

PREFACE

Advanced applications of DSP has gained acceptance over the years. Speech processing techniques like speech analysis and synthesis, speech coding and decoding, speech and speaker recognition can have myriads of application at home as well as in business and commercial environments.

Chapter 1 attempts to introduce concept of DSP and its scope. Chapter 2 has been written with the purpose of acquainting the reader with the mechanism of human speech production. In chapter 3 the various speech technology areas have been discussed. Chapter 4 to 7 have covered the various fields in such a way that their applications have become apparent. Last chapter deals with application of speech processing techniques in a home environment.

INDEX

1. Scope	03
2. Summary	04
3. Introduction	05
4. Speech processing	06
5. Speech technology areas	08
6. Speech coding	09
7. Speech analysis and synthesis	11
8. Speech and Speaker recognition	13
9. Speech enhancement	15
10. Home Genie	16

SCOPE

Our topic being Advanced DSP Applications, we have concentrated on its application in the field of Speech processing. The areas that we have covered in Speech processing are Speech analysis and synthesis, Speech coding, Speech and speaker recognition and speech enhancement.

In each of the above areas we have covered the implementation and application. Finally, we have also presented a novel idea where we have tried to combine the various speech technology areas in a bid to showcase future possibilities of DSP applications.

INTRODUCTION TO DIGITAL SIGNAL PROCESSING

Digital Signal Processing:

Digital Signal Processing can be broadly defined as the mathematical manipulation of numerical sequences representing digital signals. In other words, various manipulations performed at the bit level by digital hardware on binary bit stream or "digital signal" is called Digital Signal Processing (DSP).

Digital Signal Processing when employed for communication networks encompasses quantisation of an information-bearing signal that has an analog counterpart. This means the signal is represented using finite precision with each word representative of the value of the sample of an analog signal.

Thus, DSP includes sampling, conversion between analog and digital domains, and changes in wordlength.

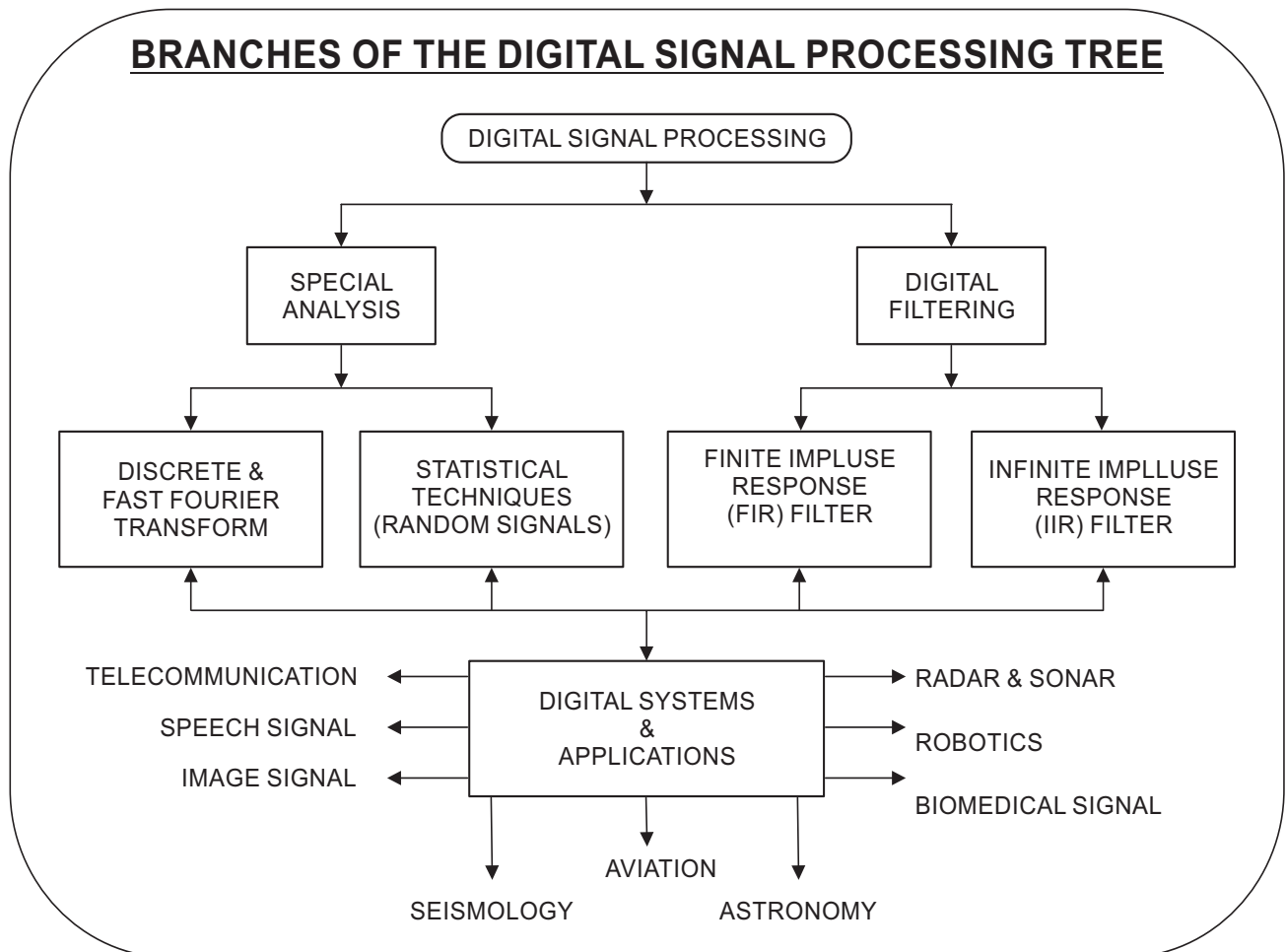
DSP employs high-speed microprocessors called Digital Signal Processors.

