Embedded Connectivity Summit 2004

Launched by Motorola

freescale™ semiconductor

TSPG Embedded Connectivity Summit

Review of Voice Compression Algorithms

Wil Yip

Email:
Tel: 480-413-3957
August 20th 2004

Launched by Motorola
freescale™
semiconductor

OUTLINE

Properties of Speech
- Voiced and Unvoiced

Basics of Speech Coding
- Waveform coding
- Parametric coding

Modern Speech Coders
- DSP Architectural Features
- Vocoder Selection Criteria
- Coding Techniques
- Algorithms
- Performance

Voice coding example: VoIP application

Speech Perception

Human ear performs spectral analysis

Critical bands are fundamental measure

- Estimated bandwidth of auditory filter bank
- Low frequencies are more important than high frequencies

Masking

- Exploits limitations of human hearing: signal-to-noise ratio greater than 20 dB in critical band cannot be distinguished
- Noise can be hidden under nearby signal

Voiced vs. Unvoiced Speech

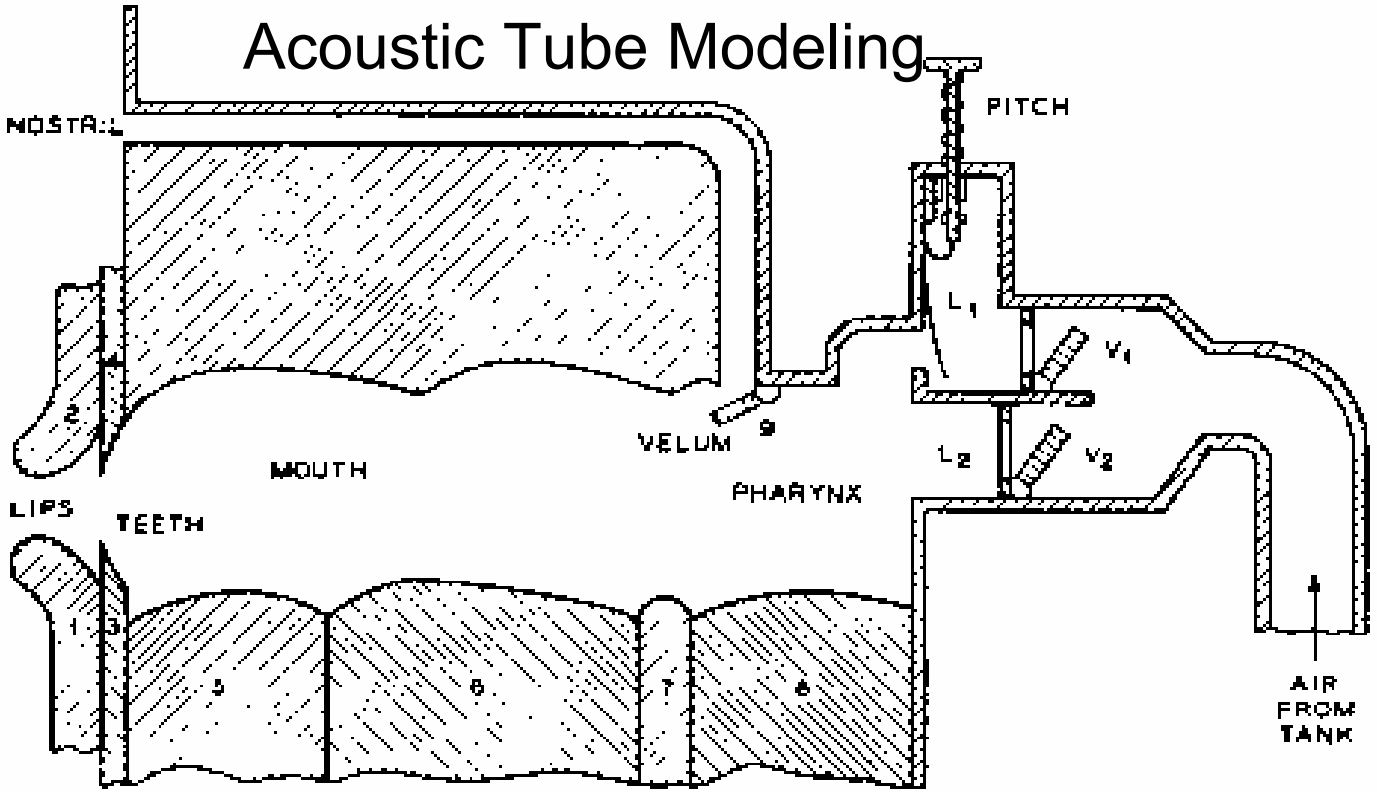
•Speech is composed of phonemes, which are produced by the vocal cords and the vocal tract (which includes the mouth and the lips).

•Voiced signals are produced when the vocal cords vibrate during the pronunciation of a phoneme.

• Unvoiced signals, by contrast, do not entail the use of the vocal cords.

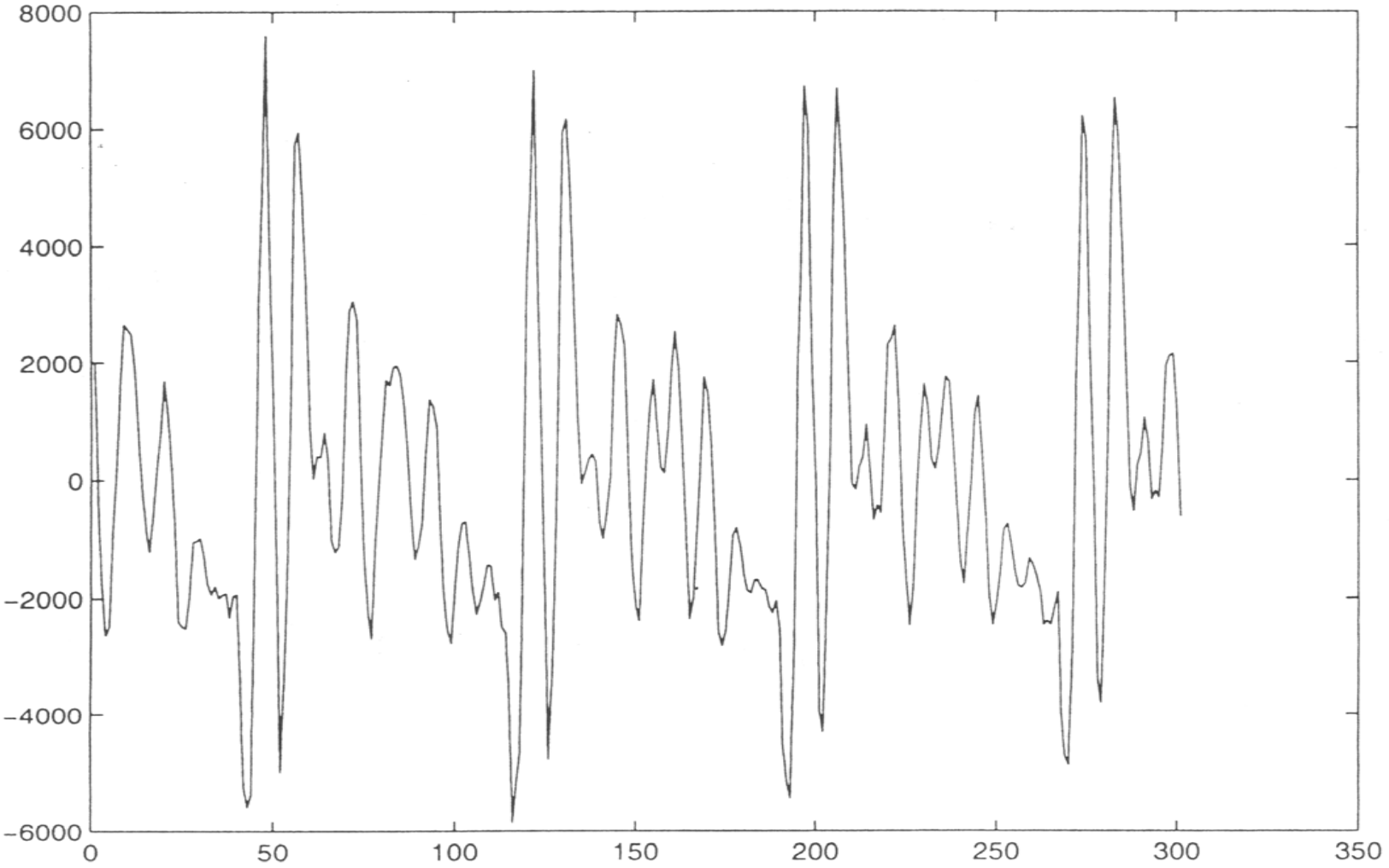
•For example, Voiced signals tend to be louder like the vowels */a/, /e/, /i/, /u/, /o/.* Unvoiced signals, on the other hand, tend to be more abrupt like the stop consonants */p/, /t/, /k/.*

