Disclaimer

- The material provided in this document is not my original work and is a summary of some one else’s work(s).
- A simple Google search of the title of the document will direct you to the original source of the material.
- I do not guarantee the accuracy, completeness, timeliness, validity, non-omission, merchantability or fitness of the contents of this document for any particular purpose.
-

Speech Coding

CODING STRATEGIES AND STANDARDS

Outline

- Introduction
- Speech Coding Techniques
- Algorithm Objectives and Requirements
- Standard Speech Coders

Introduction

- PCM systems allow perfect signal reconstruction at the repeaters of the communication systems, error correction, encryption, multiplexing, switching, and compression
- The PCM transmission bandwidth is greater than that required by the original analogue signal
- To reduce the bandwidth different coding strategies have been devised

Speech Coding Techniques

- Major speech coders have been separated into two classes
  - Waveform approximating coders: Speech coders producing a reconstructed signal which converges towards the original signal with decreasing quantization error
  - Parametric coders: Speech coders producing a reconstructed signal which does not converge to the original signal with decreasing quantization error

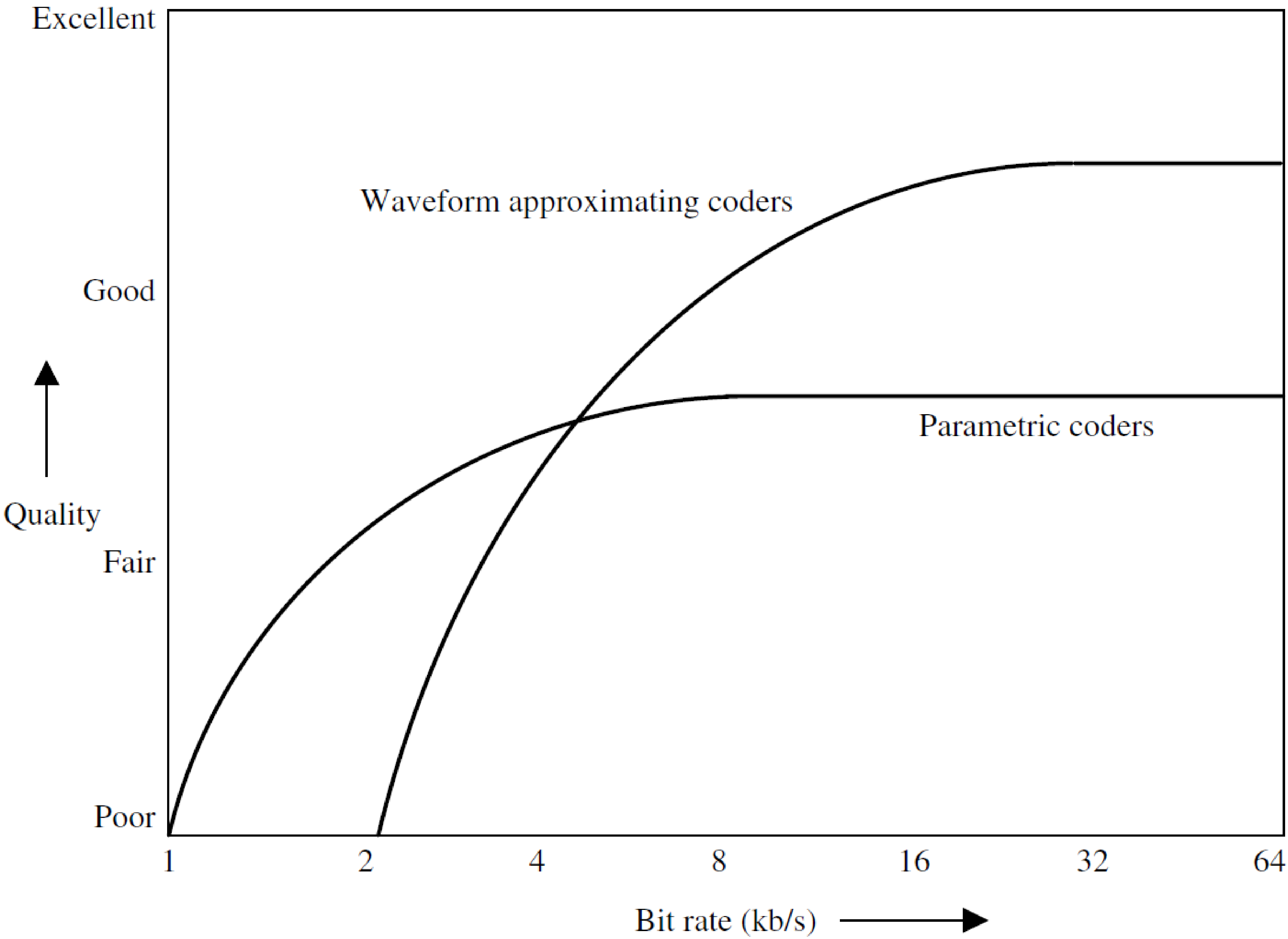

Figure 2.1 Quality vs bit rate for different speech coding techniques

Parametric Coders

- Parametric coders model the speech signal using a set of model parameters
- These coders don’t try to preserve the waveform similarity between the synthesized and the original speech signals
- These coders only preserve the features included in the speech production model, e.g. spectral envelope, pitch and energy contour, etc

Parametric Coders

- The speech quality of parametric coders do not converge towards the transparent quality of the original speech with better quantization of model parameters
- This is due to limitations of the speech production model used

Linear Prediction Based Vocoders

- LP based vocoders emulate the human speech production mechanism
- The vocal tract is modelled by a linear prediction filter
- The glottal pulses and turbulent air flow at the glottis are modelled by periodic pulses and Gaussian noise respectively

Harmonic Coders

- Harmonic coding represents the speech signal as a sum of sinusoidal components
- The model parameters, i.e. the amplitudes, frequencies and phases of sinusoids, are estimated at regular intervals from the speech spectrum
- The frequency tracks are extracted from the peaks of the speech spectra, and the amplitudes and frequencies are interpolated in the synthesis process for smooth evolution

Waveform-Approximating Coders