DSP operations fall into the following branches: -

- (1) Digital Filters
- (2) Spectral Analysis
- (3) Applications dealing with the use of DSP algorithms for solving practical problems.

Some of the operations are:

- 1) Elimination of redundant data, glitches, spikes and noise distorting the signal.
- 2) Isolation of the components of a composite signal, e.g. separation of a frequency band representing the TV signal of a particular channel.
- 3) Specific Information Extraction from a signal such as the average value, power, etc.
- 4) Information encoding e.g.; speech signal encoding for efficient transmission on a telephone network.
- 5) Information detection used for e.g.; in the detection of a surface ship in the SONAR signal.



SPEECH PROCESSING

Speech processing has been an active area for several decades with a wide variety of applications. Prior to the mid-1960's essentially all speech-processing systems were based on analog hardware implementations. As the field of digital signal processing developed, both in terms of digital hardware capabilities and the development of new signal-processing algorithms, it became clear that the associated techniques could have an impact on the area of speech processing. Many of the developments in the area of digital signal processing, were in fact, carried out within the context of speech processing, partly because the bandwidths associated with speech were well-matched to the processing speeds available.

The Speech Signal: -

Vocal cords and vocal tract are together responsible for the human speech. The vocal cords consisting of two bands of tough elastic tissue, located at the opening of the larynx, vibrate when the air from the lungs passes between them producing sound waves which are emitted from the lips and to some extent from the nose; these sound waves are heard as speech.