Vocal Track Modeling

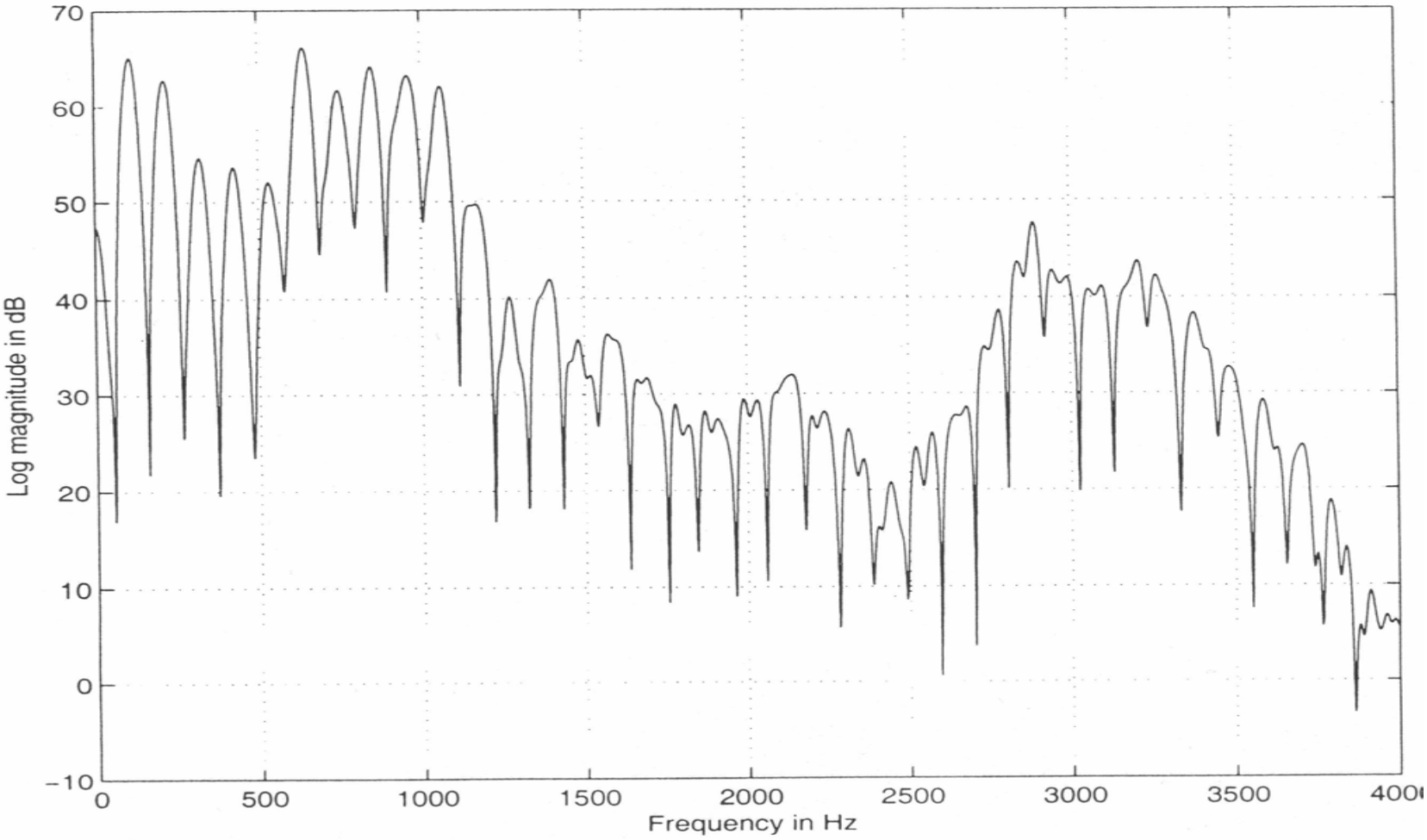


The mechanical model of speech production built by Riesz (1937).

