SORCE CODING

3.3 Source Coding

The source information has to be coded into a digital form in order for it to be further processed by the digital communication system. One of the techniques used in wireline applications is *pulse code modulation* (PCM), where the analog voice is converted into a 64-Kbps bit stream. Other wireline techniques, such as *adaptive pulse code modulation* (ADPCM) and *delta modulation* (DM), are also used. These source coding schemes for speech use what is called "waveform coding," where the goal is to replicate the waveform of the source information. This is the reason why computer modems can be used over telephones; the information contained in the waveform generated by a transmitting modem can be reliably received by the receiving modem on the other end, and the

By: Dr.Mohab Mangoud

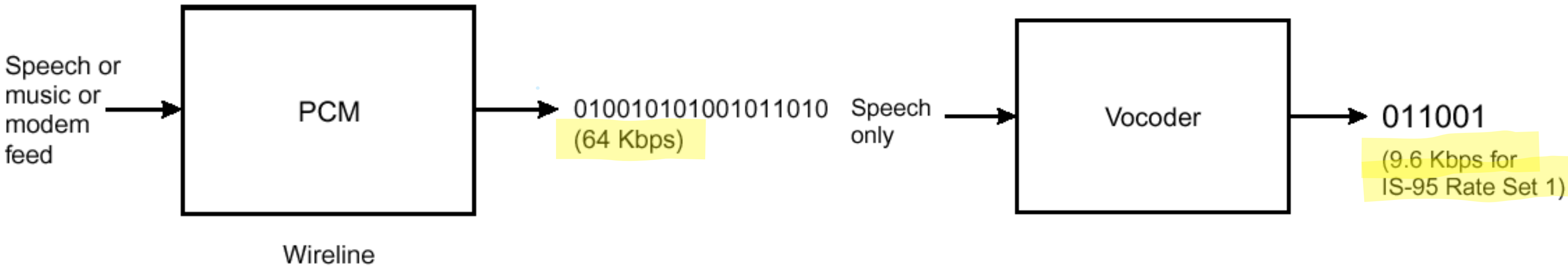
reason is that PCM attempts to replicate the waveform regardless of whether or not the information contained in the waveform is human speech or modulated pitches generated by a modem.

PCM is not feasible in wireless applications because there is a limited bandwidth available. Transmitting 64 Kbps of information over the air demands more bandwidth than can be afforded by most service providers. Therefore, alternative source coding techniques are needed to represent source information (human speech, in this case) using less bandwidth. A vocoder offers an attractive solution. It exploits the characteristics of human speech and uses fewer bits to represent and replicate human sounds. See Figure 3.3.

3.3.1 Characteristics of Human Speech

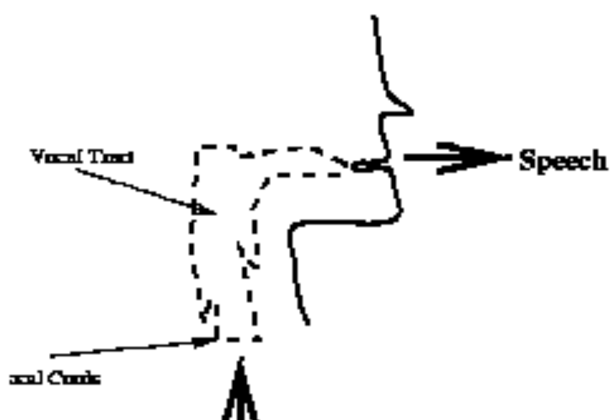
Before we discuss vocoding, it is important that we gain a basic understanding of human speech. The temporal and frequency characteristics of human sound are exploited by vocoders for speech coding. The human voice is made up of a combination of *voiced* and *unvoiced* sounds. The voiced sounds such as vowels ("eee" and "uuu") are produced by passing quasi-periodic pulses of air through the vocal tract. These sounds have essentially a periodic rate with a fundamental

By: Dr.Mohab Mangoud



1.The Basic Properties of Speech

Speech is produced when air is forced from the lungs through the vocal cords and along the vocal tract. The vocal tract extends from the opening in the vocal cords (called the glottis) to the mouth, and in an average is about 16 cm long.



