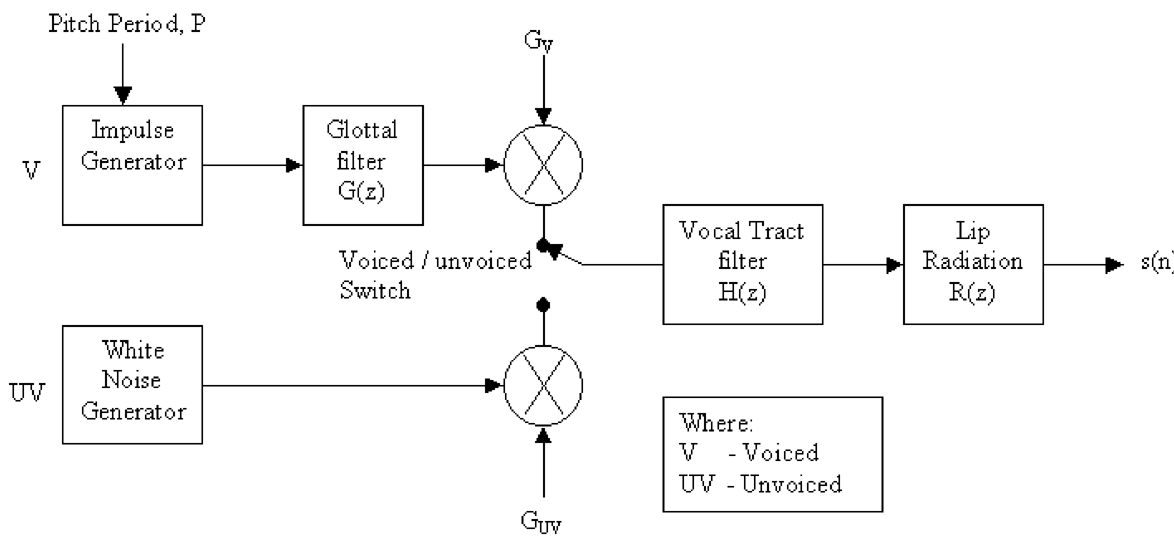


Figure 3.2: A general discrete time model for speech production

Analyzing Figure 3.2, we may derive a system transfer function for voiced and unvoiced speech as follows:

S(z) = E(z).G(z).H(z).R(z) (voiced speech)

S(z) = E(z).H(z).R(z) (unvoiced speech)

3.3 Synthetic Speech Generation

The speech coding that is extracting basic and forming parameters of speech and then reconstructing the signal with the help of the parameters follows the very process that the natural speech generation do. So the estimated generation of artificial speech over the other end of communicating devices almost exactly reproduces the original speech.

The signal processing in speech processes extracts the following three independent components of a speech sound: the excitation signal (voiced or unvoiced + pitch), the energy and the vocal tract configuration.
The voicing decision, pitch period, filter coefficients, and gain are updated for every block of input speech (called a speech frame) to track changes in the input speech.

Since the basic LPC vocoder does not produce high-quality speech, there has been significant effort aimed at improving the standard model. One well known problem with vocoder speech output is a strong buzzy quality. Formant synthesizers use a *mixed excitation* with simultaneous pulse and noise components to address this problem, an idea which has also been used in channel vocoders and LPC vocoders. Other attempts at vocoder improvement included using more realistic excitation pulses and pitch-synchronous LPC analysis.

The basic properties of the speech signal and of human speech perception can explain the principles of parametric speech coding as applied in early vocoders. Current speech modeling approaches, such as mixed excitation linear prediction, sinusoidal coding, and waveform interpolation, use more-sophisticated versions of these same concepts. Modern techniques for encoding the model parameters, in particular using the theory of vector quantization, allow the encoding of the model information with very few bits per speech frame.

3.4 Hearing Process

The ears—pinna, external auditory canal, and eardrum— are the sound-collection system for the body. Abnormalities involving the skin, cartilage, bone, and eardrum may interfere with hearing.

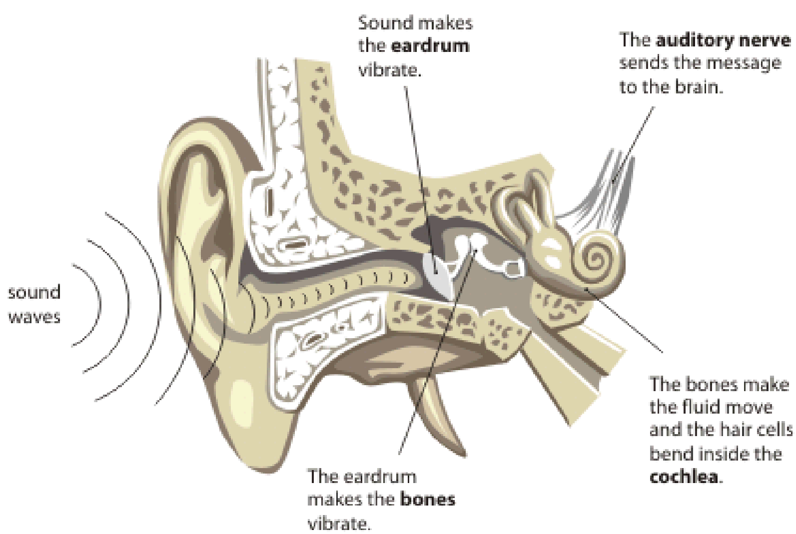


Figure 3.3: Listening mechanics

Sound waves move through the outer ear and set up vibrations in the middle ear. The vibrations are then transmitted to the inner ear, and the wave motion in the inner ear is sensed by the auditory nerve hairs in the cochlea, which transmits neural messages to the brain.

The external ear system collects sound energy for transmission into the ossicular chain (malleus, incus, and stapes) and thence to the fluids in the inner ear (cochlea). Any obstruction to the flow of energy through this system will create a hearing impairment.