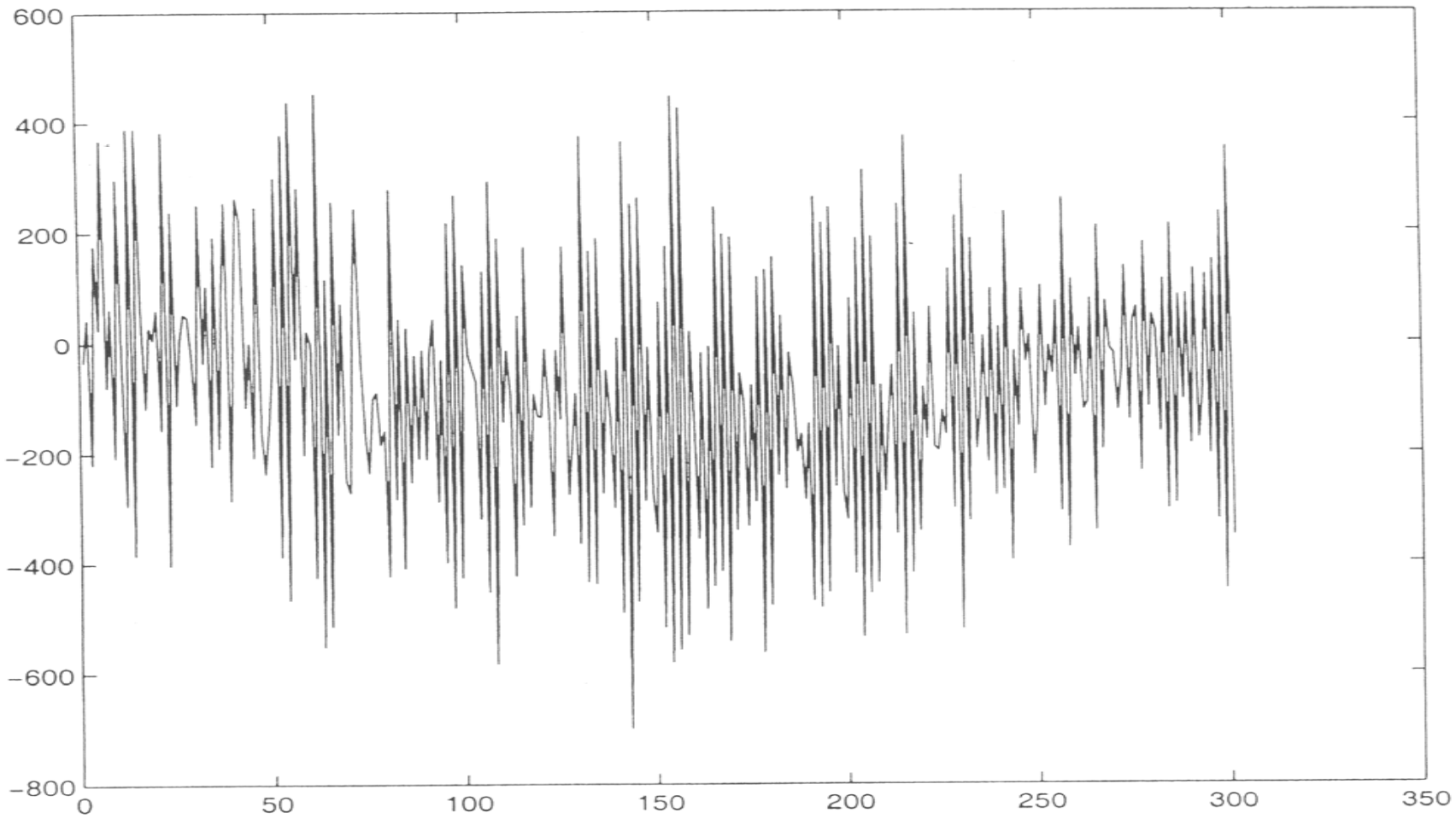
Voiced Speech



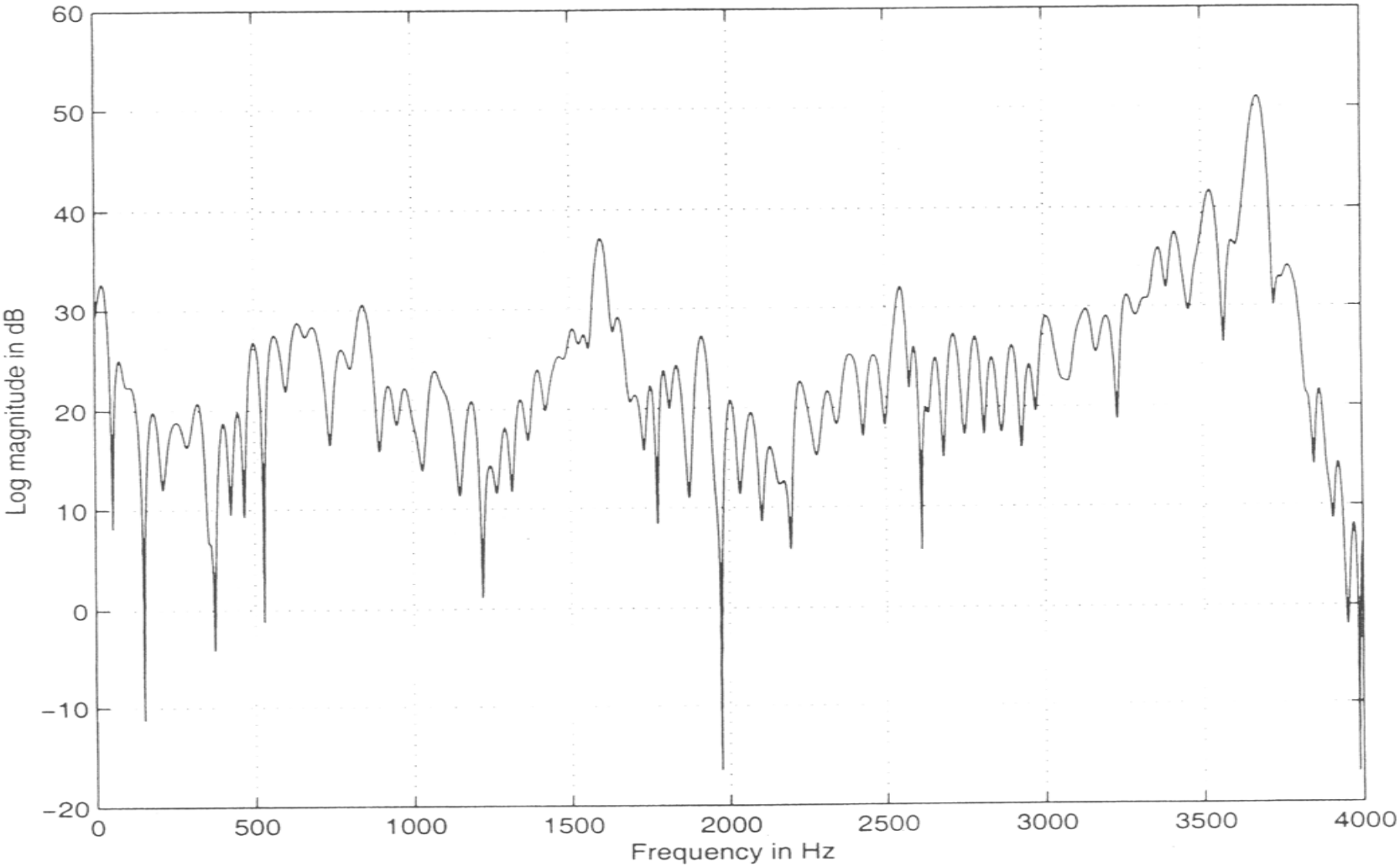
Voiced Speech



Unvoiced Speech



Unvoiced Speech



Why Speech Coding?

Need for Compression

- Efficient communications: Wireless, VoIP
- Digital storage of speech: Answering machines, Synthesis e.g.. TTS

Speech Coders are Efficient

- Speech bandwidth is about 4 KHz
- Input rate: 16 bits X 8 KHz = 128 kbps
- Coder rates: 1 – 32 Kbps
- Compression comes from speech modeling
  - **Best performance only speech signals**
  - **Minimize perceived degradation**

Speech Coders

Exploit Properties of Speech

- Vocal tract, pitch, vocal cords or excitation or residual

Waveform Coders

- Prediction and quantization schemes designed for speech signals
- Match waveform
- Medium bit rates (4 – 64 Kps)
- Examples: PCM, ADPCM, CVSD and etc.

Parametric Coders (Vocoders)

- Based on speech synthesizer model
- Analysis encodes parameters for synthesis
- Match speech characteristics but not exact waveform
- Low bit rates (1 – 4 Kbps)
- Examples: LPC, RELP, CELP, VCELP, and etc.

