CHAPTER 3

SPEECH PROCESSING

3.1 INTRODUCTION

Speech is the most important natural communication technique between human beings. It is variable in nature since the pronunciation of the words is different every time it is made. A human speech normally consists of 8000 samples per second. It is founded upon a common premise that speech is present in the band of 0-4000 Hz.[60]. Speech transfers information. To analyze the information transferred by speech, certain non-exclusive levels of descriptions are applied. To yield the original information, human brain attempts to decode the speech from different levels [105]. The analyzing of speech signals and the ensuing of processing methods are concerned with speech processing. Speech recognition, speech synthesis, speaker recognition, voice analysis and so forth are some of the applications of the speech processing [70]. From the above, the main three areas of speech processing are speech synthesis, speech recognition and speech coding.

3.1.1 Speech synthesis

The principal intention of speech synthesis is to evolve an instrument to make an intelligible speech by accepting text as the input, the lecture would be natural sounding like persons speaking [69],[104].

3.1.2 Speech Recognition

The automatic speaker recognition uses a machine to recognize a person from spoken phrase. The speaker recognition is formed by two sub fields. These are speaker verification and speaker identification. In speaker recognition, a person is authenticated by claimed identification from his/her part. In speaker identification, the system decides who the person is and there is no priori identity claim [29].

3.1.3 Speech coding

Speech coding is a method to define a speech signal by cutting down the bit rate although retaining its quality [15], [58].
3.2 SPEECH PRODUCTION AND PHYSICAL MODEL

The human speech originates when the air from the lungs is forced through the vocal tract and vocal cords. The two pairs of the mucous membrane are projected into the cavity of the larynx, which develop the vocal cords. Physical speech model is demonstrated in Fig.3.1. The glottis is the opening between vocal folds. The pharyngeal and oral cavities together are known as vocal tract. When communication takes place, the lungs function as an air supplier.

![Physical speech model](image)

Fig. 3.1. Physical speech model

When the reverberation of the folds occurs, the sound is produced by the process of vocal cords opening and closing. The sound resonance is normally known as reverberation. The sound heard by the other person is controlled by the size of the opening of the mouth, size and shaping of the teeth and lips. The sound originates when the vocal tract shape is varied by the movement of the tongue. When the velum gets down, it will change the produced sound by the nasal tract acoustically coupled with the rest of the vocal tract [108]. Fig. 3.2 shows the block model of the speech production system.
3.3 CLASSIFICATION OF SOUNDS

The human voice can be expressed as a sequence of sounds known as phonemes. Phonemes are the smallest speech units that can make a variation in the content of spoken words [116].

![Diagram of speech production]

The speech model operations are classified into two main categories:

- Voiced sounds
- Unvoiced sounds

3.3.1 Voiced sounds

The pronunciation of the voiced sounds is by the use of the vocal cords and vocal tract. The fundamental frequency of speech, expressed as pitch frequency or pitch, can be found by the vocal cords. In general, all the vowels are normally voiced signals[24],[116]. During the pitch period, a large rate of repetition is shown at regular intervals which typically lie among 2 and 20 ms, while the amplitude of the speech sound varies through that time period.

3.3.2 Unvoiced sounds
The unvoiced sounds do not use vocal cords at the time of pronunciation. So, the fundamental frequency of the speech cannot be determined in unvoiced sounds [24],[116]. Unvoiced sounds do not establish any form of repetition due to its random nature. The shape and the form of the nasal tract and vocal tracts vary constantly with time; it creates an acoustic filter with frequency response in time varying nature. The formant frequencies are denoted as the resonant frequencies at the vocal tract. It depends on the vocal tract shape and vocal tract dimension which is known as the resonant frequencies of the vocal tract [32],[107].

For speech processing applications, there are two simple methods to distinguish between voiced and unvoiced sounds.

- Short time power function: $x(n)$ is the input speech signal, which is split into small frames; (10-20 ms) then determine the power in all frames

\[
P_{av} = \frac{1}{L} \sum_{n=1}^{L} x^2(n) \tag{3.1}
\]

\[P_{av,unvoiced} < P_{av,voiced}\]
• Zero crossing rate: At \( n_0 \), the input speech \( x(n) \) has a zero crossing rate, that means
\[
x(n_0) x(n_0 + 1) < 0 \quad (3.2)
\]
The voiced signals have a greater rate of zero crossings, since they oscillate faster.