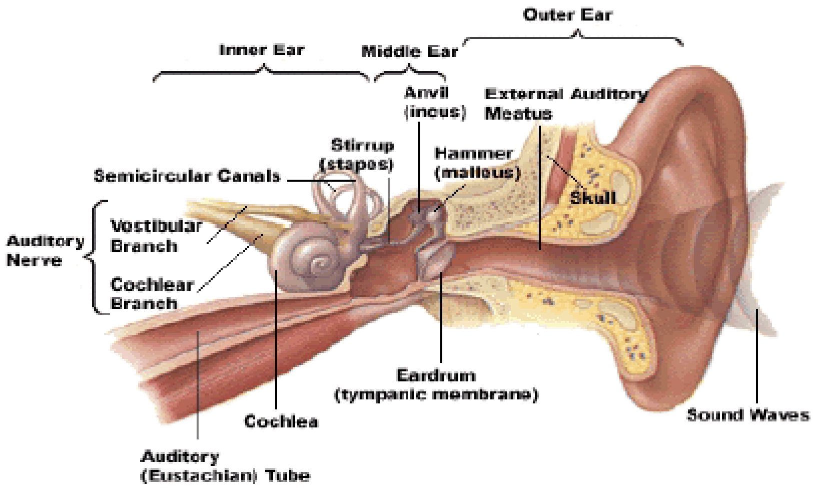


Figure 3.4: Detailed biology of sound hearing

Sound is collected in the outer ear, travels through the ossicles (or bones) of the middle ear and delivers pressure waves through the oval window of the cochlea (of the inner ear) onto the basilar membrane.

The inner ear is a fluid-filled space. The cochlea contains sensory cells that are set off by auditory stimuli. The vestibular system contains sensory cells that are sensitive to rotational motion, linear motion, and changes in the position of the head with respect to the ground.

The Temporal Lobe is the place where hearing tasks are done. The self explanatory images of figure 3.5 and figure 3.6 below shows brain nerves reliable for hearing and speech process.

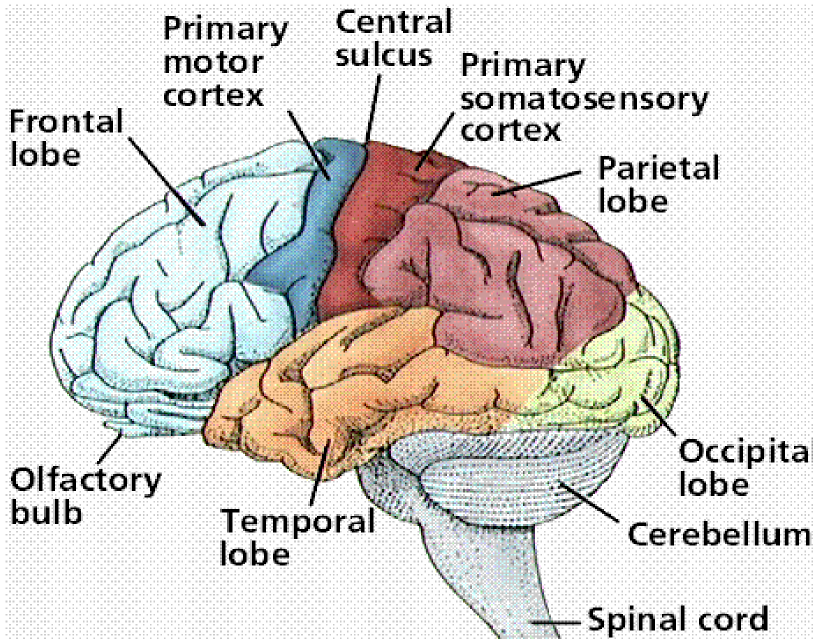


Figure 3.5: Different centers inside the brain

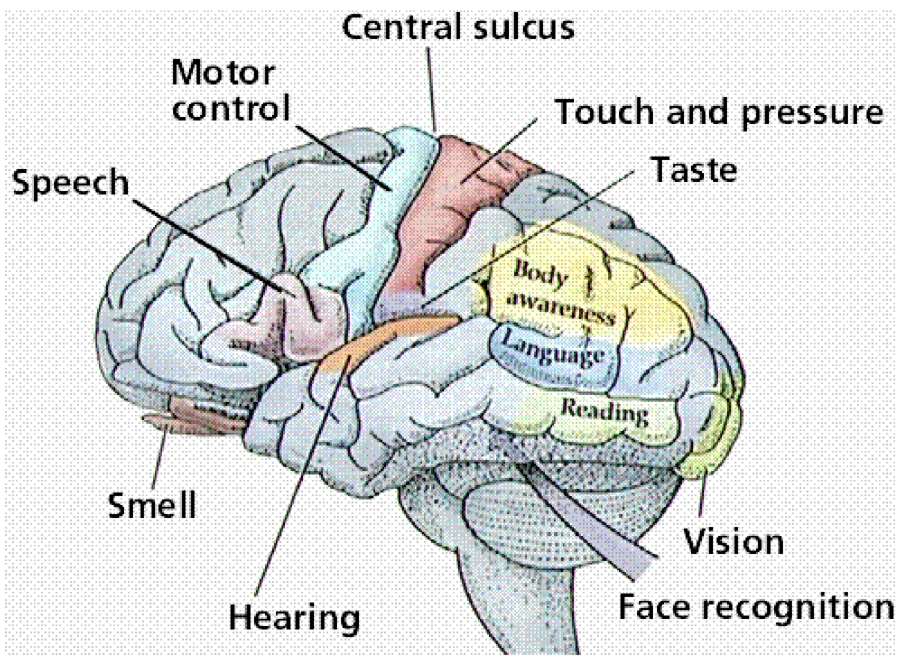


Figure 3.6: Brain positions responsible for different tasks.