Low Cost DSP Architecture

Single cycle per operation

True Parallel move capabilities

Combine parallel move with MAC instruction

Zero overhead looping

Modulo addressing

Multiple arrays indexing

Barrel shifter

Special hardware support such as fast FFT etc.

Multiple accumulators

Speech Coding Selection Criteria

Bit Rate

- **Depends on Applications:**
  - Low Bit Rate coders: 1- 4 Kbps
  - Medium Bit Rate Coders: 4 - 16 Kbps
  - High Bit Rate Coders: 16 - 64 Kbps

Speech Quality

- **Subjective**
  - Listening Test (A/B test)
  - Mean Opinion Score (MOS) (G.711 “toll” Quality 4.0)
  - Diagnostic Acceptance Measure/Diagnostic Rhyme test (DAM/DRT)
- **Objective**
  - Signal-to-noise ratio (S/N) per frame
  - Average S/N
  - Segmented S/N

Speech Coding Selection Criteria

Delay

- **In general, we want VERY low Delay – 0 delay?**
  - Delay of 5 – 50 ms voice processing
  - Delay of 5 – 150 ms buffer delay
  - Delay will affect echo cancellation
  - Application specific e.g. Satellite Communication, cellular, VoIP & etc.

Complexity

- **MIPS and memory**
  - 1 to 50 MIPS (single instruction per operation)
  - 1 - 30 K program & 1 - 30 K Data memory
- **Implementation**
  - Fixed-point vs. Floating-point,
  - Parallel moves,
  - mostly assembly language,
  - Hand coded – fine tuning

Speech Coding Selection Criteria

Error Protection Scheme

- **Build-in error correction**
  - Selective protection
  - Error correction and Error detection
- **Depends on channel coder**
  - Selective Detection
  - Channel coder (FEC) for the rest

Tandeming Performance

- **Going through different systems**
  - e.g. CDMA – GSM
  - Quality degraded due to digital to analog to digital conversion

Standard Compliance

- **ITU standards e.g. G.711, G.726, G.729, G.723.1 and etc.**
- **Wireless standards e.g. GSM – FR, HR, EFR, EVRC & etc.**

Coding Techniques

Waveform Coding Techniques:

- **Designed to reproduce the time input signal**
- **Higher quality at low compression ratios**
- **Most of the very high quality audio compression techniques are waveform coding**
- **Examples of Voice Waveform coders:**
  - Pulse Code Modulation (PCM) G.711
    - > Log vs. Linear Quantization
  - Differential PCM
  - Adaptive DPCM (ADPCM) G.726
  - Continuous Variable Slop Delta Modulation (CVSD)

Coding Techniques

Parametric Coding Techniques:

- **Using speech production models such as Acoustic Tube Modeling**
- **Generally, speech only and not for audio compression**
- **Usually, no waveform matching**
- **Sending only parameters such as:**
  - LPC coefficients
  - Pitch period and energy
  - Code book index and energy
- **Very high compression ratio and low bit rate: e.g. ~ 1- 8 kbps**
- **Examples of parametric coders:**
  - Analysis by Synthesis
    - > LPC, CELP, RELP, RPELP. ACELP, VSELP, EVRC, LD-CELP MELP and etc.
  - Others:
    - > Multi-band Excited (MBE, IMBE, AMBE)
    - > Sinusoidal Transform Coding (STC)
    - > . . .

Waveform Coding Techniques

PCM or G.711

- Pulse Code Modulation – Code amplitude of each sample
- Non Linear scale: u-law or A-law (companding) techniques
- 8-bit log quantizers and very low delay
- 16-bit linear quantizers used as input to most of the vocoders
- Higher bit rates 64Kbps and lower cost

DPCM and ADPCM G.726

- Code difference from previous sample
- Predictive coding – difference is from estimate of current sample based on multiple previous samples
- Quantization step-size and Prediction coefficients can be adaptive

CVSD

- Code Delta difference from previous multiple samples
- Delta step-size can be adaptive

Parametric Coding Techniques

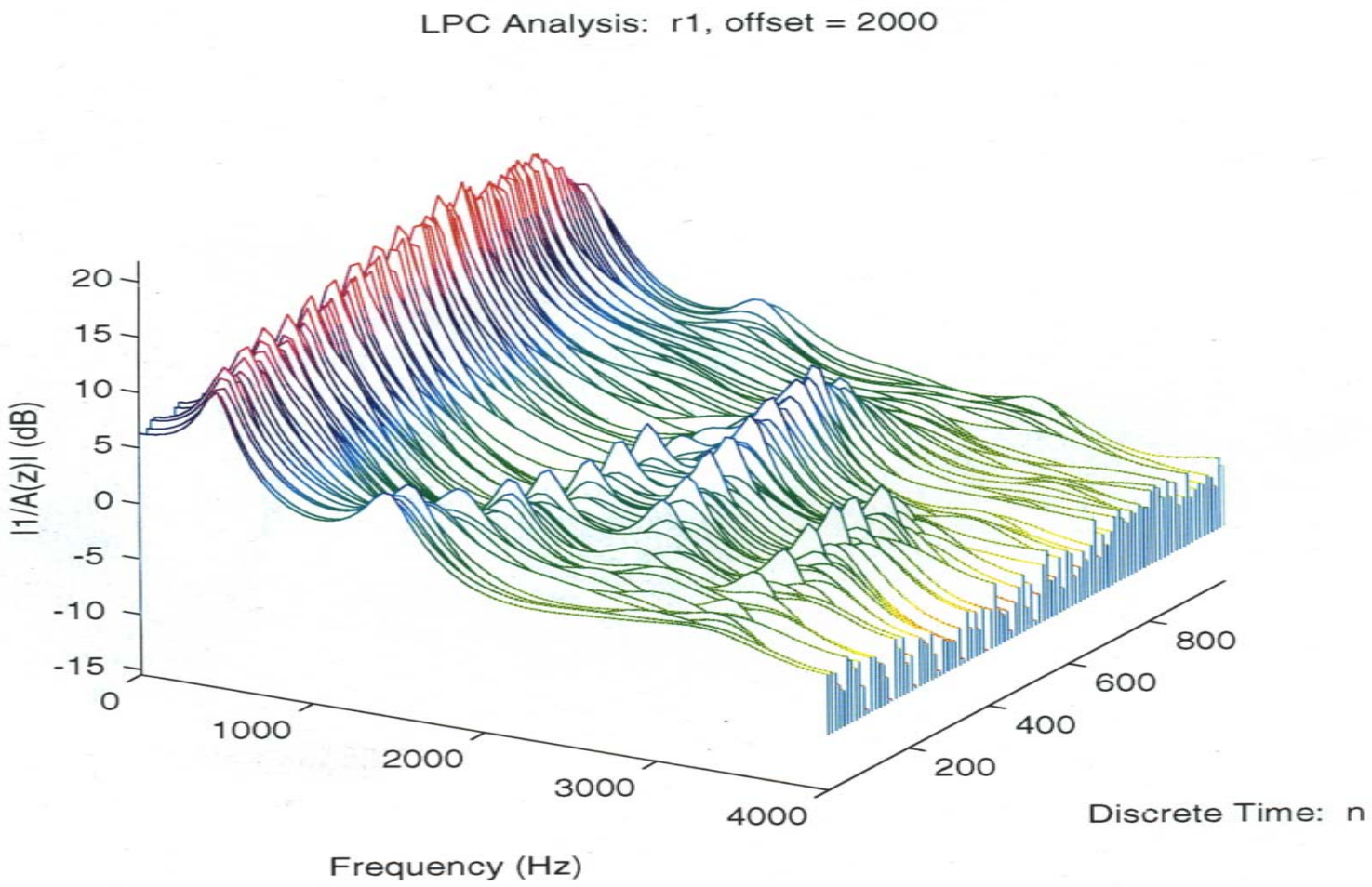
Linear Predictive Coding (LPC-10)