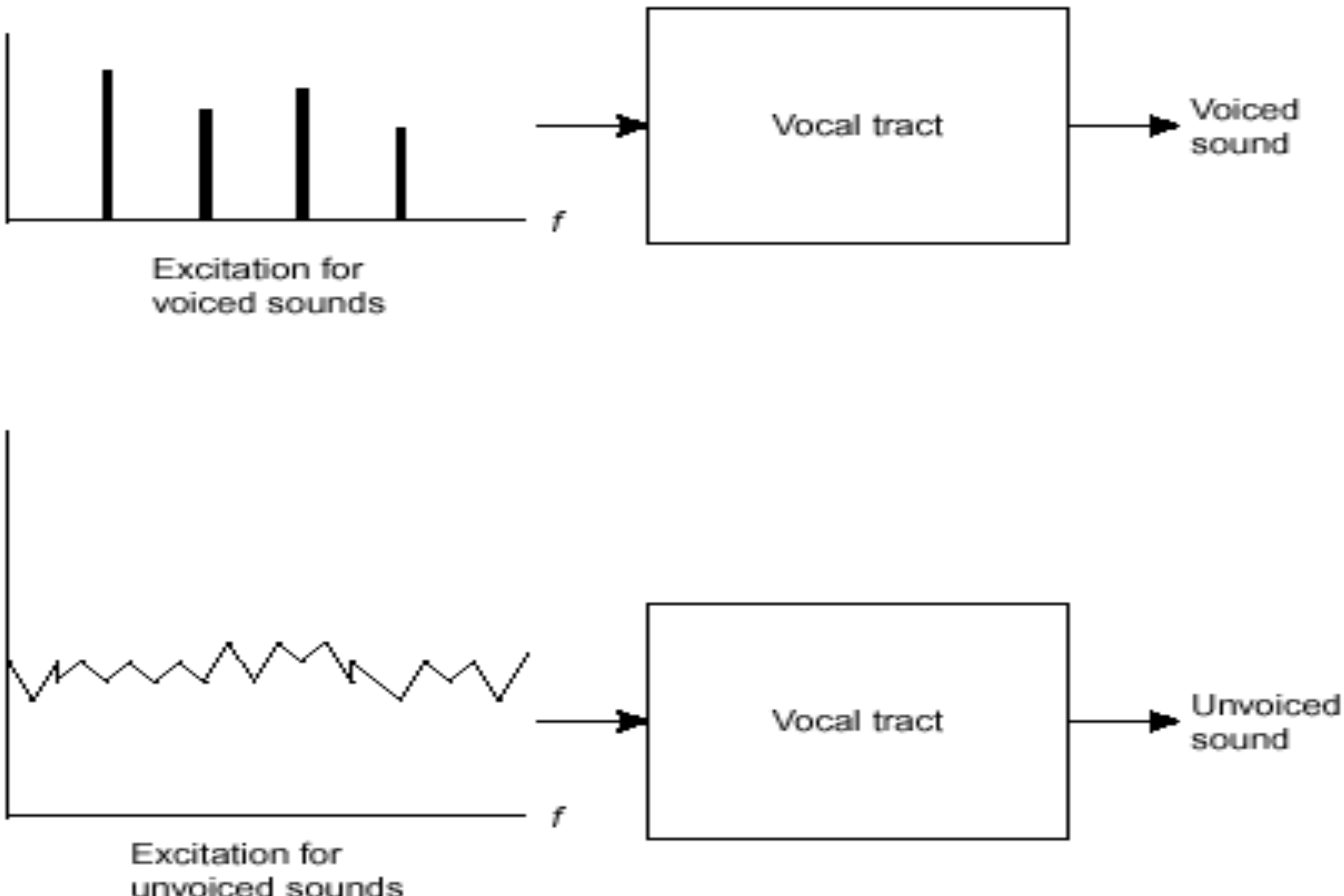
By: Dr.Mohab Mangoud

frequency. This fundamental frequency is also known as pitch. The unvoiced sounds, such as consonants ("t" and "p"), are produced by passing turbulent air through the vocal tract. These sounds are more like acoustic noise created by a closure and sudden release of vocal tract. Figure 3.4 illustrates the principle of sound generation.