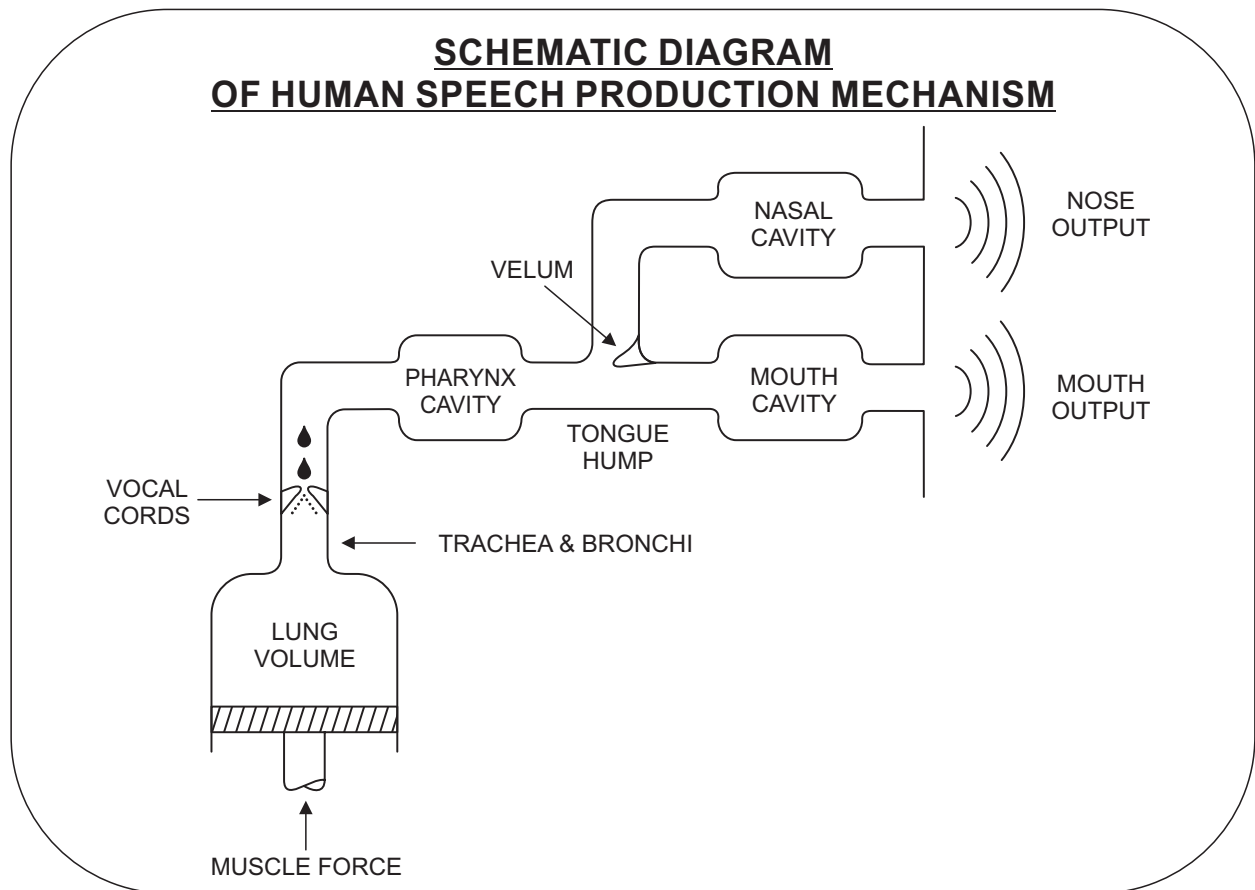
Speech sounds are of two kinds: - voiced sounds and unvoiced or fricative sounds. Quasi-periodic pulses of air exciting the vocal tract produce voiced sounds. To produce an unvoiced sound, a constriction is formed at some point along the vocal tract, generally towards the mouth. Air is forced through the constriction at high speed resulting in turbulent airflow. A broad-spectrum noise source is created which excites the vocal tract leading to the emission of an unvoiced sound.

Characteristic parameters of speech: -

There are three vital parameters for human sound: -

1. Pitch- Related to the frequency of sound (in Hertz)
2. Loudness- The physiological perception of intensity of sound (in dB)
3. Quality- Property possessed by sound by virtue of its harmonic contents

The smallest units of sound which can be recognized by contrast with their environment, are termed phonemes e.g./b/, /g/, /t/, /k/, /p/ etc. Dipones are sounds stretching from the middle of one phoneme to the center of the next.



SPEECH TECHNOLOGY AREAS

The major speech technology areas where DSP is applied are: -

1. Speech analysis and synthesis: -

Analysis of speech is concerned with study of its spectrum and extraction of time varying parameters from the signal for producing speech.

Synthesis of speech is the creation of speech-like waveforms from textual words or symbols, using a model for speech production and time varying parameters of the model. It involves the conversion of an ASCII text message to speech for artificially generating intelligible and naturally sounding voice to build talking ability into a machine.

2. Speech coding: -

Capturing the speech of a person and processing it for storage or efficiently conveying it to another person over a transmission medium such as a communication channel.

3. Speech recognition: -

Extracting meaning from a speech input whereby a person could request for information or service from a machine by talking to it.