3.5 Hearing Disability

Figure 3.7: International symbol for the hearing loss or deafness

Hearing disability may come to a person in many forms. Generally the hearing loss can be catagories into main three form naming mild, severe, profound. Sometimes in few cases the patient is completely hearing lost. Without this complete hearing loss all other cases can be handled by using hearing aids, CI, surgury, or adapting other sort of listening devices. The way these hearing devices communicate with the brain is shiwn in figure

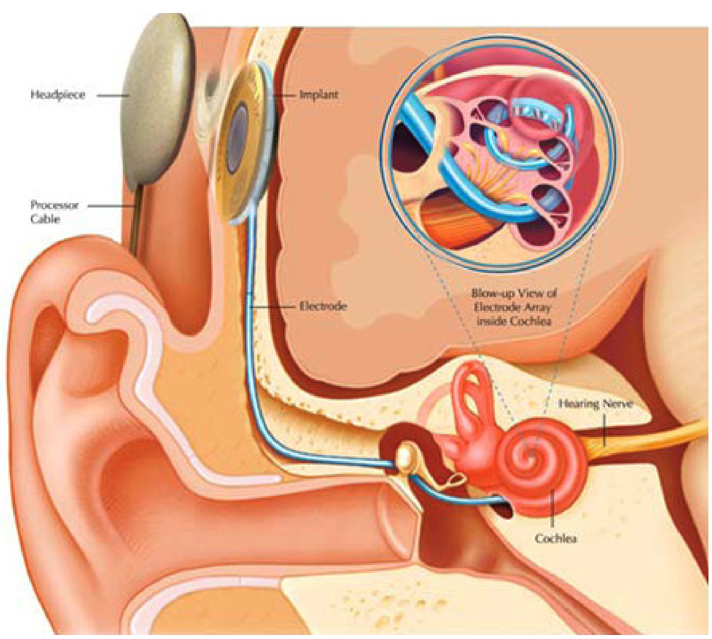


Figure 3.8: Hearing aid and CI to auditory sensors.

Chapter Four

Low band Extension of Telephony Speech

The lower band that is lost in telephone speech (50-300Hz) contributes to the speech quality and naturalness. Development of low band extension techniques are less than that of high band. In this chapter we have proposed a low band extension module to get the clear, natural and high quality speech for the hearing impaired.

4.1 Low band and High band extension of band limited Telephone speech

As mentioned earlier the bandwidth extension process described above can be used for both low band (50-300 Hz) and high band (3400-8000 Hz) extension of band limited telephone speech. Although some specific considerations have to be made in their respective cases.

Considering bandwidth extension of the low band (50-300 Hz) it contributes a vital part to speech quality and naturalness, hence more comfort for the deaf. The present extension methods used in low band extension depends on accuracy of pitch detection. Taking this matter into consideration in our thesis we tried to come up with a bandwidth extension model (for low band), which could make pitch detection with considerable accuracy. The model is as shown below.

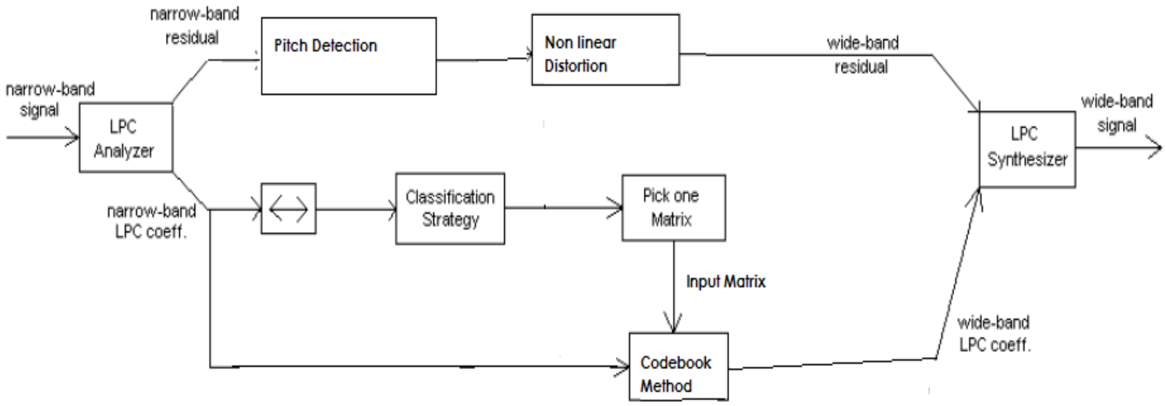


Figure 4.1: Proposed low band extension module

4.2 Proposed Algorithm for Bandwidth Extension:

In this section we discuss the proposed bandwidth extension in details. The overall system flow is shown in the above figure.

The narrow-band speech signal, with an 8 kHz sampling rate, is processed on a frame by frame basis using a 20-ms Hanning window and a 10-ms overlap between adjacent frames. The windowed speech frame is analyzed using an LPC analyzer of order 12. The output of the analyzer goes to three branches:

1 **Pitch Detection and Residual Extension:** At first after going through LPC analyzer, the narrow band telephone speech is fed through the pitch detection. Here the pitch is detected through different pitch detection techniques, the details of which are not concern of our thesis.

But it's worth mentioning over here that pitch detection is an important factor for extension of the low band part of any narrow band speech, and it should be done with considerable accuracy. Keeping this in mind in the next part we have used a Non-Linear Distortion method for the residual extension part. As discussed in the literature review part Non-Linear distortion shows better performance than the Spectral Folding method. Considering the two residual extension methods, studies have shown that Non-linear functions have been used in bandwidth extension algorithms mainly for the generation of low-frequency speech components to date [10]. Furthermore in experiments, very good results were achieved with a non-linear distortion method [10]. The details of this method have been discussed in chapter 2.

2 **Envelope Extension:** Earlier in our literature review section it has been discussed that the two widely used techniques for envelope extension are Linear Estimation method and Codebook method. Both methods have their advantages and disadvantages. It's worth mentioning that human speech has non linear characteristics, and in using a linear model to describe it may yield certain degree of distortion. Chapter 2 of this thesis discusses in details how codebook method can be more advantageous than the linear estimation method.

3 **Classification of Speech Frames:** Speech frame classification is a very important factor in the total grade of speech intelligibility and accuracy. Therefore a good classifier is essential. This issue has been discussed in details in chapter five where we have proposed a new speech frame classifier which is potentially capable of improving the accuracy.