Although human voice is time varying, its spectrum is typically stationary over a period between 20 and 40 ms. This is the reason why most vocoders produce frames that have a duration on this order. For example, the IS-95 vocoder produces frames that are 20 ms in duration.

3.3.2 Vocoders

The voice tract can be modeled by a linear filter that is time varying. That is, the filter response varies with time. This is done by periodically updating the coefficients of the filter. This filter is typically all-pole because an all-pole filter requires less computational power than a filter with both poles and zeros. Thus,



By: Dr.Mohab Mangoud

the filter modeling the vocal tract can be represented as $1/T(z)$. If we repr
the excitation signal as $E(z)$, then the spectrum of the speech signal $S(z)$ c
written as

$$\boxed{S(z) = \frac{E(z)}{T(z)}}$$

The all-pole filter $1/T(z)$ can be written as

$$\boxed{\frac{1}{T(z)} = \frac{1}{1 - \sum_{k=1}^{K} b_k z^{-k}}}$$

Equation (3.1) can also be written as

$$E(z) = S(z)T(z)$$

By: Dr.Mohab Mangoud

The all-zero filter $T(z)$ is sometimes referred to as the analysis filter, and (3.3) represents the process of speech *analysis*. The all-pole filter $1/T(z)$ is referred to as the synthesis filter; it is used in conjunction with the excitation signal $E(z)$ to synthesize the speech signal $S(z)$. Equation (3.1) thus represents the process of speech *synthesis*. This type of coding technique is sometimes called analysis-synthesis coding. Figure 3.5 shows how speech is analyzed at the transmitting end and synthesized at the receiving end. The voice encoder analyzes the speech and produces excitation parameters (such as voiced/unvoiced excitation decisions) and filter coefficients valid over the 20-ms interval. The excitation parameters and filter coefficients are the outputs of the speech encoder. In the IS-95 CDMA system, these parameters and coefficients are the encoder. In the IS-95 CDMA system, these parameters and coefficients are the information that is communicated between the transmitter and receiver. The voice decoder at the receiving end uses these parameters and coefficients to construct the excitation source and synthesis filter. The result is estimated speech $\tilde{S}(z)$ at the output of the voice decoder.

Linear-predictive coding (LPC) is widely used to estimate filter coefficients. A feedback loop in the encoder is used to compare actual voice and replicated voice. The difference between actual voice and replicated voice is the *error*. LPC is set up to generate filter coefficients such that this error is minimized. These filter coefficients, along with excitation parameters, are then used by the decoder for speech synthesis.

The IS-95 CDMA system uses a variant of the LPC called *code-excited linear prediction* (CELP). Instead of using the voiced/unvoiced decision, CELP