4. Speaker recognition: -

Deciding whether or not an unknown speech specimen was spoken by the individual speaker whose identity was claimed.

5. Speech enhancement: -

Minimization of the derogatory effects of noise on the performance of speech communication systems by signal estimation, source coding and signal classification based on statistical models and performance measures.

SPEECH CODING

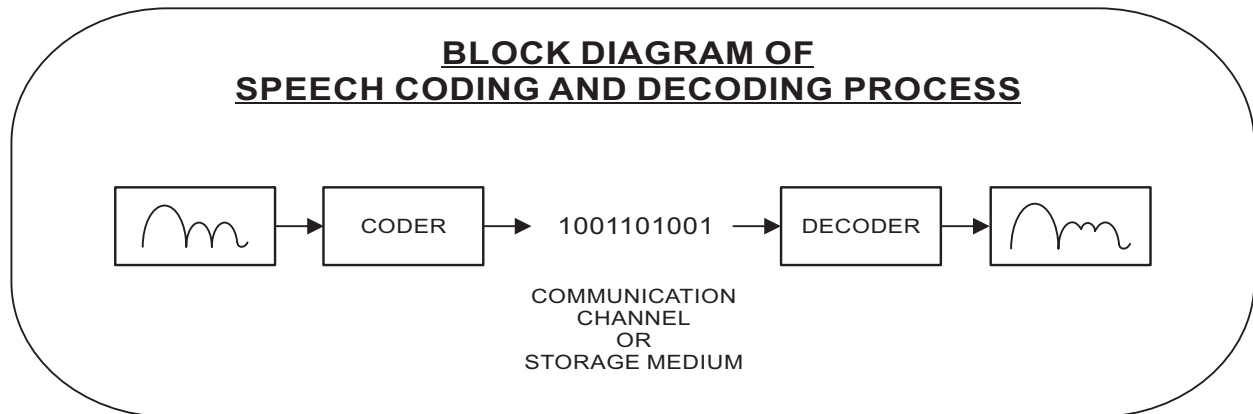
Speech coding deals with compression of speech so that bandwidth required for transmission is reduced. This also helps in storage at the lowest bit rate and with a desired level of quality. In speech coding, a speech signal is converted to a digital representation, which is a sequence of binary digits. This is done by direct application of sampling process. Different methods are employed for speech coding.

They are as follows: -

1. Waveform coding – In waveform coding the signal is preserved, i.e. the decoded signal is approximated as closely as possible to the input signal. It transmits signal at a low bit rate of 12kbits/sec. The different waveform coding techniques are PCM, DPCM, APCM, Linear predictive coding, frequency domain coding etc.
2. Baseband coding – This technique is used for lower bit rates of around 8kbits/sec. But in this case, quality is sacrificed, as there is a small amount of distortion. This method is used in communication applications.
3. Narrowband coding – Below 8kbits/sec baseband decoders do not perform satisfactorily. Narrowband coding is used to transmit signal intelligibly at low bit rate by using production and perceptual aspects of speech.

Applications of Speech coding: -

- ☐ It is used in digital transmission of telephony, narrow band cellular radio, military communication and secrecy missions.
- ☐ Voice mail sent on telephone network, voice encryption, integrated voice and data transmission over packet network.
- ☐ Speech storage and its synthesis from stored speech make use of speech coding.
- ☐ Speech coding can be used in future personal communication.



SPEECH ANALYSIS AND SYNTHESIS

Speech analysis involves the extraction of the time-varying filter parameters from the speech waveform.

Speech analysis methods commonly used are:

1. Homomorphic filtering: -

In homomorphic filtering two components of signal excitation function and vocal tract impulse response to be convolved are added to give the output called complex cepstrum. The additive components can be separated by linear filtering.

2. linear predictive analysis: -

In this method the speech waveform of vocal tract response and excitation function are convolved to obtain parameters of speech by two methods: Autocovariance and autocorrelation. The spectral matching properties of linear prediction method suit the characteristics of speech hence it is very strong method of speech analysis.