Chapter Five

Speech Frame Classification and Mapping Based on Fuzzy Logic

In the recent generation of very low bit-rate speech coding schemes, one of the most delicate issues is to adapt the appropriate signal excitation to the LPC filter modeling the vocal tract. The problem essentially consists of the need for a good, efficient speech frame classifier. This chapter discusses the speech frames, speech frame classification, feature extraction, mapping methods in speech coding and a comparison of them in sections 5.1, 5.2, 5.3, 5.4 respectively and finally proposes a new speech frame classifier and mapping method comparing with the existing technology. The Fuzzy logic aspects have been discussed in section 5.5 and its successive sub-sections and the potential improvement has been shown in section 5.6.

5.1 Speech Frames

The principle of speech coding is to estimate and transmit the sine wave parameters for each speech frame, and then to synthesize speech using these parameters [9].

Several processing steps occur in extracting spectral-based features. First, the speech is segmented into frames by a 20 ms window progressing at a 10 ms frame rate. A speech activity detector is then used to discard silence/noise frames. Typically, the speech detector discards 20–25% of the signal from conversational telephone speech.

Next, mel-scale cepstral feature vectors are extracted from the speech frames. The mel-scale cepstrum is the discrete cosine transform of the log-spectral energies of the speech segment *Y*. The spectral energies are calculated over logarithmically spaced filters with increasing bandwidths (*mel-filters*). For band limited telephone speech, cepstral analysis is usually performed only over the mel-filters in the telephone pass band (300–3400 Hz). Lastly, delta cepstral values are computed using a first-order derivative of an orthogonal polynomial temporal fit over ±2 feature vectors (two to the left and two to the right over time) from the current vector.

Finally, the feature vectors are channel-normalized to remove linear channel convolutional effects. When training and recognition speech are collected from different microphones or channels (e.g., different telephone handsets and/or lines), this is a crucial step for achieving good recognition accuracy. However, this linear compensation does not completely eliminate the performance loss under mismatched microphone conditions, so more sophisticated compensation techniques such as feature mapping, where transformations are trained to map features coming from particular channels (e.g., microphones) into channel independent features, are also applied.

There are many variants of these cepstral features, such as using linear prediction coding (LPC) spectral analysis instead of fast Fourier transform (FFT)-based spectral analysis, but the basic steps are similar and performance does not vary significantly.

5.2 Speech Frame Classification

The goal of speech classification is to distinguish the silence, unvoiced, and voiced (SUV) regions in the speech signals. The best known operation of this kind is the discrimination between voiced and unvoiced regions, which can be found at the speech vs. sound classification, an issue of great importance in many areas of speech processing. It is well known the algorithm for locating the beginning and ends of an utterance, used in the automatic recognition of isolated words, based on zero-crossing rate and energy. Thus, a frame of speech is considered to be voiced if the energy level is high and the zero-crossing rate is low. On the other hand, if the energy level is low and zero-crossing rate is high the frame is identified as either unvoiced or background silence.

Although this classification scheme performs well for isolated words, it cannot be applied to continuous utterances. In this latter case the problem is that in a sentence the silent regions appear between words, and it is very easy to misclassify them as unvoiced portions. Since this remark is valid also for the unvoiced areas it appears that we need to add additional rules to the above classification scheme in order to better discriminate between silent and unvoiced regions.

The distinguishing characteristic of both waveform comparison and frequency-domain techniques is that the basic measure operates on a per-frame basis and that they use simple schemes to combine the estimated per-frame distortions. The most commonly employed method for the construction of a global objective distortion measure over a number of *N* frames can be computed by arithmetically averaging the per-frame computed distance measures. In a more-sophisticated form of this basic measure, unequal

contributions to perception from each speech frame can be taken into account.
An unequal distribution can be related to frame energy and/or voicing.

The seven parameters identified for voicing classification are as follows: the normalized first autocorrelation coefficient and error prediction, the first coefficient of the LPC filter, the zero *crossing* rate, the number of maxima and minima in the frame (i.e. the zero crossing rate of the differential speech signal), the energy level and the pitch gain.

For voiced speech frames, the excitation is a train of unit impulses with spacing $P0$ (as estimated from the cepstrum), while for unvoiced frames, the excitation is a discrete random noise sequence.

With recent advances in machine learning techniques, vector space modeling has emerged as a promising alternative solution to multiclass speech classification problem. The vector space approach inherits several attractive properties that we discover in text-based information retrieval. For example, it handles the fusion of different types of language cues in a high-dimensional vector seamlessly. As opposed to similarity-based likelihood measurement, the vector space approach is motivated by discriminative training, which is aimed at minimizing misclassification error. In this chapter, we are interested in practical issues related to vectorization of spoken document and vector-based classifier design.

5.3 Feature Extraction

The first stage of feature extraction is framing. The signal is split into overlapping segments of about 25 ms each. These segments are short enough that, within each frame of data, the speech signal is approximately stationary. Each frame is passed through a discrete Fourier transform (DFT), where the time-domain signal becomes a complex-valued function of discrete frequency.

The feature extraction stage seeks to provide a compact encoding of the speech waveform. This encoding should minimize the information loss and provide a good match with the distributional assumptions made by the acoustic models. Feature vectors are typically computed every 10 ms using an overlapping analysis window of around 25 ms. One of the simplest and most widely used encoding scheme is *mel frequency cepstral coefficients (MFCCs)*. These are generated by applying a truncated cosine transformation to a log spectral estimate computed by smoothing an FFT with around 20 frequency bins distributed nonlinearly across the speech spectrum. The nonlinear frequency scale used is called a *mel scale* and approximates the response of the human ear. The cosine transform is applied in order to smooth the spectral estimate and decorrelate the feature elements.
Further psychoacoustic constraints are incorporated into a related encoding called *perceptual linear prediction (PLP).*

Our brain monitors the temporal intensity pattern in each frequency band characterized by a *critical band* filter. By comparing these temporal patterns across a few center frequencies that are not too closely spaced, a reliable estimate of the presented vowel is possible. Such principles are both implemented in perception models and in feature extraction strategies.