By: Dr.Mohab Mangoud

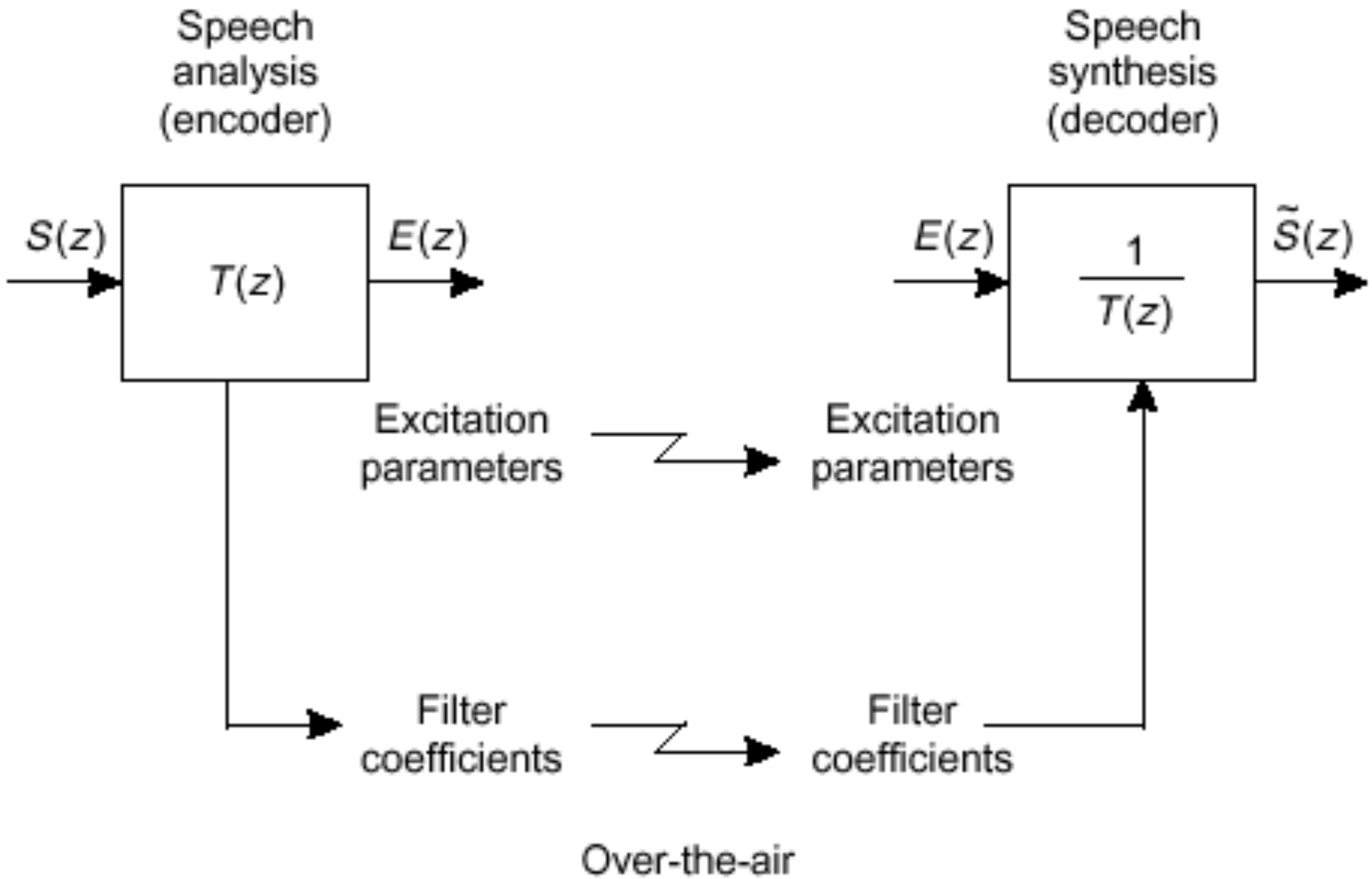


Figure 3.5 Process of replicating human speech.

has a different form of excitation for the all-pole filter. Specifically, the CELP decoder uses a codebook to generate excitation inputs to the synthesis filter. For a complete description of CELP, see Schroeder and Atal [2].