- Vocal Track and Cords modeling
- Short Term prediction – spectrum analysis
- Long Term prediction – pitch analysis
- White noise excitation
- 2400 bps

Linear Predictive Analysis-by-Synthesis

- Close-loop analysis
- Generate coded speech for each possible excitation
- Choose minimum squared error
- Vector quantization modeling
- Very HIGH complexity
- Medium bit rates (4 – 16 Kbps)

LPC Analysis



Linear Predictive Analysis-by-Synthesis

Excitation Types

- **Residual Excitation (RELP)**
  - Vector quantized residual
  - Residual vector contains amplitude and phase information
- **Multi-pulse (MP)**
  - Excitation is set of pulses
  - Remaining samples are zero
  - Pulses are sequentially optimized
  - May have different amplitudes
- **Regular-pulse (RPE)**
  - Multi-pulse with constrained positions
- **Codebooks (CELP)**
  - Stochastic: Gaussian sequences
  - Trained excitation vectors
  - Algebraic: equal-amplitude pulses (ACELP)

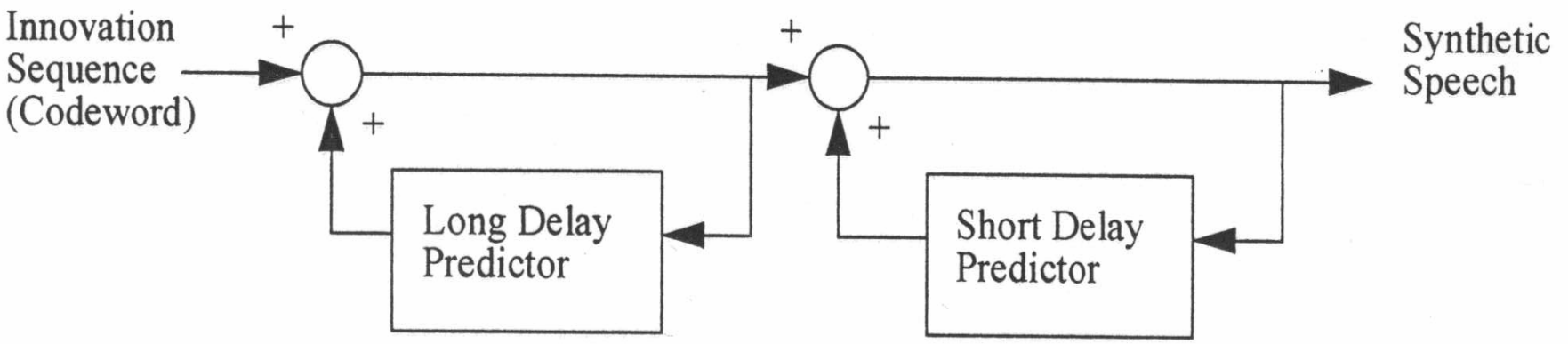
Parametric Coder Algorithms

USFS-1016 CELP

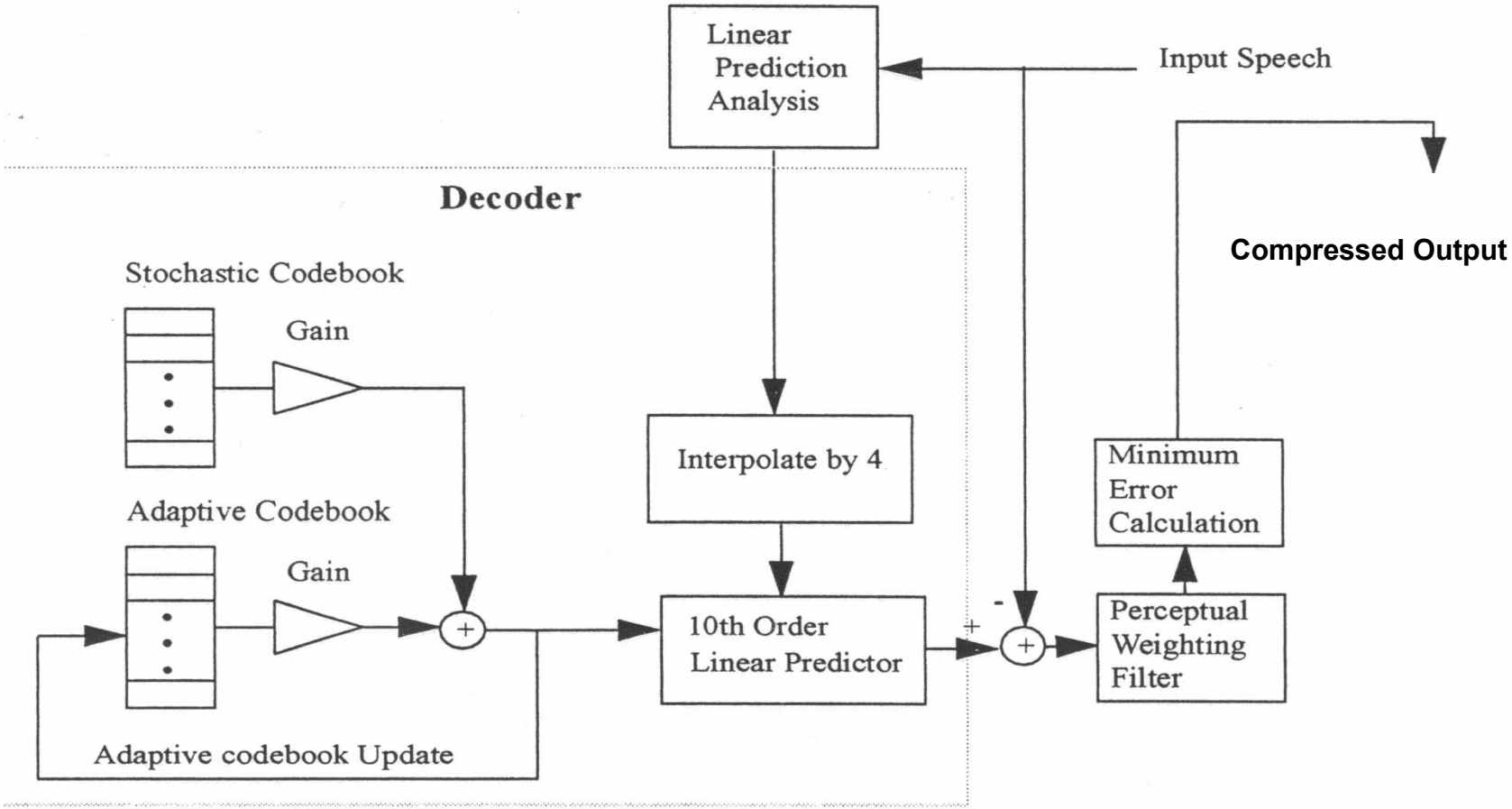
- Vector Quantization Technique
- Frame based processing with four sub-frames
- 10^{th} order LPC filter for the Short-Term Spectrum
- Adaptive Codebook VQ to Model Pitch
- Fixed Stochastic Codebook to vector quantize Residual of Short-Term and Long-Term pitch VQ
- Uses Perceptually Weighted Distortion Measure to Select optimal code words

