DESIGN OF A CELP CODER AND A STUDY OF ITS PERFORMANCE USING VARIOUS QUANTIZATION METHODS

EECS 651: PROJECT PRESENTATION

UNIVERSITY OF MICHIGAN, ANN ARBOR

APRIL 18, 2005

By

Awais M. Kamboh

Krispian C. Lawrence

Aditya M. Thomas

Philip I. Tsai

PROJECT GOALS

- To design and implement a CELP coder in matlab
- To use different quantization methods to quantize the LP parameters of the coder
- To evaluate the performance of the coder in terms of MSE and 'perceptual MSE' using the various methods of quantization

Presentation Outline

- Introduction to Speech coding
- CELP
- CELP coder
- Quantization Methods
- Results and Comparisons
- Conclusions and recommendations
- Q&A

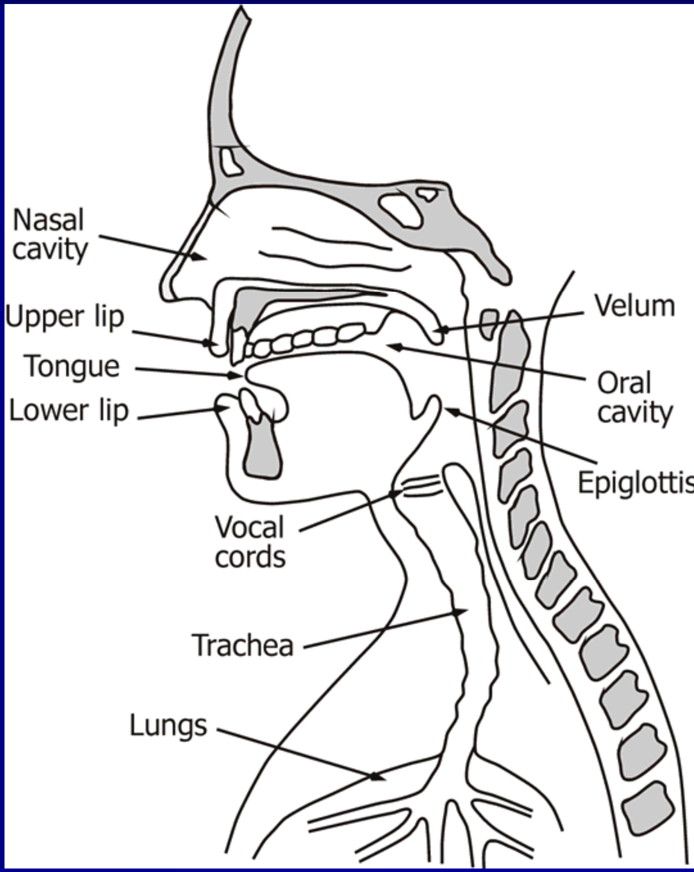
Introduction to Speech Coding

- Concerned with obtaining compact digital representation of voice signals for more efficient transmission or smaller storage size.

- Objective is to represent speech signal with minimum number of bits yet maintain the perceptual quality.

Speech Production

- Speech
  - Air pushed from the lungs past the vocal cords and along the vocal tract
  - The basic vibrations – vocal cords
  - The sound is altered by the disposition of the vocal tract (tongue and mouth)
- Model the vocal tract as a filter
  - The shape changes relatively slowly
- The vibrations at the vocal cords
  - The excitation signal



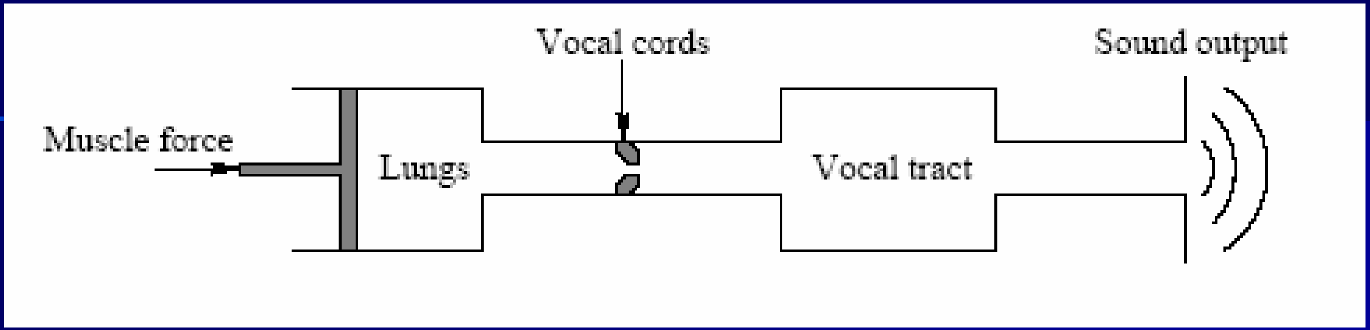
Speech sounds

- Voiced sound
  - The vocal cords vibrate open and close
  - Quasi-periodic pulses of air
  - The rate of the opening and closing – the pitch
- Unvoiced sounds
  - Forcing air at high velocities through a constriction
  - Noise-like turbulence
  - Show little long-term periodicity
  - Short-term correlations still present
- Plosive sounds
  - A complete closure in the vocal tract
  - Air pressure is built up and released suddenly