5.4 Mapping

Each sound is represented by a single model with several entry and exit states. Speech mapping is done by estimating the probability of these states and relating them to a wideband reference.

A general approach to simulate this process is through statistical mapping and data mining [e.g., Gaussian mixture models (GMMs) or neural networks (NNs)], or clustering. Another way is the recognition of the importance of the high-level cognitive process and applies a statistical data-mining approach. A large pool of candidate features is created and the ones that lead to the most accurate prediction of perceived quality are selected.

Another approach is to use perceptually motivated spectral envelope representations in combination with a mapping from this representation to the quality measure through a GMM. A similar approach but with mappings based on hidden Markov models (HMMs) and neural networks often is used. A recent development simplifies this method and combines the GMM mapping with a small set of features selected for optimal performance from a large set of features. As a result, the algorithm has very low complexity, requiring only a small fraction of the computational capability of a mobile phone.

5.4.1 Codebook vs. Linear Estimation

Figure 5.1 shows a scenario of computational difficulty of linear method in comparison to a non linear method. If the solid squares and circles are to be computed, the first linear approach will be certainly produce large error, but the second approach of non linear calculation will do it more accurately.

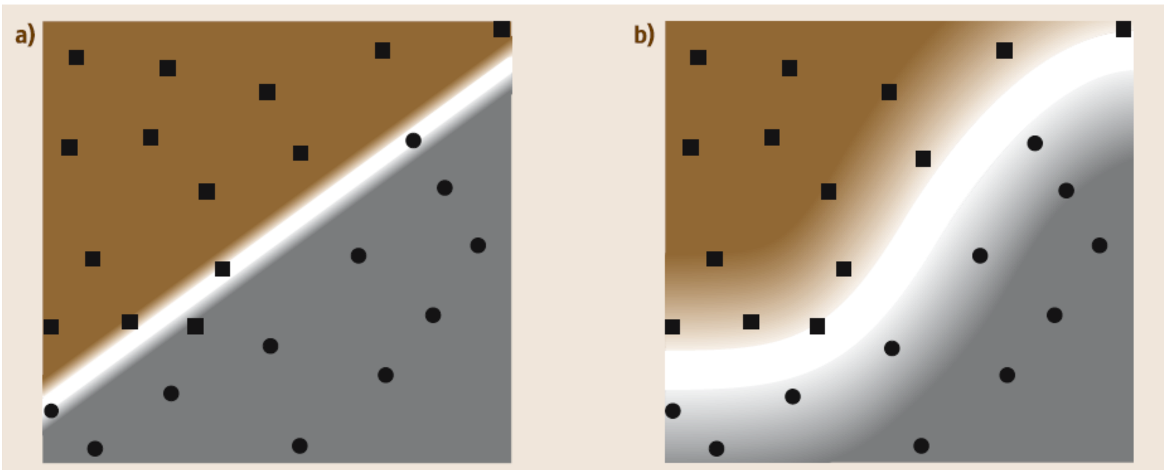


Figure 5.1: non linear vs. linear scaling

Linear estimation method has been used widely for mapping the narrow band feature to the wide band estimations. The reason for its popularity is its ability to mathematically analyze the parameters and estimate linearly. But deriving a matrix by vector quantization of LPC coefficients, the LPC must be converted to linear spectral frequency (LSF) for stability of the system. This extra task makes the process complex and heavy. However the LPC to LSF conversion does not guarantee the desired stability every time.

Despite of its less memory and computational leniency linear estimation method is not the best way to estimate speech parameters, which is non linear in nature. The codebook method, that could be a good candidate, takes more memory and returns less accuracy rate. But using fuzzy logic rather than HMM in codebook method this limitations can be overcome and the accuracy can be higher.

5.4.2 Hidden Markov Model

The foundations of modern HMM-based continuous speech recognition technology were laid down in the 1970s by groups at Carnegie-Mellon, IBM, and Bell Labs.

The Hidden Markov Model is a finite set of *states*, each of which is associated with a (generally multidimensional) probability distribution []. Transitions among the states are governed by a set of probabilities called *transition probabilities.* In a particular state an outcome or *observation* can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are ``hidden'' to the outside; hence the name Hidden Markov Model.

A hidden Markov model is a statistical model in which the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters. The extracted model parameters can then be used to perform further analysis, for example for pattern recognition applications.

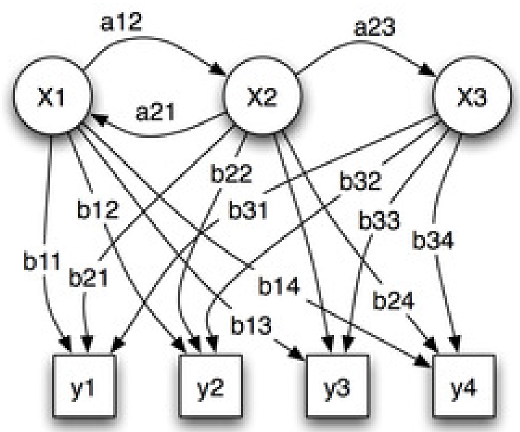


Figure 5.2: Probabilistic parameters of a hidden Markov model
x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

The HMM is a result of the attempt to model the speech generation statistically, and thus works in the side of speech generation. During the past several years it has become the most successful speech model used in speech coding. The main reason for this success is its ability to characterize the speech signal in a mathematically tractable way.

In a typical HMM based systems, the HMM stage is proceeded by the preprocessing (feature extraction) stages. Thus the input to the HMM is a discrete time sequence of parameter vectors. The parameter vectors can be supplied to the HMM, either in vector quantized form or in raw continuous form. Here important point is how the HMM deals with the stochastic nature of the amplitudes of the feature vectors which is superimposed on the time stochasticity.

5.4.3 Limitations and Disadvantages

Although use of HMM technology has greatly contributed to recent advantages in speech coding, there are some natural limitations of this statistical model.