- Waveform coders minimize the error between the synthesized and the original speech waveforms
- The early waveform coders include companded Pulse Code Modulation (PCM) Adaptive Differential Pulse Code Modulation (ADPCM) etc.
- The recent waveform-approximating coders based on time domain analysis by synthesis such as CELP, explicitly make use of the vocal tract model and the long term prediction to model the correlations present in the speech signal

Hybrid Coding of Speech

- When the bit rate is reduced, the perceived quality of these coders tends to degrade more for some speech segments while remaining adequate for others
- To circumvent this problem, hybrid coders that combine different coding principles to encode different types of speech segments have been introduced
- A hybrid coder can switch between a set of predefined coding modes

Algorithm Objectives and Requirements

- Factors which influence the choice of coding algorithm includes
  - Quality and Capacity: Lowering the bit rate of the speech coder, causes degradation of quality to a certain extent
  - Coding Delay: CD may be algorithmic, computational or due to transmission. CD creates echo problems and subjective annoyance
  - Channel and Background Noise Robustness: By employing built-in robustness, less FEC can be used and higher source coding capacity is available to give better speech quality. Maintaining good speech quality under noisy conditions, good quality background noise regeneration by the coder is also an important requirement

Algorithm Objectives and Requirements

- Factors which influence the choice of coding algorithm includes
  - Complexity and Cost: complexity/power consumption, and hence cost, is still a major problem in portable applications. Silence compression is used to improve the situation
  - Tandem Connection: The ability of an algorithm to cope with tandeming with itself or with another coding system is important
  - Voiceband Data Handling: Voice connections are regularly used for transmission of digital data, e.g. modem, facsimile, and other machine data, an important requirement is an algorithm's ability to transmit voiceband data

Standard Speech Coders

- Standardization is essential in removing the compatibility and conformability problems of implementations by various manufacturers

ITU-T Speech Coding Standard

Table 2.1 ITU-T narrowband speech coding standards

Speech coder	Bit rate (kb/s)	VAD	Noise reduction	Delay (ms)	Quality	Year
G.711 (A/μ-Law PCM)	64	No	No	0	Toll	1972
G.726 (ADPCM)	40/32/24/16	No	No	0.25	Toll	1990
G.728 (LD-CELP)	16	No	No	1.25	Toll	1992
G.729 (CSA-CELP)	8	Yes	No	25	Toll	1996
G.723.1 (MP-MLQ/ACELP)	6.3/5.3	Yes	No	67.5	Toll/ Near-toll	1995
G.4k (to be determined)	4	–	Yes	~55	Toll	2001