Fig. 3.4 Representation of a speech sound. (a) The speech waveform, (b) Spectrogram, (c) The pitch track.

Fig. 3.4 illustrates the speech sound. Fig.3.4 (a) shows the original speech waveform, plot Fig.3.4 (b) shows the spectrogram of the speech signal, which represents the distribution of energy in the frequency domain. The pitch path of the sound signal is illustrated in Fig.3.4 (c). It is realized that the pitch lies only in the voiced part of the speech. [59]

3.4 HUMAN SPEECH PRODUCTION MODEL

Fig. 3.5 illustrates a human speech production model. Here, the nose and mouth can be expressed as time varying filter. Linear predictive codings are used to bring down the sum of the squared deviation among the input speech and the forecast speech. The frame size of the estimated speech signal for LPC is 20 ms, which is quasi stationary in this time frame [32].

The time varying filter transfer function is \( H(z) \) and is given by
\[
H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \quad (3.3)
\]

\( G \) and \( a_k \) are the prediction coefficients of the transfer function.
3.5 SPEECH CODING SCHEMES

Speech codes are generally separated into three types: source coders, hybrid coders, and waveform coders. The waveform coders produce very respectable quality of spoken language; they use high bit rates. Vocoder use very low bit rate, but they produce speech which sounds synthetic. Hybrid quality speech produces a better quality speech with an intermediate bit rate.

![Diagram of human speech production and electrical system speech model](image-url)

Fig. 3.5 Model of (a) human speech production (b) electrical system speech model
The Fig. 3.6 shows the variation of speech of the speech coders with reference to the bit error rate.

- Waveform coders
- Vocoders
- Hybrid coders

### 3.5.1 Waveform coders

The waveform coders are complication less coders, which can produce a high quality speech signal at the rate and above 16kbits/Sec. If the data rate is less than this level, the recovered signal quality will degrade immediately. These coders are not speech-specific, which can operate on any input waveform with the amplitude and bandwidth bounded by certain limits. This operation is carried out by sampling and quantization. Quantization distortion occurs during the process of quantization. To assess the speech quality, Signal to Noise ratio (SNR) is used. The generally used waveform coding techniques explained below [7],[108]:

#### 3.5.1.1 Pulse code modulation (PCM)

PCM is the most elementary kind of waveform coding, which involves sampling and quantization of the input wave form. Narrow band signals are sampled and band-limited at 4 kHz. To obtain a good quality voice communication in linear quantization, approximately 12 bits/samples are needed, to produce a bit rate of 96kbits/Sec. The bit rate of the samples can be lowered by the utilization of non-
uniform quantization. The simplicity, high quality and low delay are the advantages of the pulse code modulation [7],[116].

3.5.1.2 Delta modulation

To accomplish the digital transmission of analog signals, delta modulation uses a single bit PCM code. In the case of conventional PCM, to represent a sample value, multiple bit codes are used since each code is a binary representation of the sign and magnitude of a particular sample. In delta modulation, a single bit is sent rather than a coded representation. The delta modulation algorithm is more elementary. This single bit specifies that the sample is smaller or larger than the previous sample. If the previous sample is larger than the current sample logic 0 is transmitted, and if the previous sample smaller than the current sample, logic 1 is used.

3.5.1.3 Adaptive differential PCM

Another type of waveform coder is the adaptive differential pulse code modulation (ADPCM). It quantizes among the input speech signal and the predicted signal. If the forecasting is accurate and the variation among the input speech signal and the forecast signal will have a less variance than the input speech signals, then it will be able to quantize with lower bits that would be lower than that required for the original speech signals. The quantized difference signal is appended to the predicted signal to make the reconstructed signal. The adaptive prediction and quantization are the components depending on the operation of the coder [7].

3.5.2 Vocoder

The human voice is the main characteristics of the digitally encoded speech signal. Analyzer and synthesizer are the parts of the vocoder. From the original speech, the analyzer extracts some parameters and transmit these parameters instead of the original signal. The parameters are sent instead of speech to the transmitter. At the receiving terminal, by using these parameters, the original speech signal is restored. Normally, the vocoders are operating in the range below 4.8 kbps. Mean opinion score(MOS), diagnostic rhyme test and diagnostic acceptability test are the subjective measurement methods used to measure the distortion in the vocoders. The primary advantage of the vocodres is their power to reconstruct the intelligent speech
at very low bit rate. Two different types of vocoders are Homomorphic vocoder and Linear predictive vocoder [61].