Speech synthesis involves the generation of the correct acoustic frequency of human words as language depending upon the required application. The main factors affecting speech synthesis are: -

1. Quality of synthesized speech

2. Speech fluency

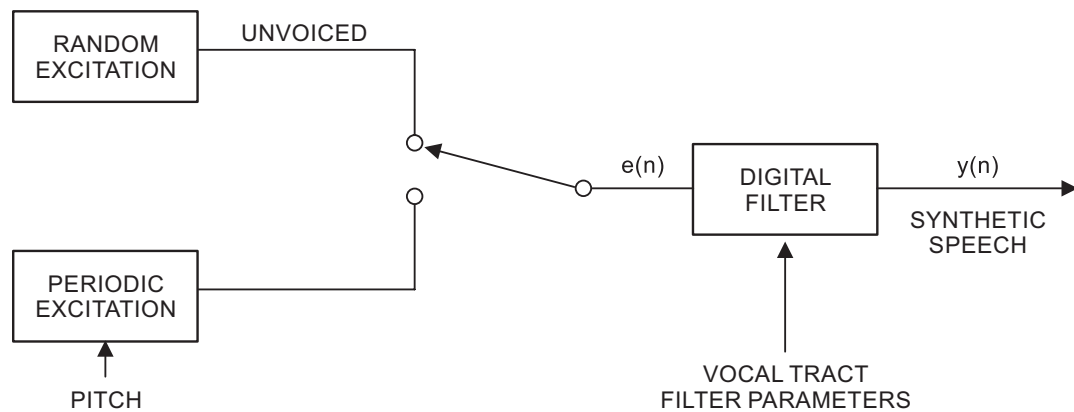
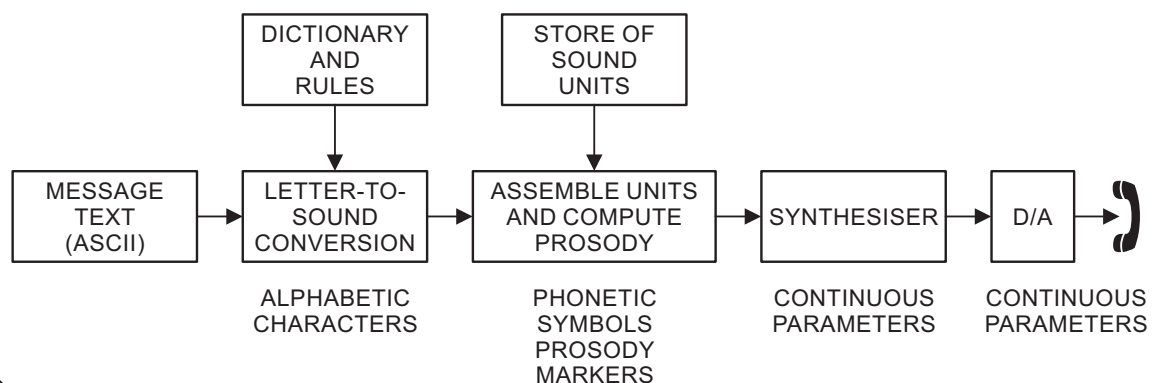
3. Naturalness in pronunciation of individual words and whole sentences.

4. Complexity and cost of hardware realization.

In text to speech synthesis, synthesized speech is generated when the required word on sentences is either typed or spelt. As per the input the encoded word, phrase and sentence are concatenated and speech is synthesized. In voice modeling synthesis the signal is analysed and stored and then reconstructed at any time and place.

Applications:

- ☐ It is used in a non-interactive automatic intercept system for telecommunications such as announcement systems about time, warnings, voice output of e-mail, voice alarms.
- ☐ It is also used in interactive systems such as learning systems. Reading machines for dumb or blind (handicapped). This is the same system that is used by the world renowned scientist, Stephen Hawking who cannot talk due to a paralysis attack. The text that is typed by him is converted to speech using speech synthesis technique.
- ☐ It is used in data enquiry service e.g. railways, flight information, sports scores etc. which requires unlimited & unconstrained text speaking capability.

BLOCK DIAGRAM OF SPEECH SYNTHESIS SYSTEM**BLOCK DIAGRAM OF FULL TEXT-TO-SPEECH SYNTHESIS SYSTEM**

SPEECH AND SPEAKER RECOGNITION

In speech recognition we extract the information from the speech while in speaker recognition we establish the identity of the individual speaker.