By: Dr.Mohab Mangoud

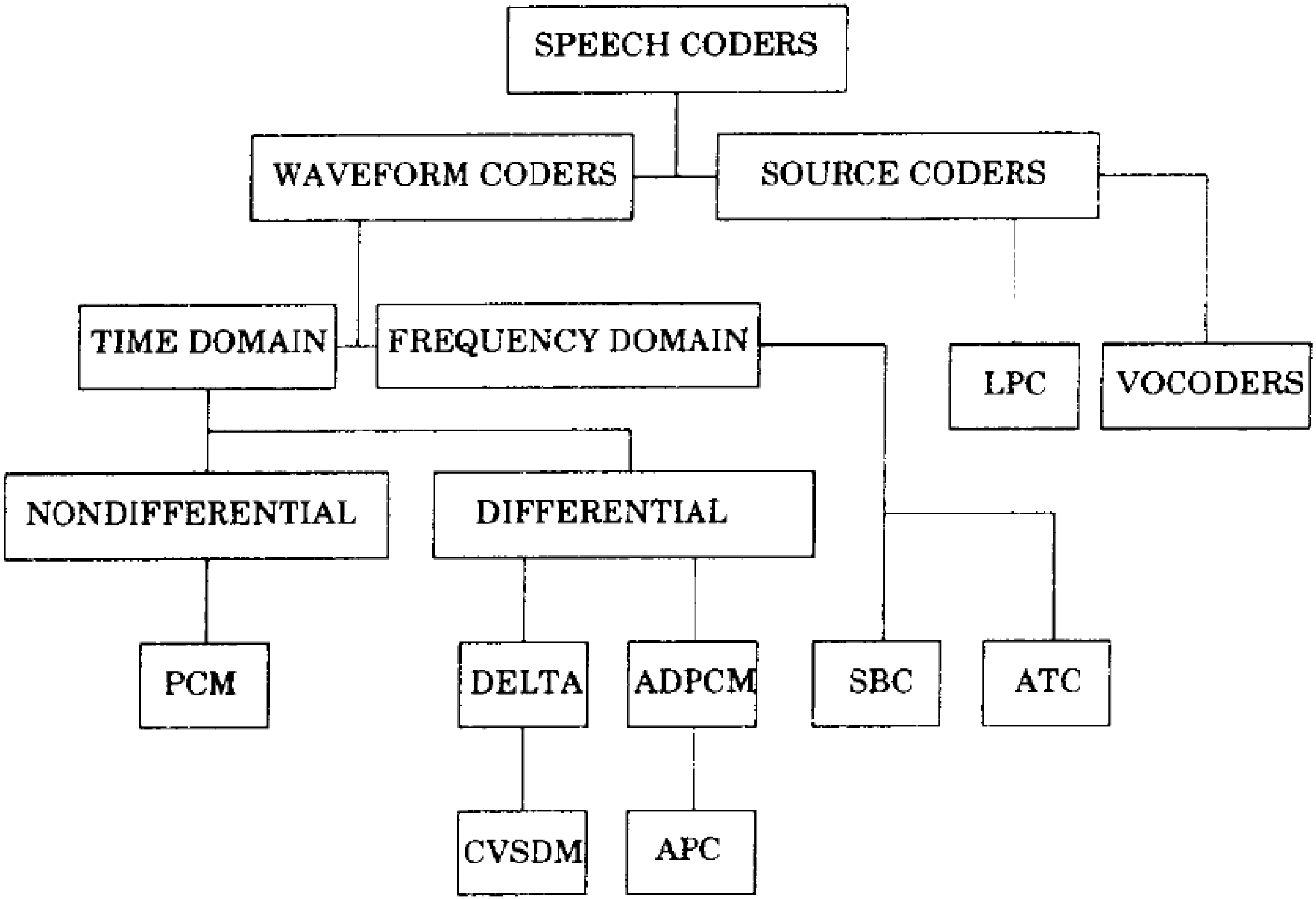


Figure 7.1
Hierarchy of speech coders (courtesy of R.Z. Zaputowycz).

64 kbits/s PCM Codecs (ITU G.711):

Pulse Code Modulation (PCM) codecs are the simplest form of waveform codecs. Narrowband speech is typically sampled 8000 times per second, and then each speech sample must be quantized. If linear quantization is used then about 12 bits per sample are needed, giving a bit rate of about 96 kbits/s. However this can be easily reduced by using non-linear quantization.

For coding speech it was found that with non-linear quantization 8 bits per sample was sufficient for speech quality which is almost indistinguishable from the original. This gives a bit rate of 64 kbits/s, and two such non-linear PCM codecs were standardized in the 1960s.

By: Dr.Mohab Mangoud

DPCM & ADPCM

By: Dr.Mohab Mangoud

If the predictions are effective then the error signal between the predicted samples and the actual speech samples will have a lower variance than the original speech samples. Therefore we should be able to quantize this error signal with fewer bits than the original speech signal. This is the basis of Differential Pulse Code Modulation (DPCM) schemes - they quantize the *difference* between the original and predicted signals.

The results from such codecs can be improved if the predictor and quantizer are made adaptive so that they change to match the characteristics of the speech being coded. This leads to Adaptive Differential PCM (ADPCM) codecs. In the mid 1980's the CCITT standardised a ADPCM codec operating at 32 kbits/s, which gave speech quality that was very similar to the 64 kbits/s PCM codecs. Later ADPCM codecs operating at 16,24 and 40 kbits/s were also standardised. For theory:

4.Source Codecs (FS1015)

Source coders operate using a model of how the source was generated, and attempt to extract, from the signal being coded, the parameters of the model. It is these model parameters which are transmitted to the decoder. Source coders for speech are called vocoders, and work as follows:

The vocal tract is represented as a time-varying filter and is excited with either a ***white noise source***, for unvoiced speech segments, or a ***train of pulses separated by the pitch period*** for voiced speech.