European Digital Cellular Telephony Standards

Table 2.2 ETSI speech coding standards for GSM mobile communications

Speech coder	Bit rate (kb/s)	VAD	Noise reduction	Delay (ms)	Quality	Year
FR (RPE-LTP)	13	Yes	No	40	Near-toll	1987
HR (VSELP)	5.6	Yes	No	45	Near-toll	1994
EFR (ACELP)	12.2	Yes	No	40	Toll	1998
AMR (ACELP)	12.2/10.2/7.95/ 7.4/6.7/5.9/ 5.15/4.75	Yes	No	40/45	Toll ~ Communi- cation	1999

North American Digital Cellular Telephony Standards

Table 2.3 TIA/EIA speech coding standards for North American CDMA/TDMA mobile communications

Speech coder	Bit rate (kb/s)	VAD	Noise reduction	Delay (ms)	Quality	Year
IS-96-A (QCELP)	8.5/4/2/0.8	Yes	No	45	Near-toll	1993
IS-127 (EVRC)	8.5/4/2/0.8	Yes	Yes	45	Toll	1995
IS-733 (QCELP)	14.4/7.2/3.6/1.8	Yes	No	45	Toll	1998
IS-54 (VSELP)	7.95	Yes	No	45	Near-toll	1989
IS-641-A (ACELP)	7.4	Yes	No	45	Toll	1996

Secure Communication Telephony

Table 2.4 DoD speech coding standards

Speech coder	Bit rate (kb/s)	VAD	Noise reduction	Delay (ms)	Quality	Year
FS-1015 (LPC-10e)	2.4	No	No	115	Intelligible	1984
FS-1016 (CELP)	4.8	No	No	67.5	Communication	1991
DoD 2.4 (MELP)	2.4	No	No	67.5	Communication	1996
STANAG (NATO) 2.4/1.2 (MELP)	2.4/1.2	No	Yes	>67.5	Communication	2001

Satellite Telephony

Table 2.5 INMARSAT speech coding standards

Speech coder	Bit rate (kb/s)	VAD	Noise reduction	Delay (ms)	Quality	Year
IMBE	4.15	No	No	120	Communication	1990
AMBE	3.6	No	No	–	–	–

Selection of a Speech Coder

- Quality measurements based on SNR can be used to evaluate coders that preserve the waveform similarity, usually coders operating at bit rates above 16 kb/s
- For parametric coders, perception-based subjective measures are more reliable. The Mean Opinion Score (MOS) is a widely-used subjective quality measure
- MOS is costly and Perceptual Evaluation of Speech Quality is a recent method for speech quality assessment.

Selection of a Speech Coder

Table 2.6 Mean Opinion Score (MOS) scale

Grade (MOS)	Subjective opinion	Quality
5 Excellent	Imperceptible	Transparent
4 Good	Perceptible, but not annoying	Toll
3 Fair	Slightly annoying	Communication
2 Poor	Annoying	Synthetic
1 Bad	Very annoying	Bad