Speech recognition:

It includes recognition of message, talker identity determination, understanding the meaning of message and operate machines on the instructions.

There are two methods of recognition:

1. Pattern recognition: In this method either isolated words or connected words are recognized by converting speech signal into digital and comparing this compressed signal with the stored vocabulary patterns. The closely identical pattern to input speech is determined gives the response for recognized word otherwise the user is asked to repeat the word in a different way.
2. Continuous recognition system: This system uses smaller elements like phonemes, starting from spectral and phoneme levels, these models progress towards the word and grammatical levels.

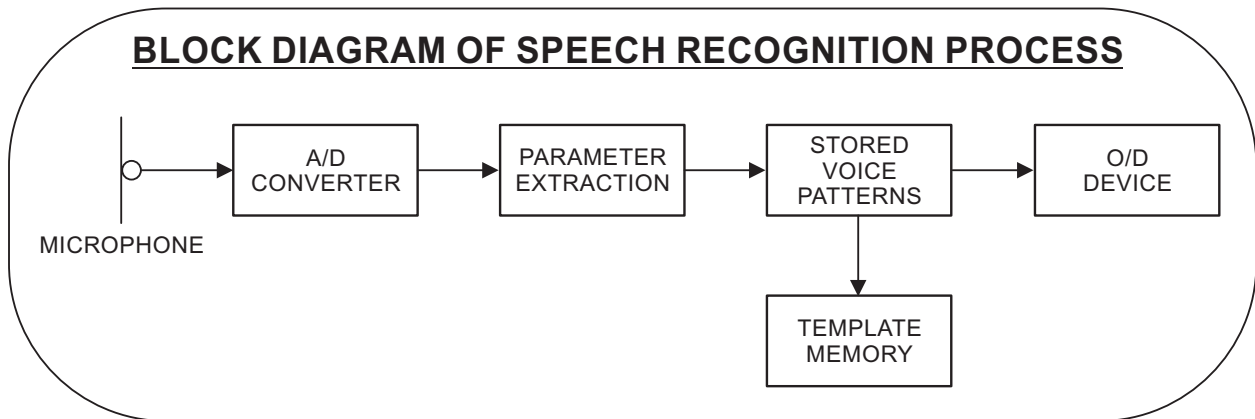
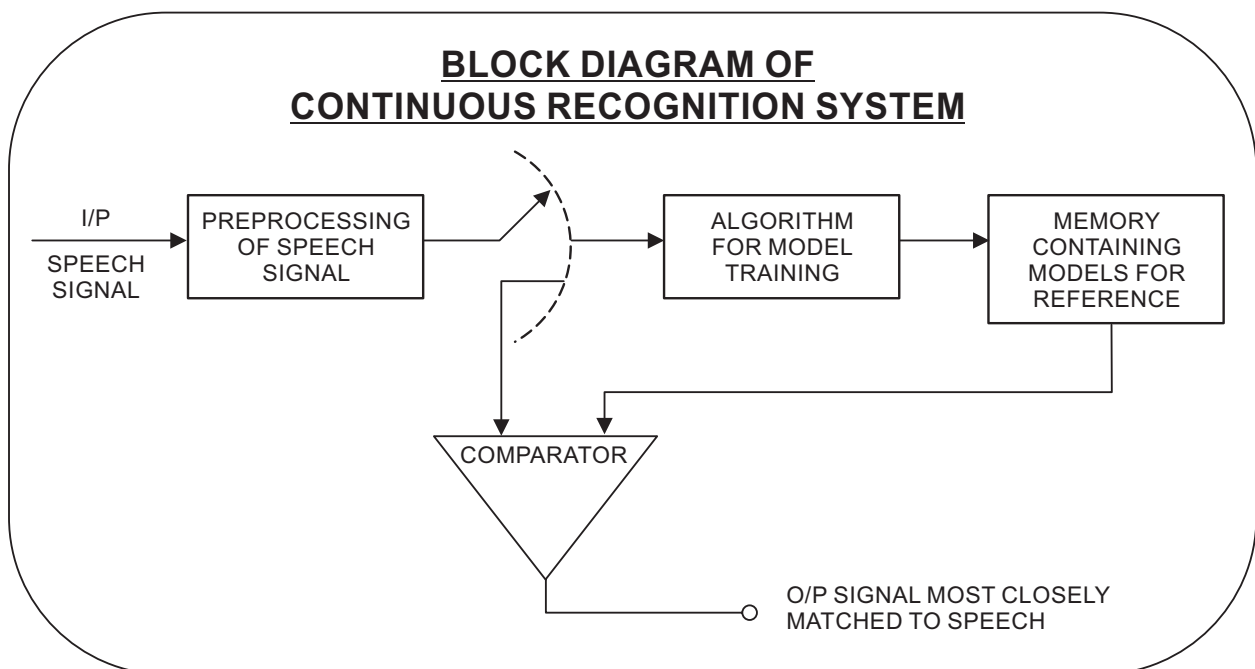
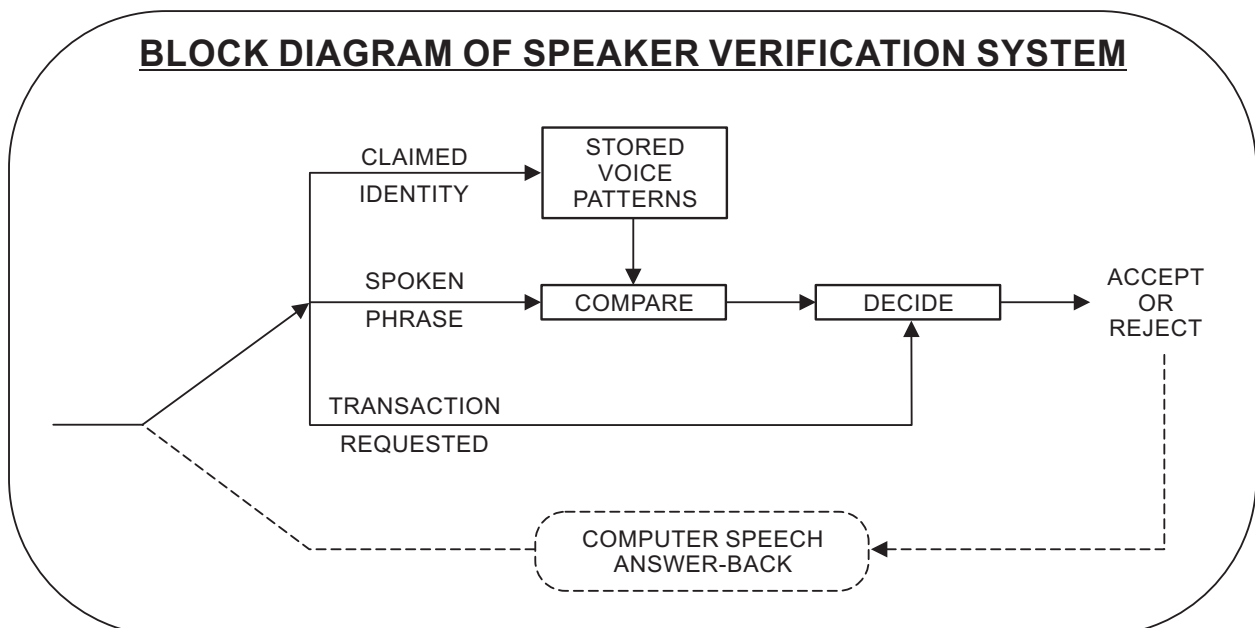
Speech understanding technology recognizes the word in a speech, understands it's meaning and responds by doing the action as instructed by accessing a knowledge based design.