3.5.2.1 Homomorphic vocoder

The homomorphic vocoder is based upon nonlinear signal processing techniques that yields an improved model of human speech. The vocoder expects that the speech represents a convolution process of the excitation signal and vocal tracts impulse response in the time domain. By using a low-pass filter, it is possible to sort out the excitation signal and the impulse response of the vocal tract since the variation of the impulse response is slow when compared with the variation of the excitation signal. This will enable to produce a more accurate pitch information, which is transmitted along with the coefficients of the signal representing that the accurate pitch information can be conveyed with a data rate of 4 kbps.

3.5.2.2 Linear predictive coder

The linear predictive vocoder is a one of the popular type vocoders and produces a better quality speech at low bit rate. The linear predictive coder uses a different type of vocal tract. By this model each sample of speech shows a linear combination of earlier samples using an infinite impulse response filter to model speech. To reduce the error among the input speech signal and predicted speech, this vocoder uses the coefficients of the filter. After finding the filter coefficients, these are quantized and transferred to the receiver vocoder. The receiver uses these coefficients to get the predicted speech of 20 ms. The reconstructed output obtained by this process results in good quality voice [7],[116].

![LPC employed in vocoders](image)
The design demonstrates the speech production model using LPC employed in vocoders. This vocoder reduces the errors between the original and predicted signal by estimating the coefficients of the filter.

### 3.5.3 Hybrid coder

Hybrid coders try to fill the gap among the source coders and waveform coders. The commonly used and most successful hybrid coders are Analysis by Synthesis (Abs) coder. It uses the same type of linear prediction filter of the vocal tract like in LPC vocoders. Instead of applying voiced/unvoiced model, select an excitation signal to match the reconstructed signal to the input speech signal. The model of Abs coder is shown in Fig.3.8.

Abs coder splits the input speech into 20 ms frames. The parameter of each frame is calculated for a synthesis filter, and then the filter excitation is measured. Then, the excitation signal is passed through the synthesis filter, it minimizes the error among the input speech signal and the predicted signal.

Fig.3.8 Analysis-by-synthesis hybrid coder

### 3.6 SPEECH EVALUATION METHODS

The need of speech quality evaluation is increasing day by day due to the rapidly increasing usage of speech processing algorithms. Quality and intelligibility are the two main attributes of the speech signal, but both are not equivalent. Therefore, different evaluation methods are applied to access the intelligibility and quality of the
processed signal. Quality is more subjective in nature, so the reliable evaluation of quality is difficult.

The assessment of quality of speech can be measured by two methods. These are:

- **Subjective quality assessment methods:** Here the original speech and the reconstructed speech is compared by a group of listeners on a pre-determined scale.
- **Objective quality assessment methods:** Objective quality method evaluates the original and processed signal by a mathematical comparison.

### 3.6.1 Subjective quality assessment method

The subjective quality is the quality indicator of the communication services such as voice, audio and video. Subjective quality assessment methods are applied to calculate the subjective quality of the speech signal. The subjective quality assessment methods use subjective evaluation tests to measure the quality of the speech experienced by the users. The most widely used method is the “opinion rating”. The performance of the system is rated under a five point scale ranging from excellent to bad. These methods can be mainly divided into five categories [87].

- MOS (Mean Opinion Score)
- CMOS (Comparison Mean Opinion Score)
- DMOS (Degradation Mean Opinion Score)
- PC (Pair Comparison)
- Equivalent Q value

#### 3.6.1.1 Mean Opinion Score (M.O.S)

The most commonly used subjective quality measure is M.O.S. It uses “Opinion Score” to measure the quality of the system. MOS can be categorised into two types by depending on the quality factors to be accessed [50],[51].