Selection of a Speech Coder

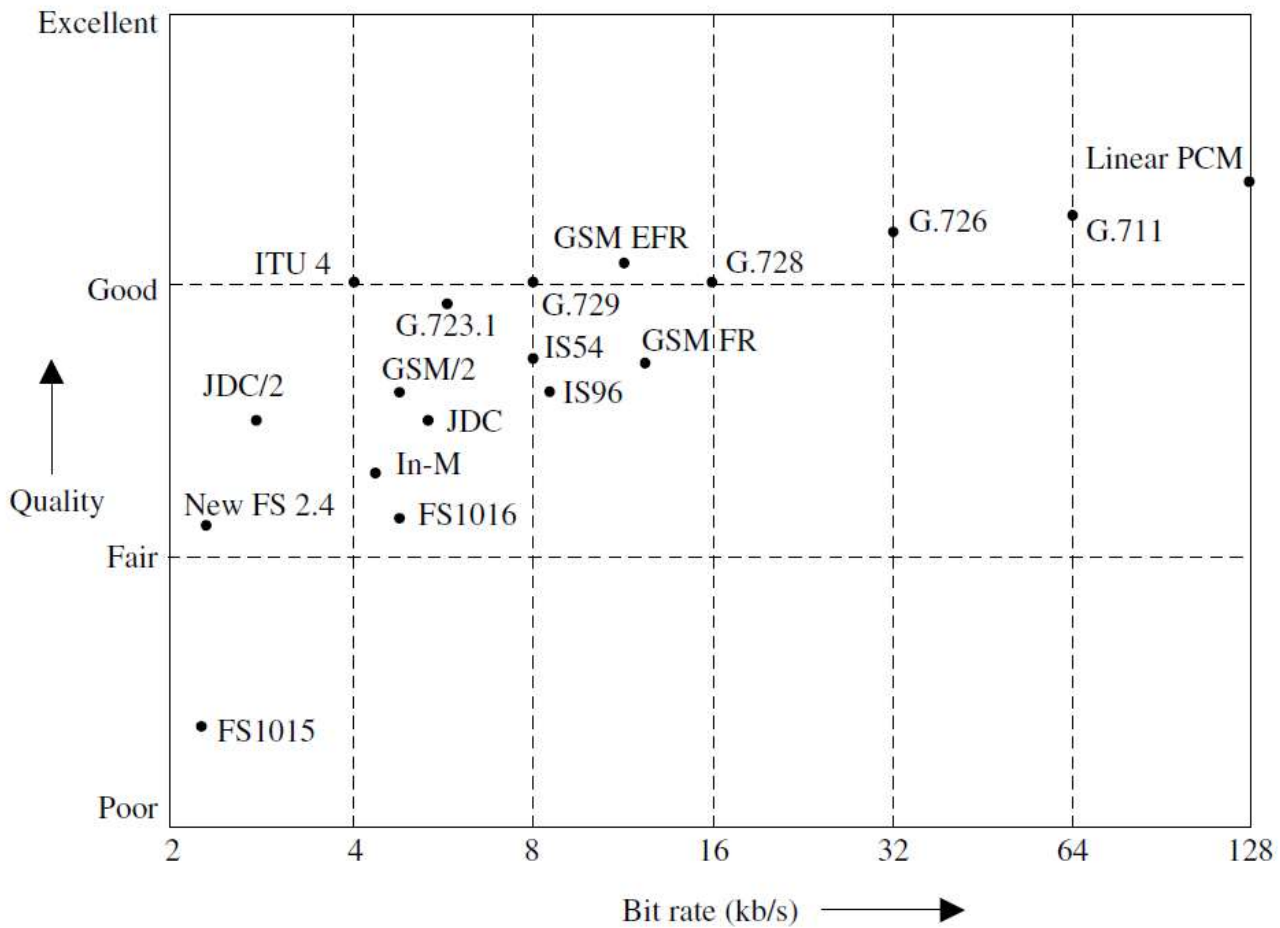

Figure 2.2 Performance of telephone band speech coding standards (only the top four points of the MOS scale have been used)

Selection of a Speech Coder

Table 2.7 Comparison of telephone band speech coding standards

Standard	Year	Algorithm	Bit rate (kb/s)	MOS*	Delay+
G.711	1972	Companded PCM	64	4.3	0.125
G.726	1991	VBR-ADPCM	16/24/32/40	toll	0.125
G.728	1994	LD-CELP	16	4	0.625
G.729	1995	CS-ACELP	8	4	15
G.723.1	1995	A/MP-MLQ CELP	5.3/6.3	toll	37.5
ITU 4	–	–	4	toll	25
GSM FR	1989	RPE-LTP	13	3.7	20
GSM EFR	1995	ACELP	12.2	4	20
GSM/2	1994	VSELP	5.6	3.5	24.375
IS54	1989	VSELP	7.95	3.6	20
IS96	1993	Q-CELP	0.8/2/4/8.5	3.5	20
JDC	1990	VSELP	6.7	commun.	20
JDC/2	1993	PSI-CELP	3.45	commun.	40
Inmarsat-M	1990	IMBE	4.15	3.4	78.75
FS1015	1984	LPC-10	2.4	synthetic	112.5
FS1016	1991	CELP	4.8	3	37.5
New FS 2.4	1997	MELP	2.4	3	45.5

Thank You