The disadvantage of the rule-based approaches is that intonational (continuous linked forms of speech) contours (outline of speech parameters) lack the richness and variability of natural contours.

In (HMM-based) speech synthesis and vocoders the main goal is high quality speech, and as such, features are chosen that are well suited for re-synthesis. So during feature extraction it discards several important frames.

A major limitation is the assumption that successive observations (frames of speech) are independent, and therefore the probability of a sequence of observations can be written as a product of probabilities of individual observations.

The Markov assumption, that the emission and the transition probabilities depend only on the current state, itself is incorrect. The probability of being in a given state at time t only depends on the state at time t-1is clearly inappropriate for speech sounds.

The observation sequence used for training is finite. For this it either has to increase the size of the training observation set or reduce the number of states. The first option is often impractical and the second one is fixed because every state has its own physical reasons. A suitable solution of this problem increases the complexity. Along with these limitations the following will state more disadvantages of HMM:

- They make very large assumptions about the data.

- The number of parameters that need to be set in an HMM is huge. For example, the very simple three-state HMM there are a total of 15 parameters that need to be evaluated. For a simple four-state HMM, with five continuous channels, there would be a total of 50 parameters that would need to be evaluated.
- The amount of data that is required to train an HMM is very large.
- HMMs only use positive data to train. HMM training involves maximizing the observed probabilities for examples belonging to a class. But it does not minimize the probability of observation of instances from other classes.

5.5 Fuzzy Logic

Fuzzy logic is a true extension of conventional logic. In reality, a rule cannot be defined for each possible case. Exact rules that cover the respective case perfectly can only be defined for a few distinct cases. These rules are discrete points in the continuum of possible cases and humans approximate them. This approximation, and likewise the abstraction and thinking in analogies, are only rendered possible by the flexibility of 'human logic.' To implement this human logic in engineering solutions, a mathematical model is required and Fuzzy logic has been developed as such a mathematical model. Fuzzy logic can be viewed as an extension of multi-valued logic.

Fuzzy systems are suitable for problems requiring approximate rather than exact solutions, which can be represented through descriptive or qualitative expressions, in a more natural way than mathematical equations.

Through simple fuzzy rules, the Fuzzy Voicing Detector (FVD) system achieves a sophisticated speech classification, returning a range of continuous values between the two extreme classes of voiced/unvoiced.

As compared with traditional algorithms, the FVD correctly classifies typically difficult sound and, on account of its fuzzy nature, maintains good performance even in presence of background noise.

Although a continuous voicing range between 0 and 1 is very useful in certain coding *aspects,* in this thesis we present a possible FVD application for classification of the speech segment into three phonetic categories: Voiced. Unvoiced and Mixed (i.e. speech segments generated by simultaneous vocal cord vibration and vocal tract turbulence). Precise identification of this third category, in fact allows a significant improvement in perceptive quality.

5.5.1 Using Fuzzy

Though fuzzy logic analyzes human-like very complex systems but its working principle is based on very simple rules. In applications like control analysis simple and brief sets of fuzzy logic can gain expected smart and controlled output.

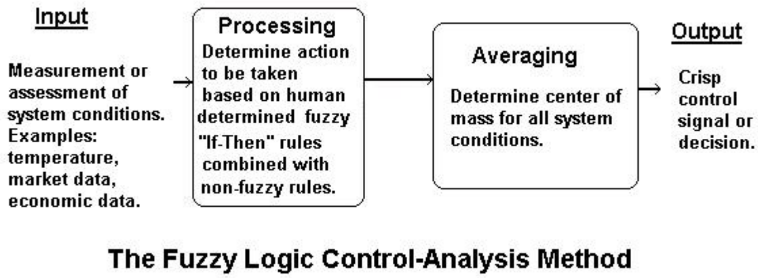


Figure 5.3: Fuzzy Logic Method

The natural speech and hence the characteristics of speech frames have fuzziness. One of the main causes of degradation in the quality of speech coding is the inherent limits of traditional method in adapting the signal excitation of the LPC filter to the non- stationary characteristics of the speech waveform. A more sophisticated speech frame classification would**,** in fact, make it possible to adapt an appropriate signal to the filter modeling the vocal *tract* by better identifying how much is random in the speech segment being dealt with.

In recent years, speech processing specially recognition technology has developed rapidly, yet, to apply them to practical systems problems occur. One of the main reasons is that speech processing technique is based on accurate mathematical results and binary logic. This method does not fit in with the human thought and judgment pattern which are based on fuzzy logic.

Speech frame classification and mapping method based on fuzzy logic rules is mainly in an abstract form till now. No specific technical rules or methods have been developed yet. Some works used fuzzy vector quantization (FVQ)[ ] in using fuzzy logic for speech coding.

Fuzzy logic along with the concept of neural networks and probabilistic reasoning represents one of the main computer methodologies that enable imprecise, incomplete and uncertain information treatment, as also as work with the complex, non-linear problems. These methodologies are commonly named by soft computing.