**Listening quality:**

This measures the quality of speech experienced by the users when listening to the speech. It is not able to consider the quality factors such as echos and delays that cannot be measured in the evaluation test. It is a one way speech quality measure, where the speech is played back to the evaluator.
**Conservation quality:**

This measures the quality of speech experienced by the users when sharing a conversation. This method can measure the quality factors such as echos and delays in the conversations, but it requires specialized equipment to do the conversations in real time. The overall evaluation flow is the same for both the listening quality and the conversation quality. The listening quality as well as the conversation quality use Absolute Category Rating (ACR) test. It is a five-point scale and the scores are assigned from one to five. The MOS score is achieved after getting the ACR score from the sufficient number of evaluators.

<table>
<thead>
<tr>
<th>Category</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td>1</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
</tr>
<tr>
<td>Fair</td>
<td>3</td>
</tr>
<tr>
<td>Good</td>
<td>4</td>
</tr>
<tr>
<td>Excellent</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3.1 ACR rating categories

![Fig. 3.9 Flow to obtain MOS](image)

**3.6.1.2 Degradation Mean Opinion Score (DMOS)**

DMOS evaluates the degree of degradation by comparing the accessed signal with the original signal. The evaluators listen to the input speech signal and then listen to the speech to be evaluated after a lag of 0.5-1 sec. DMOS uses five grade scale from 1-5. DMOS score is the average score from the sufficient number of evaluators [55].
Table 3.2 DCR rating categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degradation is very annoying</td>
<td>1</td>
</tr>
<tr>
<td>Degradation is annoying</td>
<td>2</td>
</tr>
<tr>
<td>Degradation is slightly annoying</td>
<td>3</td>
</tr>
<tr>
<td>Degradation is audible but not annoying</td>
<td>4</td>
</tr>
<tr>
<td>Degradation is inadiable</td>
<td>5</td>
</tr>
</tbody>
</table>

The DCR tests are more convenient than ACR test with low level of degradation in the assessment of the systems. Anyhow, DCR tests proceed with ACR test to achieve the same number of rating scores.

### 3.6.1.3 Comparison mean opinion score (CMOS)

The comparison mean opinion score is used for the analysis of listening quality. The evaluators make their assessments by comparing the original speech and the accessed speech. Anyhow the ordering of the two speeches is changed randomly, and the evaluators do not know which one is the speech to be accessed. The evaluators use a comparison category rating (CCR) of the second speech sample and the first speech sample for the quality assessment [49],[52].

Fig. 3.10 Flow to obtain DMOS

Fig. 3.11 CMOS flow diagram
This method uses a seven point scale to assign the scores ranging from -3 to +3 to express which sound is better and how much is shown in the Table 3.3. CMOS score is the average score from a sufficient number of evaluators.

Table 3.3 CCR rating categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much worse</td>
<td>-3</td>
</tr>
<tr>
<td>Worse</td>
<td>-2</td>
</tr>
<tr>
<td>Slightly worse</td>
<td>-1</td>
</tr>
<tr>
<td>About the same</td>
<td>0</td>
</tr>
<tr>
<td>Slightly better</td>
<td>1</td>
</tr>
<tr>
<td>Better</td>
<td>2</td>
</tr>
<tr>
<td>Much Better</td>
<td>3</td>
</tr>
</tbody>
</table>

3.6.1.4 Pair Comparison (PC)

The pair comparison method is used to assessing the all the available combinations of the two speech samples for deciding the superiority of the speech samples. The evaluators listen to two speech samples, separated by a fixed interval, for each combination and decide that which sample has better quality. This method can obtain a reliable result, and also it is very easy to evaluate for the evaluators.

3.6.1.5 Equivalent Q value

![Fig.3.12 Concept of MNRU](image-url)
The equivalent Q value method calculates the Q value of the Modulated Noise Reference Unit (MNRU), which is equivalent to the quality of the accessed speech. The MRNU adds the noise which is proportional to the amplitude of the speech signal. Q value is defined as the ratio of the speech signal to the added noise, which is expressed in dB. Higher value of Q indicates higher quality signal and a smaller ratio of added noise, and a smaller value of Q indicates low quality signal [38],[56],[95].

3.6.2 Objective quality assessment methods

The subjective quality assessment method is a reliable method for the assessment of speech quality but these methods are expensive and more time consuming. Also, these methods require a set of trained evaluators for measuring the speech quality. Therefore, the need of objective assessment arises. A non-intrusive measure is a different type of measurement, which does not need the access of the input speech signal. Fig.3.13 illustrates the computation of the conventional measure (also known as intrusive) and non-intrusive measure.