Speaker recognition:

Speech technology is also used in speaker recognition. In speaker recognition, the time-alignment procedure similar to the one used in speech recognition is used. The speech features of the customer's sample are compared to the voice pattern corresponding to the claimed identity and if a suitable match is obtained, the identity claim is verified.

Application:

- ☐ Speech and speaker recognition can be used to directly control machines through human voice. E.g. Building aids for the handicapped like voice-operated motorized wheelchair, voice dialing of carphone, voice command in aeroplane cockpit.
- ☐ Information retrieval systems such as voice controlled railway enquiry, house banking from remote location etc.
- ☐ Loading information systems e.g. form or medical report filling, Dictaphones etc.
- ☐ Speaker verification for security systems.

BLOCK DIAGRAM OF SPEECH RECOGNITION PROCESS**BLOCK DIAGRAM OF CONTINUOUS RECOGNITION SYSTEM****BLOCK DIAGRAM OF SPEAKER VERIFICATION SYSTEM**

SPEECH ENHANCEMENT

Speech enhancement methods are used to maintain the naturalness of the speech. They also help to improve certain aspects of speech like intelligibility and quality. Hence the processed speech signal is better in terms of quality, intelligibility and naturalness than unprocessed one. The methods used in speech enhancement are:

1. Spectral subtraction
2. Source coding
3. Signal estimation
4. Signal classification based on statistical models and performance measures.

Microphone enhancements:

Another advance that'll help speech interaction with PCs comes not in speech recognition itself, but in how the sound gets into your PC in the first place. At present, noise-canceling microphone headsets are needed to get a good signal-to-noise ratio for accurate recognition. But these headsets are inconvenient, uncomfortable.

An alternative comes in the form of an array of microphones, arranged along one side of your monitor or keyboard. The signals arriving at each microphone are processed and used to cancel out all sounds except the speaker's voice. The way it works is that the sound of your voice takes a slightly different amount of time to reach each of the microphones. By delaying some of the signals digitally, they can be made to 'line up'. Other sounds come from other points in space, and arrive at each of the microphones in a different order to your voice. After passing through the delay stage, these other signals cancel themselves out.

Applications:

- ☐ It is used in improving the performance of equipments working in noisy environments of offices, factories and streets such as payphones, pilots speech in cockpit.
- ☐ Betterment of normal unimpaired speech for hearing impaired
- ☐ Noise reduction in teleconferencing systems to prevent its broadcasting from one location to another.
- ☐ Removing disturbances of noisy radio channels.

HOME GENIE!!!

During the last few years, there have been sporadic and isolated attempts to apply DSP in several fields of areas of Speech processing. Here we present a conglomeration of several applications in an attempt to develop a home, which will literally be “at your command”.

6:30 a.m.—You wake up to the sweet voice of your alarm clock. The sweet voice is a result of **advanced speech synthesis technique**. You could select the voice that you liked best from a database of voices collected from volunteers all over the world.

On your way to the bathroom you give a voice command to the central processing unit placed conveniently in the center of your home. The message is relayed to your geyser, which heats up your water so that you can have a refreshing bath. **Speaker recognition technique** ensures that the unit accepts a command from its master only, as it recognizes your voice. You had previously trained your system so that it recognizes not only your voice but also the voices of your household members. The system has already stored your voices in a database. This database is then used to verify the speaker’s identity. A microphone array around the unit implements **speech enhancement** so that the right order goes to the right appliance.

7:30 a.m. — You would like to know what’s going on around in the world. A simple command to your central processing unit switches on your television set.

That’s not all you can even listen to news from various websites on your computer. This is possible using **text to speech synthesis technique**.

Next on your agenda is the checking of e-mails. But you have to arrange your briefcase before you reach your office. Here, again text to speech synthesis comes handy. Your computer reads out your mails for you.

8:30 a.m.— You are ready to leave your home to begin your day at work. Before leaving for work you would like to know all your appointments and important meetings for the day. Your digital diary doubles up as your personal assistant and reads out your schedule. You are worried about the safety of your home. But a single command to your DSP chip inside the central processing unit locks all your windows and safety locks on your cupboards. After you step out of your home the front door is locked on your command. The command is interpreted using **speech recognition** and **speech analysis**. You leave for your office peacefully as you know that the system will fail to recognize any

other person's voice and would not open the door for any one other than yourself. As mentioned earlier it could be trained to recognize the voices of your family members.

6:30 p.m. — You are back home after a day's hard work. You are so exhausted that you will feel glad if someone switches on the lights and fans for you. Thanks to **Speech coding methods** the fans and lights can hear your wish and immediately turn on for your convenience.

The above applications are just few of the possibilities presented by the use of speech processing. The various speech processing techniques can be used to make your life easier and simpler whether it be at your home or at workplace. It can provide a strong platform for the personal communicator of the future.