Therefore the information which must be sent to the decoder is the filter specification, a voiced/unvoiced flag, the necessary variance of the excitation signal, and the pitch period for voiced speech. This is updated every 10-20 ms to follow the non-stationary nature of speech.

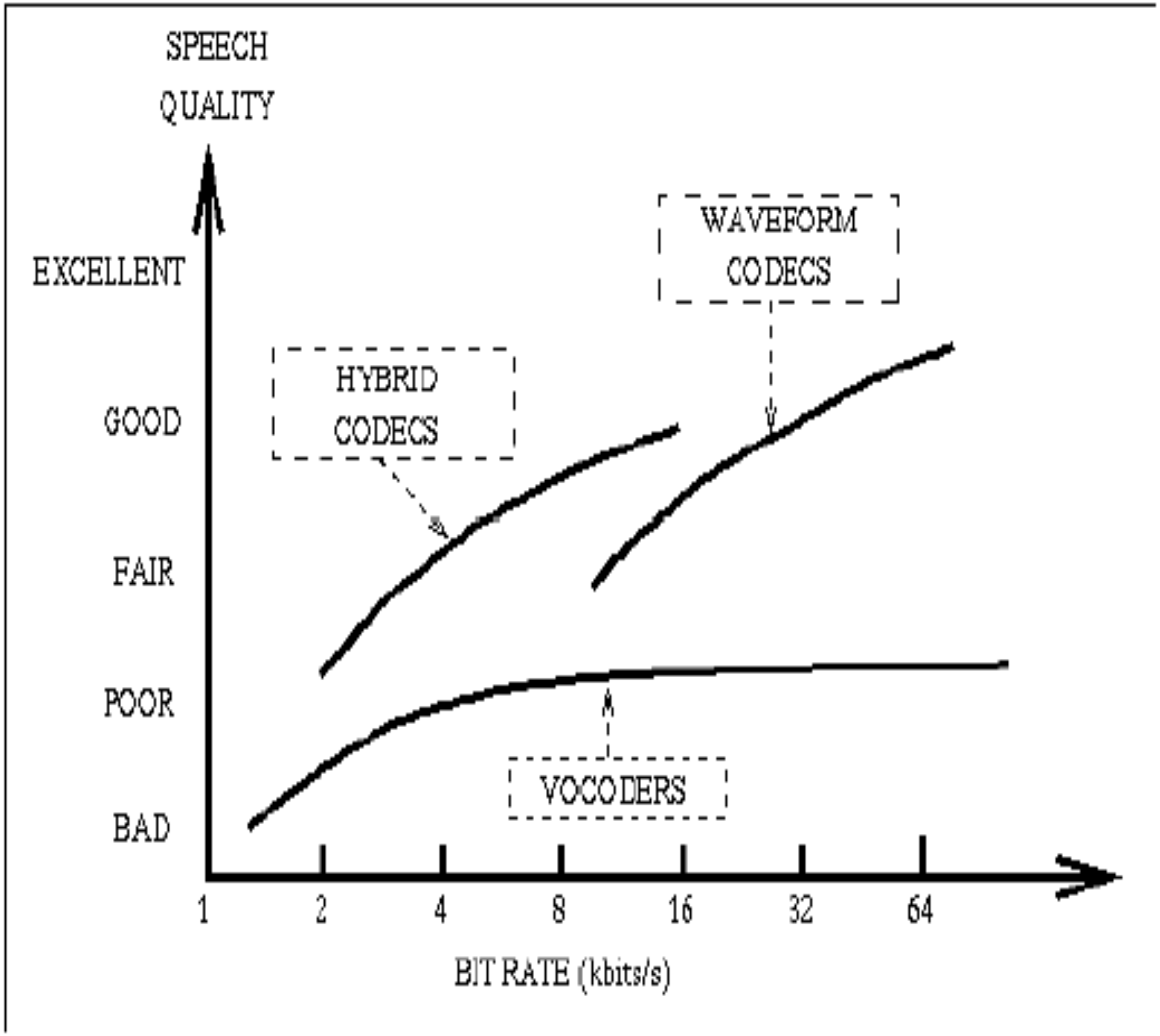


Figure 5 : Speech Quality Versus Bit Rate For Common Classes of Codecs

By: Dr.Mohab Mangoud

Summary of Common Speech Coding Standards:

Standard	year	Algorithm	Bit_rate	application	MOS	Delay
G.711	1972	Mu&A-law, PCM	64kbps	Network transmission	4.3	0.125ms
G.721	1984,87	ADPCM	32kbps	undersea cable	4.0	0.125ms
G.722	1988	Subband ADPCM	48-64kbps	ISDN,Vconf.	4.0	0.2ms
G.726,727	1988	VBR-ADPCM	16-24-32-40kbps	low-tier PCS/cordless	2,3.2,4,4.2	0.125ms
G.728	1992	LD-CELP	16kbps	bi-directional networks	4.2	0.625ms
G.729	1995	CS_ACELP	8kbps	2G cellular	4.0	15ms
G.723.1	1995	MP_MLQ ACELP	5.27/6.3kbps	Videophone H.323, H.324	3.5-3.7	37.5ms
GSM-FR	1989	LTP_RPE	13kbps	Euro_cellular	3.7	20ms
GSM_EFR	1995	ACELP	13kbps	Euro_cellular	4.0	20ms
IS-54	1989	VSELP	8kbps	NA-TDMA	3.5	20ms
IS-96	1993	QCELP	1.2,2.4,4.8,9.6kbps	NA-CDMA	3.3	20ms
GSM_HR	1994	VSELP	5.6kbps	Euro_cellular	3.5	24.5ms
DoD_FS1015	1996	LPC-10	2.4kbps	secure teleph.	≤ 3.0	25ms
DoD_FS1016	1990	CELP	4.8kbps	secure teleph.	3.0	45ms
G.722.2	2001	AMR_WB ACELP	6.6-23.85kbps	VoIP,Vconf., 3G cellular	3.7-4.4	15_25ms

By: Dr.Mohab Mangoud

7.8 Choosing Speech Codecs for Mobile Communications

Choosing the right speech codec is an important step in the design of a digital mobile communication system [Gow93]. Because of the limited bandwidth that is available, it is required to compress speech to maximize the number of users on the system. A balance must be struck between the perceived quality of the speech resulting from this compression and the overall system cost and capacity. Other criterion that must be considered include the end-to-end encoding delay, the algorithmic complexity of the coder, the d.c. power requirements, compatibility with existing standards, and the robustness of the encoded speech to transmission errors.

Example 7.4

A digital mobile communication system has a forward channel frequency band ranging between 810 MHz to 826 MHz and a reverse channel band between 940 MHz to 956 MHz. Assume that 90 per cent of the band width is used by traffic channels. It is required to support at least 1150 simultaneous calls using FDMA. The modulation scheme employed has a spectral efficiency of 1.68 bps/Hz. Assuming that the channel impairments necessitate the use of rate 1/2 FEC codes, find the upper bound on the transmission bit rate that a speech coder used in this system should provide?

Solution to Example 7.4

Total Bandwidth available for traffic channels = $0.9 \times (810 - 826) = 14.4$ MHz.

Standard	Service Type	Speech Coder Type Used	
			Bit Rate (kbps)
GSM	Cellular	RPE-LTP	13
CD-900	Cellular	SBC	16
USDC (IS-54)	Cellular	VSELP	8
IS-95	Cellular	CELP	1.2, 2.4, 4.8, 9.6
IS-95 PCS	PCS	CELP	14.4
PDC	Cellular	VSELP	4.5, 6.7, 11.2
CT2	Cordless	ADPCM	32
DECT	Cordless	ADPCM	32
PHS	Cordless	ADPCM	32
DCS-1800	PCS	RPE-LTP	13
PACS	PCS	ADPCM	32

Number of simultaneous users = 1150.
Therefore, maximum channel bandwidth = 14.4 /1150 MHz = 12.5 kHz.
Spectral Efficiency = 1.68 bps/Hz.
$\eta \quad \times \quad B$
Therefore, maximum channel data rate = 1.68×12500 bps = <u>21</u> kbps. $\eta = \frac{C}{B}$
FEC coder rate = 0.5.
Therefore, maximum net data rate = 21×0.5 kbps = 10.5 kbps.
Therefore, we need to design a speech coder with a data rate less than or equal to 10.5 kbps.