![Fig. 3.13 Intrusive and non-intrusive measures](image)

The objective speech quality measures are implemented by dividing the original speech signal into 10-30 ms frames, then determining the distortion measure among the original and the processed signals. Speech distortion can be calculated by averaging the speech distortion measures of every speech frame. The distortion measure can be done either in time domain or in frequency domain. The objective quality measures are classified into six classes [25],[39],[67].

- Time and frequency Signal to Noise Ration measures
- Spectral Distance Measures Based on LPC
- Perceptually-Motivated Measures
- Perceptual Evaluation of Speech Quality (PESQ) Measure
- Bark Distortion Measures
  - Composite Measures
  - Non-intrusive Objective Quality Measures
  - Evaluation of Objective Quality Measures

### 3.6.2.1 Time and frequency Signal to Noise Ratio measures

Time domain or frequency domain can be used to evaluate the segmental signal to noise ratio. The time domain is the simplest objective measure to the assessment of the speech enhancement. The segmental noise ratio can be expressed as [44]

$$ SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left( \frac{\sum_{n=Nm}^{Nm+N-1} x^2}{\sum_{n=Nm}^{Nm+N-1} (x(n)-\hat{x}(n))^2} \right) $$  \hspace{1cm} (3.4)

where the original speech signal is represented as $x(n)$, and the enhanced signal is denoted as $x^*(n)$, the frame length of the signal is $N$ (it may chosen 15-20 msec), and the total number of frames in the signal is $M$. The segmental SNR can be continued in the frequency domain and it produces the frequency-weighted segmental SNR [110].

$$ f_{w}SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left( \frac{\sum_{j=1}^{K} B_j \left[ \frac{F^2(m,j)}{(F(m,j)-\hat{F}(m,j))^2} \right]}{\sum_{j=1}^{K} B_j} \right) $$  \hspace{1cm} (3.5)

where $B_j$ - weight placed on the $j$th frequency band

$K$- is the number of bands

$M$ - total number of frames in the signal

$F(m,j)$ – filter bank amplitude of the clean signal in the $j$th frequency band at the $m$th frame

$F^*(m,j)$ - filter bank amplitude of the enhanced signal in the same band

The frequency-based segmental SNR shows more flexibility to place different weights for various frequency bands of the spectrum.
3.6.2.2 Spectral Distance Measures Based on LPC

Certain objective measures recommended depend on the difference among all pole models of the input speech and predicted speech signals. The various types of LPC-based objective measures include: the itakura-saito measure (IS), cepstral distance measure and the log-likelihood ratio measure (LLR). It is derived from the cepstral coefficients.

3.6.2.3 Perceptually-Motivated Measures

Perceptual Evaluation of Speech Quality (PESQ) Measure:

In perceptually motivated measures, perceptual evaluation of speech quality (PESQ) is the commonly used measuring method. Among all the objective measures examined, the PESQ is the most commonly used assessment tool for speech enhancement as suggested by ITU-T for quality assessment of speech of 3.2 kHz narrow band speech codes and handset telephony [54,94,113]. The PESQ score is calculated by using the average disturbance value $A_{\text{ind}}$ and the average disturbance value $D_{\text{ind}}$ as follows [23],[94].

$$PESQ = a_0 + a_1 D_{\text{ind}} + a_2 A_{\text{ind}} \quad (3.6)$$

Fig. 3.14 Perceptual evaluation of speech quality

The structure of the PESQ is illustrated in Fig. 3.14. The original and degraded signals are first equalized to a standard level and then filtered by a filter. Then, align these signals in time to precise time delays and prepared through an auditory transform to access loudness spectra, which is used in the next stage of PESQ calculation for a measure of audible error [94], [113].
3.6.2.4 Composite Measures

Composite measures are formed from the assembling of multiple objective measures. Regression analysis is applied to calculate the optimum combination of objective standard for maximum correlation. The choice of objective criteria for composite measures depends basically on the observational evidence and suspicion. Composite measure evaluation is performed by MARS technique [87].

3.6.2.5 Non-intrusive Objective Quality Measures

Many of the objective speech quality evaluation methods are “intrusive” as they have need to clean or original signal. These methods predict the quality of speech by evaluating the distortion among the clean and forest signals and then mapping the calculated distortion value to a quality metric. In some of the applications, the input signal is not available in these cases, and non-intrusive speech quality measure is highly desirable for the continuous evaluation of the speech quality [25], [39].