Code-Excited Linear Prediction (CELP)

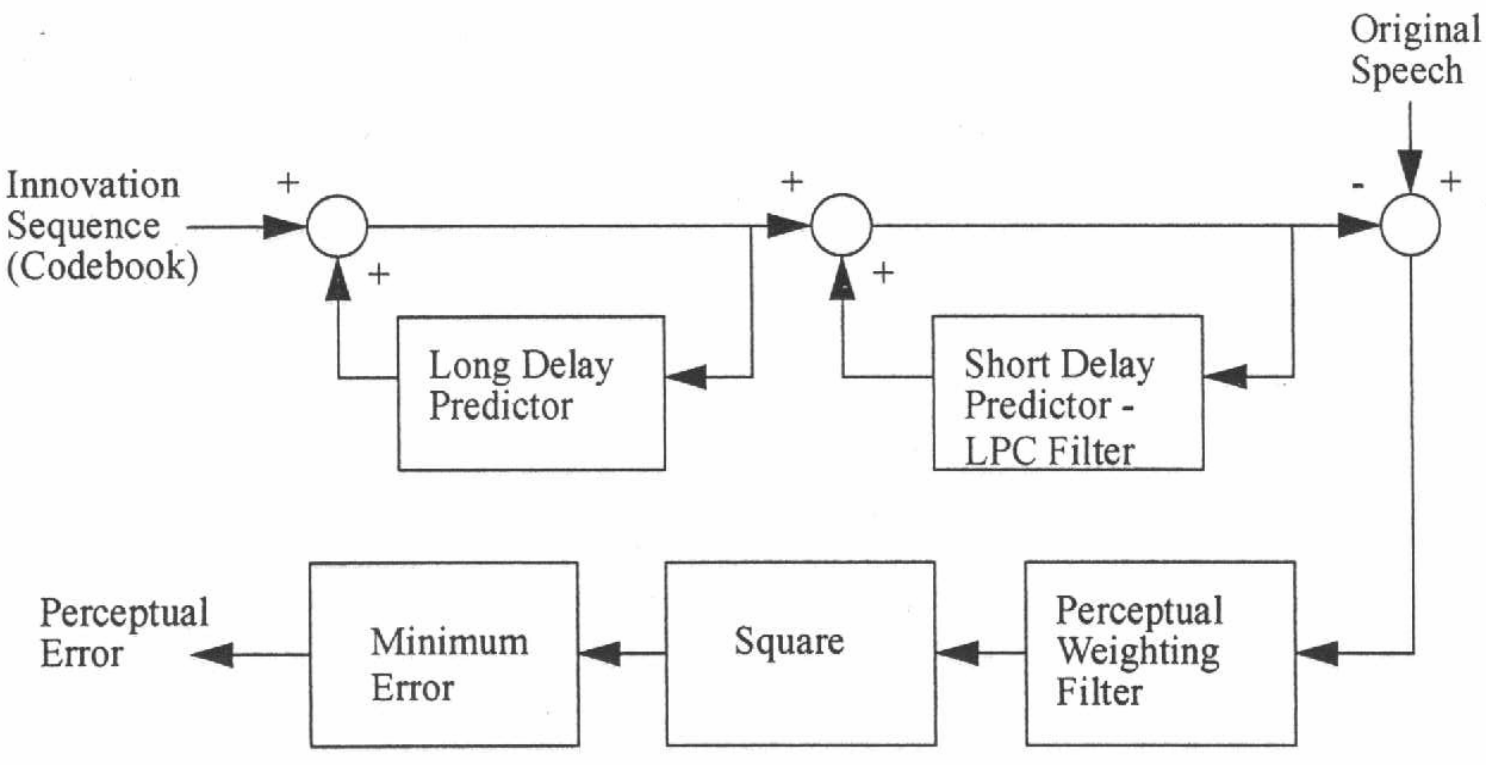
CELP Speech Synthesis Model



Typical CELP Encoder Block Diagram



Analysis by synthesis Method

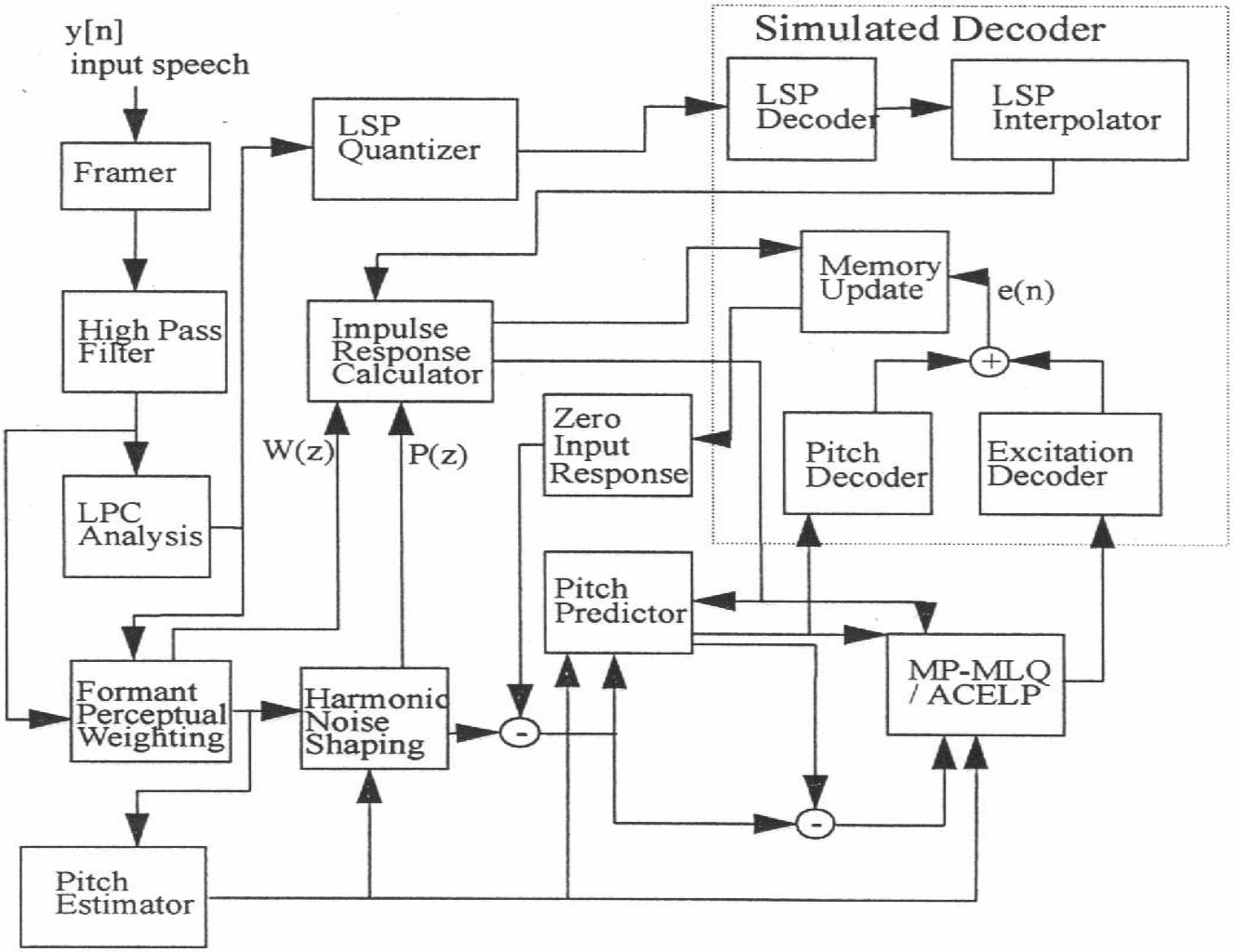


Parametric Coder Algorithms

ITU G.723.1a MP-MLQ/ACELP

- Dual-rate speech coder using multipulse maximum likelihood quantization (MP-MLQ) excitation for 6.3 Kbps and Algebraic code excited linear prediction (ACELP) for 5.3 Kbps.
- 30ms frame with four sub-frames of 7.5 ms each