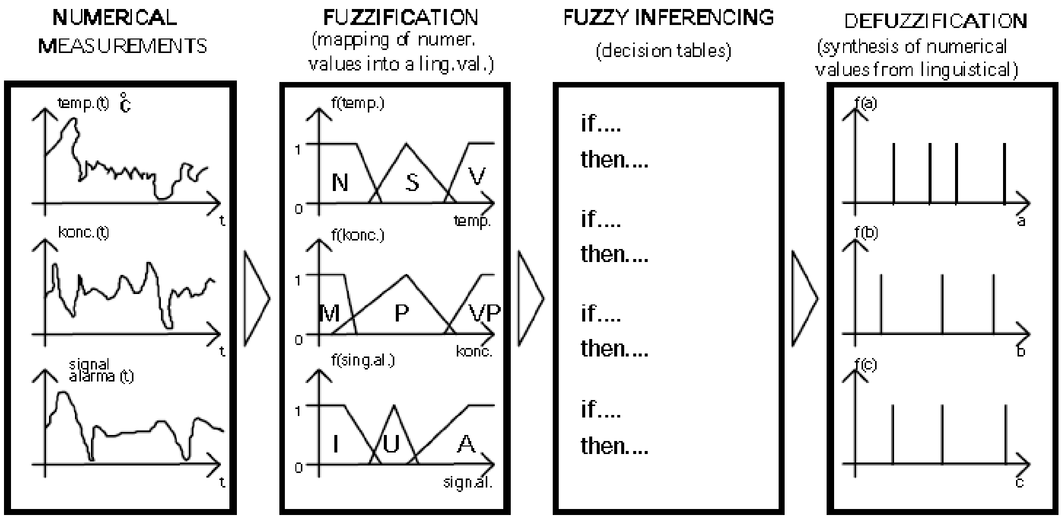


Figure 5.4: fuzzy concept of working

Concept of speech signal fuzzification and defuzzification for a crisp output can be found in recent researches.

5.5.2 Proposed Method

The need to analyze speech frames with fuzzy rules is very obvious. Speech is non linear and fuzzy logic is expert method in recognizing random and non linear observations. The technique proposed, with an estimated low level of complexity, has certain advantages over the main voicing determination algorithms (VDAs), which are based on either simple threshold analysis algorithms or more complex algorithms based on pattern recognition methods. The reasons for possible misclassification errors of VDAs may be different and even contradictory. In general, as they are not very robust in the presence of a high level of background noise, they cause voiced to unvoiced errors, which degrade the speech quality, whereas unvoiced to voiced errors involve a higher coding rate. Weak fricative sounds, for example, which have a low energy level and a random character, are often confused with background noise. On account of its fuzzy nature our system is more robust than traditional solutions, maintaining good performance even in the presence of background noise.

Our proposed mapping method based on the fuzzification technique using CMM can be illustrated in figure 5.5.

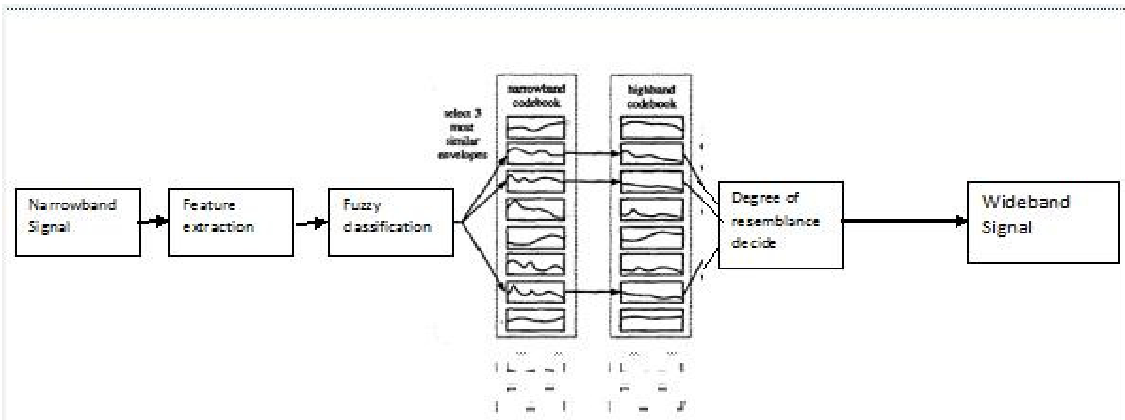


Figure 5.5: Proposed classifier and mapping method based on fuzzy logic

Here after the feature extraction from narrow band all the parameters will be inputted to fuzzy decision process, the fuzzy voice detector will classify the voice. Then the parameters of classified voice will be fed into the codebook mapping method where the mapping decisions will be made by fuzzification and will be linked to the corresponding wideband codebook. Then deciding upon the degree of resemblance an average crisp output will be given by the system.

5.6 Improvements

Multilevel speech classification method based on fuzzy logic has been proposed in [24], some specific improvements could be shown there. Having identified a set of significant parameters in the speech frame and having generated a suitable series of learning and testing patterns, using a tool which learns by example, using simple set of six *fuzzy* rules with a testing and learning normalized mean square error of less than 10 % was obtained. This shows that accuracy rate of above 90% is easily obtainable in fuzzy classifiers.

The machine learning approaches (using decision trees, neural nets and linear regression) have the advantage of being able to produce natural-sounding intonational contours, because they learn the mapping between annotated text and corresponding natural pitch contours in the training phase. However, the machine learning approaches have the disadvantage of requiring large amounts of training data to cover the combinatorial space of phoneme sequences and prosodic contexts.

The proposed classifier should be able to efficiently and smartly classify the speech signal from the extracted parameters. It is computationally easy, depends on simple fuzzy rules, but capable of recognize the complex and random nature of speech features. The existing codebook method will be mapping more accurately with the use of fuzzification - defuzzification technique, with less memory taken. This will eliminate the chances for the system to be unstabilized.

The implementation of proposed fuzzy logic based classification and mapping is not developed yet so, the specific statistics based on test results are not available yet. But the proposal certainly has the potentials to improve the entire speech coding and decoding system comparing to the existing methods, which we have explored to be limited and degraded in many aspects.

Chapter Six

Conclusion

Within the timelines we have tried to cover a widely expended topicIn this thesis we have covered the following topics and problems:

a. Listening problems and solutions
b. Bandwidth extension methods
c. Bandwidth extension to the lower band detecting the pitch
d. Natural and synthetic Audio generation process
e. Natural and lost hearing process
f. Evaluating mapping methods
g. Speech frame classification method
h. Mapping decision method proposal

There are two aspects where this thesis has contributed improvements:

- Low band extension for the deaf
- Speech frame classifier and mapping method based on fuzzy logic

For several decades, spanning the evolution from circuit switched telephony to IP telephony, digital voice communications have relied on narrow band (300-3400 Hz) speech. In that time tele or videoconferencing applications with a need for more elaborate sound content have become more popular. The extra content necessitates a wider audio band, e.g. covering musical instruments, sound effects etc, than that necessary for speech.