![Non-intrusive technique](image)

Fig.3.15 Non-intrusive technique

3.6.2.6 Evaluation of Objective Quality Measures

Statistical analysis is applied to estimate the correlation between the values of the subjective scores and the objective measures. This can be done by two different ways. One is figure of merit and the other one is correlation of objective measures with subjective listening test. Figure of merit is the commonly used statistical analysis for the objective quality measure.
**Figure of merit:**

Statistical analysis is used to attain the Figure of Merit. This analysis is necessary for the validity of evaluation method of objective measures. The figure of merit can be calculated by two different ways in statistical analysis. The first one is Pearson's correlation coefficient and the second one is standard error of the estimate [87].

The Pearson's correlation coefficient is applied to achieve the correlation between subjective listening scores and the objective measures. The second figure of merit is a measuring method of the standard deviation of the prediction error achieved by employing the objective measures to forecast the subjective listening scores.

### 3.6.3 Speech intelligibility measures

Speech intelligibility is an evaluation of the effectiveness of understanding speech, i.e. a standard of how much effectively speech can be communicated through the system [20]. This measurement is commonly represented as a percentage of precisely understood messages. The speech intelligibility usually comes together with speech quality [84],[85].

\[
\text{Speech intelligibility} = \frac{100}{T} \cdot R \tag{3.7}
\]

T – number of speech samples in the speech

R – number of correct speech samples

![Fig.3.16 Elements in the speech intelligibility system](image)

In general, two different methods are applied to evaluate the speech intelligibility. These are:
3.6.3.1 *Subjective intelligibility assessment*

The subjective intelligibility evaluation is based on words, sentences and phonemes. In between, these three speech materials show a fixed relation. Various scoring methods are used for tests with sentences. Commonly, the mean opinion score is used. The intelligibility score shows a 5-point scale; bad, poor, good, fair and excellent. Speech Reception Threshold (SRT) is a very reproducible test based on sentence intelligibility. In SRT, noise masked sentence is presented to a listener [34],[83]. The listener needs to retrieve the sentence properly. If the listener produces a correct answer, the next sequence is presented with greater noise level 2dB. This will continue till the response of the subject is incorrect.

3.6.3.2 *Objective intelligibility assessment*

The beginning of the development of the objective intelligibility measure was the Articulation index (AI), which evaluates the intelligibility of speech for different types of transmission channels. AI score is expressed in the ranges from 0 to 1. The AI score 0.3 or below is considered poor, the score between 0.3 to 0.5 is fair, 0.5 to 0.7 is good and above 0.7 is excellent [46],[47]. Two generally used objective measures are the Speech Intelligibility Index (SII) and the speech transmission index (STI). The STI standard is based on the generation and analysis of an artificial test signal that replaces the original speech signal. The analysis indicated in the range of 0 to 1. The STI is used perfectly for noise, echos, reverberation, non-linear distortion and bandpass limiting. STI is regulated by IEC standard [21]. STI is calculated by using the weighted sum of modulation transfer index (MTI). The modulation transfer index is derived from the Modulation Transfer Function [46],[47].

\[
STI = \frac{SNR_{av} + 15}{30} \quad (3.8)
\]

\(SNR_{av}\) is the average value of SNR
Table 3.4 Speech Transmission Index ratings

<table>
<thead>
<tr>
<th>STI Value</th>
<th>&lt;0.30</th>
<th>0.30-0.45</th>
<th>0.45-0.60</th>
<th>0.60-0.75</th>
<th>&gt;0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>STI Rating</td>
<td>Bad</td>
<td>Poor</td>
<td>Fair</td>
<td>Good</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

The SII is obtained by the computation of the physical parameters of the channel. The SII can be used for noise and bandpass limiting. The non-linear distortion and temporal effects are not exactly involved. By ANSI standard SII is regulated [10],[103].

![Relation between AI and STI](image)

**Fig 3.17 Relation between AI and STI**

**3.7 Summary**

The area of speech processing such as speech production mechanism, and classification of sounds are explained in this chapter. Also, different types of speech evaluation methods are explained for assessing intelligibility, quality and privacy.