- Complexity ~ 30 MIPS

G.723.1a Encoder Block Diagram

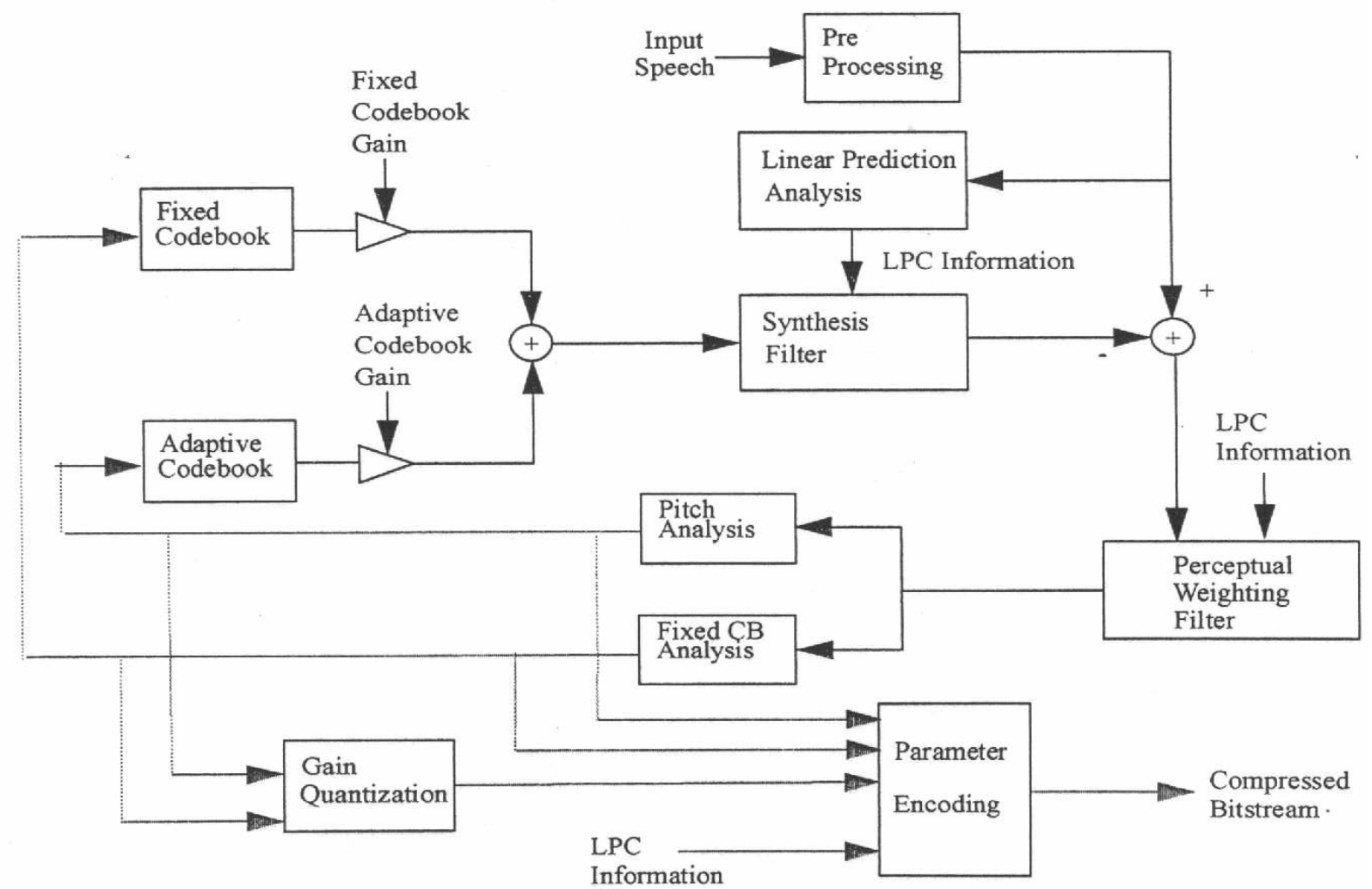


Parametric Coder Algorithms

ITU G.729a/b CS-ACELP

- 10ms frame with two 5ms subframes
- Open Loop pitch analysis first and optimize with closed-loop pitch together with codebook gain
- Spectral Envelope: differentially quantized with 18 bits
- Excitation: adaptive (pitch) and fixed codebook
- Gains: vector quantized
- Adaptive post-filtering
- Moderate complexity ~20 MIPS