These new codecs support not only speech but also mixed content or music. Mixed content can mean advertising, ring tones, music on hold and even movie trailers.

So we propose bandwidth extension implemented not only for the telephone receiver but also in the hearing aid/devices.

With the help of digital fuzzy microprocessors and fuzzy microchips using fuzzy logic gates the proposed classifier and mapping algorithm can be implemented.

Devising the whole system/module consisting BWE (low and high), Fuzzy classifier and mapping algorithm, and implementing these in hearing instruments is our future plans.

The entire quality improvement of the speech coding process will also be a great future aspect.

References

[1] Dr. Mark Ross, "When a Hearing Aid is Not Enough," *US Dept. of Education, NIDDR, NY*, November 1999.

[2] J. Makhoul and M. Berouti, "High-frequency regeneration in speech coding systems," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 1979.

[3] Arttu Laaksonen, "Bandwidth extension in high-quality audio coding," *MSc. Thesis, Dept. of ECE, Helsinki University of Technology*, 2005

[4] H. Qian, O. Poroy and P. Loizou, "A Bluetooth-based wireless phone adapter for cochlear implant users," *Conference on Implantable AuditoryProstheses*, 2001.

[5] J. M. Valin and R. Lefebvre, "Bandwidth extension of narrowband speech for low bit-rate wideband coding," *Proc. IEEE Workshop on Speech Coding*, pp. 2000.

[6] Nels Rohde, Svend Aage, "Artificial Bandwidth Extension of Narrowband Speech," *MSc. Thesis, Dept. of Electronic Systems, Aalborg University*, 2007.

[7] Bernd Iser, Wolfgang Minker, Gerhard Schmidt, "Bandwidth Extension of Speech Signals," *Springer Science+Business Media, LLC*, 2008.

[8] Haifeng Qian, "Bluetooth Receiver and Bandwidth Extension Algorithms for Telephone- Assistive Applications," *MSc. Thesis, MSEE Dept.,The University of Texas at Dallas*, May 2002.

[9] Jacob Benesty, M. Mohan Sondhi, Yiteng Huangv (Editors), “Springer Handbook of Speech Processing,” *Springer Science+Business Media, LLC*, 2008.

[10] Erik Larsen, Ronald M. Aarts, “Audio Bandwidth Extension,” *John Wiley & Sons Ltd*, 2004.

[11] WAI C. CHU, “Speech Coding Algorithms,” *Wiley-Interscience*, 2003.

[12] Y. M. Cheng, D. O'Shaughnessy and P. Mermelstein, “Statistical recovery of wideband speech from narrowband speech,” *IEEE Transactions on Speech and Audio Processing*, vol. 2, October 1994.

[13] S. Chennoukh, A. Gerrits, G. Miet and R. Sluijter, “Speech enhancement via frequency bandwidth extension using line spectral frequencies,” *Proc. of IEEE Int.Conf. on Acoustics, Speech and Signal Processing*, vol.1, 2001.

[14] J. Epps and W. H. Holmes, “A new technique for wideband enhancement of coded narrowband speech,” *Proc. IEEE Workshop on Speech Coding*, 1999.

[15] J. C. Haartsen, “The Bluetooth radio system,” *IEEE Personal Communications*, Feb. 2000.

[16] P. Jax and P. Vary, “Wideband extension of telephone speech using a hidden markov model,” *Proc. IEEE Workshop on Speech Coding*, 2000.

[17] G. Miet, A. Gerrits and J. C. Valiere, “Low-band extension of telephone-band speech,” *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2000.

[18] L. R. Rabiner, “A tutorial on Hidden Markov Models and selected applications in speech recognition,” *Proceedings of IEEE*, 1989.

[19] Peidong Song, Yoshinao Aoki, "Application of Fuzzy set theory in speech recognition," *Hokkaido University, Japan*.

[20] Jerry M. Mendel, IEEE Fellow, "Fuzzy Logic Systems for Engineering: A Tutorial," 1995.

[21] R. Culebras, J. Ramirez, J.M. Gorriz, J.C. Segura, "Fuzzy Logic Speech/Non-speech Discrimination for Noise Robust Speech Processing," *Proc. of ICCS Int. Conf.*, 2006

[22] Miloš Manić, Dragan Cvetković, Momir Praščević, "INTELLIGIBILITY SPEECH ESTIMATION USING FUZZY LOGIC INFERENCING," *In the scientific journal FACTA UNIVERSITATIS, UNIVERSITY OF NIŠ*, July 1999.

[23] Albino Nogueiras, Asunción Moreno, Antonio Bonafonte, and José B. Mariño, "Speech Emotion Recognition Using Hidden Markov Models," *Proc. of Eurospeech 2001 – Scandinavia*, 2001.

[24] Francesco Beritelli, Salvatore Casale, Marco Russo, "Multilevel Speech Classification Based on Fuzzy Logic," *Proc. of IEEE Int. Conf.*, 1995

Online resources

[25] "Specification of the Bluetooth system," http://www.bluetooth.com, November2008.

[26] "Speech Coding", http://freespace.virgin.net/sanjeev.sarpal/speech_coding.htm, December 2008.

[27] http://www.wikipedia.com, May 2008- January 2009 (based on keyword search)

[28] http://www.fuzzy-logic.com, January 2009

[29] http://cse.unsw.edu.us/~waleed/phd/html/node36.htm, December 2008

Field surveys

[30] Siemens Inc., Dhaka, Bangladesh

1. Nihad Sulatana, Head of Sales, Biomedical Instruments Dept., June 2008

2. Siemens Mobile Servicing centre, June 2008

3. Siemens Hearing Aid Centre, July 2008

[31] Bangladesh University of Engineering Technology (BUET)

1. BUET Library, June 2008

2. Professor Dr. Sattay Prasad Mojumder, June 2008

3. Professor Dr. Aynal Haque, July 2008

[32] School for Assisting Hearing Impaired Children(SAHIC), Mohakhali, Dhaka, August 2008