G.729a/b CS-ACELP Encoder Block Diagram

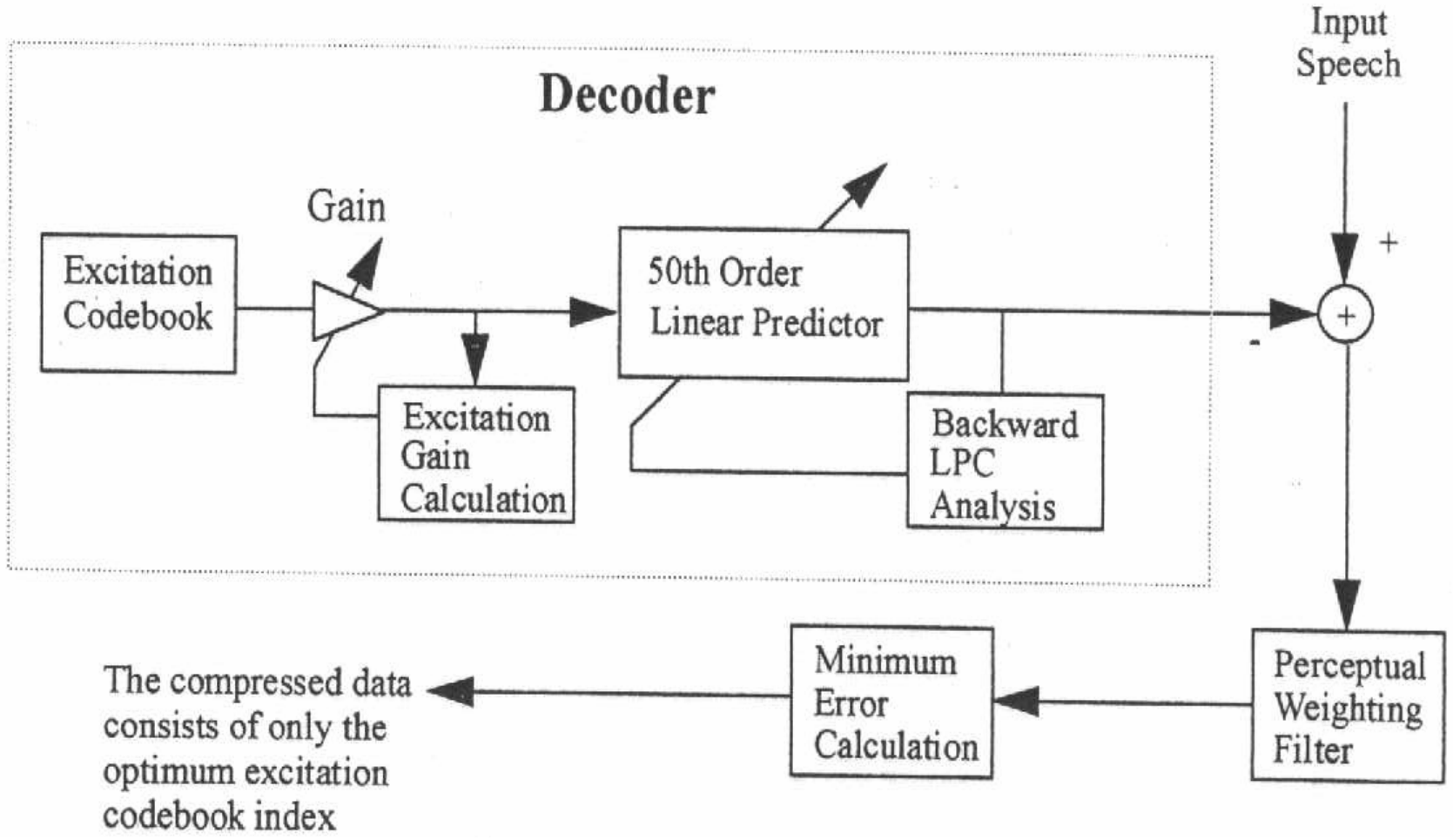


Parametric Coder Algorithms

ITU G.728 16 Kbps LD-CELP

- Low-delay code excited linear prediction (LD-CELP)
- Very short frame results in low delay
- 2.5 ms frame with four sub-frame of 625 ms each (5 samples each)
- 50^{th} order LPC filter to model both short-term Spectrum and Long-Term pitch
- Uses Backward Adaptation of Gain and LP Coefficients
- Uses Fixed Stochastic Codebook (128 code words) and 3-bit gain to vector quantize residual of Short-Term LP
- 10-bit code vectors are transmitted over the channel

G.728 LD-CELP Encoder Block Diagram



Parametric Coder Algorithms

IS54-B VSELP

- Vector Sum Excited Linear Prediction (VSELP)
- Operates on 20ms frame with four sub-frames
- CELP based technology similar to USFS-1016
- Codebook are sequences of vector pulses
- Vector can be added or subtracted to form code excitation

Parametric Coder Algorithms

Other Vocoders:

Sinusoidal Transform Coder

- Speech is modeled as sum of sine waves
- Amplitudes are quantized, not phases
- Variable cutoff frequency: high-frequencies are replaced with random noise signal

Multi-band Excited (MBE,IMBE,AMBE)

- Different form o sinusoidal model
- Up to 15 frequency bands each with binary voicing decision

Sub-band coders:

- Divide the speech spectrum into different bands of frequency
- Each band is coded based on perceptual sensitivities

Mean Opinion Score (MOS)

Score	Opinion Scale	Listening: Effort Scale
5	Excellent	No effort required
4	Good	No appreciable effort required
3	Fair	Moderate effort required
2	Poor	Considerable effort required
1	Bad	Difficult to understand

Performance

Coding Technique	Standard	Bit Rate	Speech Quality (MOS)	Complexity (MIPS)	Full-Duplex Delay (ms)
A-Law & U-Law PCM	G.711	64Kbps	4.0	0	0
ADPCM	G.726	16/24/32/40Kbps	4.1	10-16	0
LPC	USFS-1015	2.4Kbps	2.3	6-12	65
CELP	USFS-1016	4.8Kbps	3.2	13-25	105
VSELP	IS-54B/GSM HR	8.0Kbps	3.5	20-25	60
LD-CELP	G.728	16KBPS	4.0	25-45	7.5
MP-MLQ	G.723.1a	6.4/5.3Kbps	3.9	22-28	100
CS-ACELP	G.729	8Kbps	4.0	20-28	30
GSM-EVRC	ACELP	13.3-6.2-2.7-1.0Kbps	4.0	25-35	45

MIPS and Memory Requirements for HAWK (VoIP)

Components	Program	Data Memory				MIPS
	Code (words)	Total (words)	Tables (words)	Stack (1) (words)	Static (2) (words)	
G.711	164	26	0	26	0	0.50
G.726	1,205	382	220	16	146	15.58
G.723.1A	7,749	11,745	9,439	1,366	940	23.60 (peak) 18.50 (typical)
G.729AB	8,516	4,911	2,665	1,108	1,138	12.97
G.168 (16 ms echo span)	1,666	1,035	237	20	778	6.50
Overhead (Estimate)	4,000	1,200	0	200	1,000	4.00

Notes:

(1) Stack (Scratch or Local) Data Memory is per System Memory Requirement

(2) Static (Global) Data Memory is per Channel Data Memory Requirement

VoIP Voice processing Bundling

		DSP56852	DSP56853	DSP56854	DSP56855	DSP56857	DSP56858
	Program RAM	6K	12K	16K	24K	40K	40K
Voice Processing	Data RAM	4K	4K	16K	24K	24K	24K
Bundle #		Wedge	Leonard	Sutton	Duval	Sawgrass XL1	Sawgrass XL
(1) G.711/G.726/G.168		**2**	**3**	**4**	**5**	**7**	**12**
(2) G.711/G.723.1a/G.168		**0**	**0**	**2**	**3/4**	**3/4**	**3/4**
(3) G.711/G.729ab/G.168		**0**	**0**	**4**	**4**	**4**	**4**
(4) G.711/G.726/G.723.1a/G.729ab/G.168		**0**	**0**	**0**	**4**	**4**	**4**
(5) G.711/G.168		**2**	**3**	**8**	**16**	**16**	**16**

Assuming:

1) 4 MIPS per Channel Overhead Processing Estimate

2) Overhead Memory Estimate as shown below

3) Running All Internal Memory