Code-Excited Linear Predictor (CELP)

- Variants of CELP (LD-CELP, ACELP etc.)
- Main difference in generation of excitation signal, Filters and Bit rate.
- Performance
  - 4kbps or lower bit-rates give synthetic quality speech / mechanical speech.
  - Most modern CELP variants produce relatively higher bit-rates and good quality speech.
  - Performance cannot be judged by MSE alone.

Linear Predictive Coding.



- Lungs generate an excitation signal which is modeled as white noise.
- Vocal cords either remain open or vibrate with some frequency, called ‘Pitch’.
- The resulting speech is either unvoiced or voiced respectively.
- Vocal tract acts as an IIR filter.

CELP Parameters (In this Implementation)

- **Excitation Signal:** A number of signals are stored in a codebook. We choose the signal that best suits a particular chunk of data (frame).
- **LP Coefficients:** The coefficients of vocal tract filter.
- **Gain**: Represents the loudness/energy of speech.
- **Pitch Filter Coefficient**: We determine pitch by modeling it as a long delay correlation filter which produces quasi-periodic signals when excited.
- **Pitch:** Pitch of the sound. In the range 50Hz to 500Hz. In this case it is referred to as Pitch Delay measured in # of samples

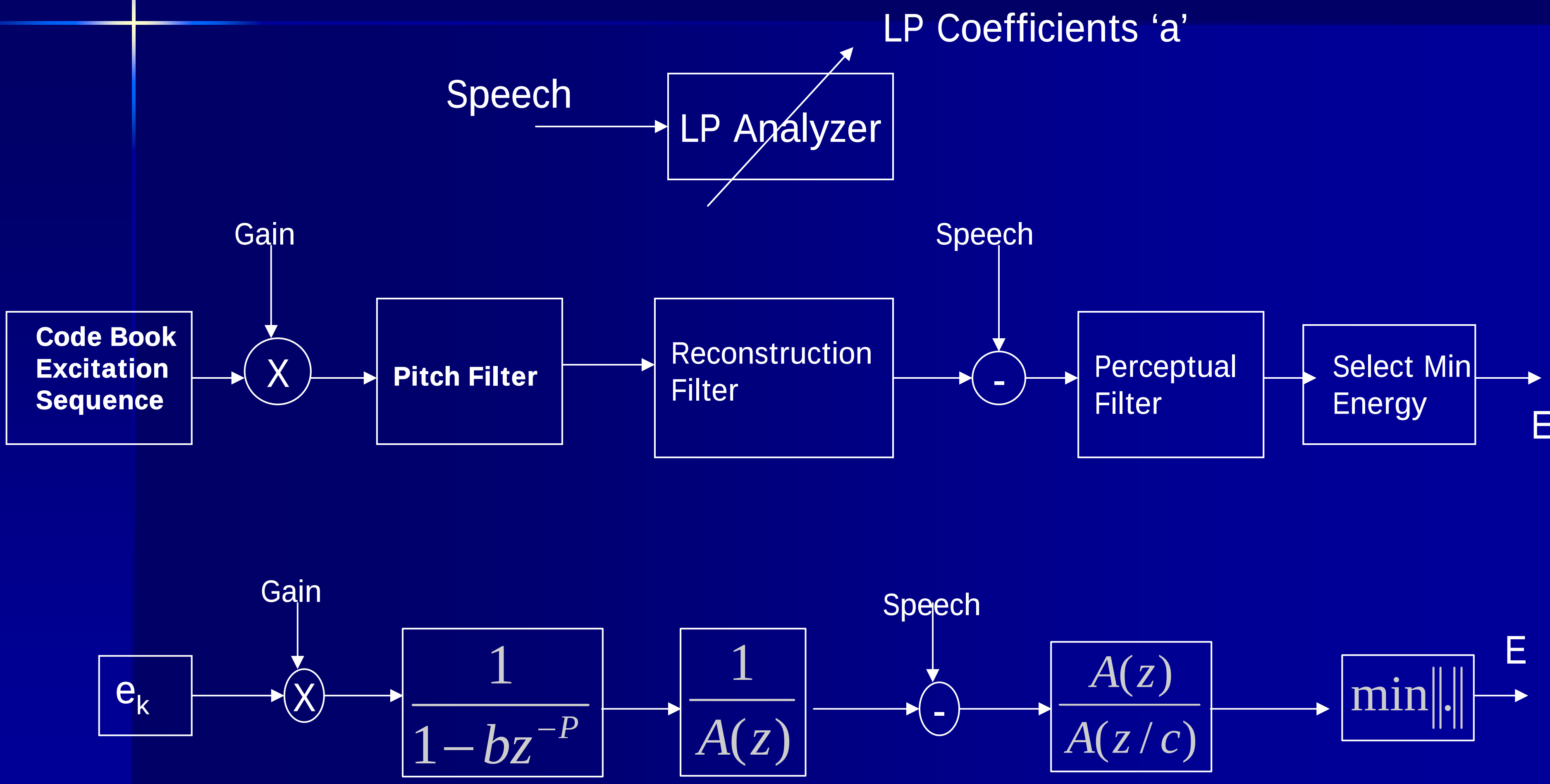
Rate of CELP

Frame Size: 160 samples. (20 ms)
Subframe Size: 40 samples (5 ms)

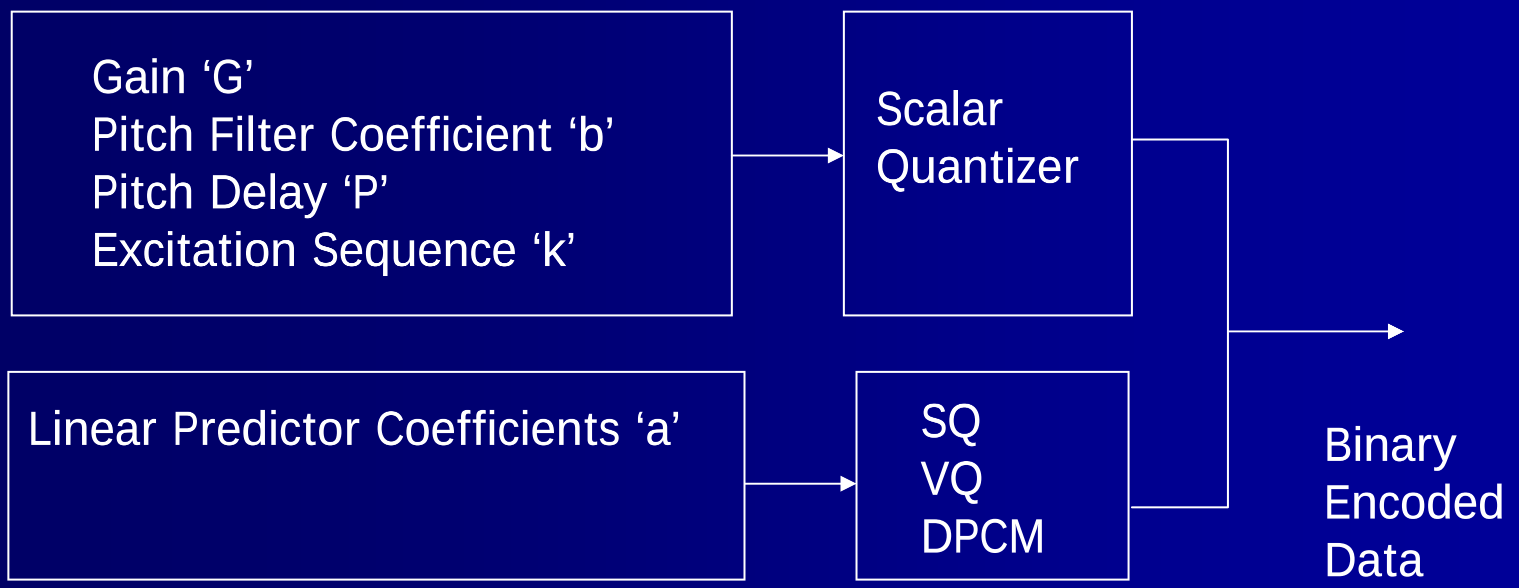
LP coefficients are transmitted once per frame. All others are transmitted once per subframe.

Code Book : 512 entries; 9 bits
Gain: Generally between -2 to +2: 8 bits
Pitch: 50Hz to 500Hz =>
    16 to 160 samples (at 8KHz Sampling): 8 bits
Pitch filter Coeff: 0 to 1.4: 6 bits
LP Coefficients: Different for different Rates.

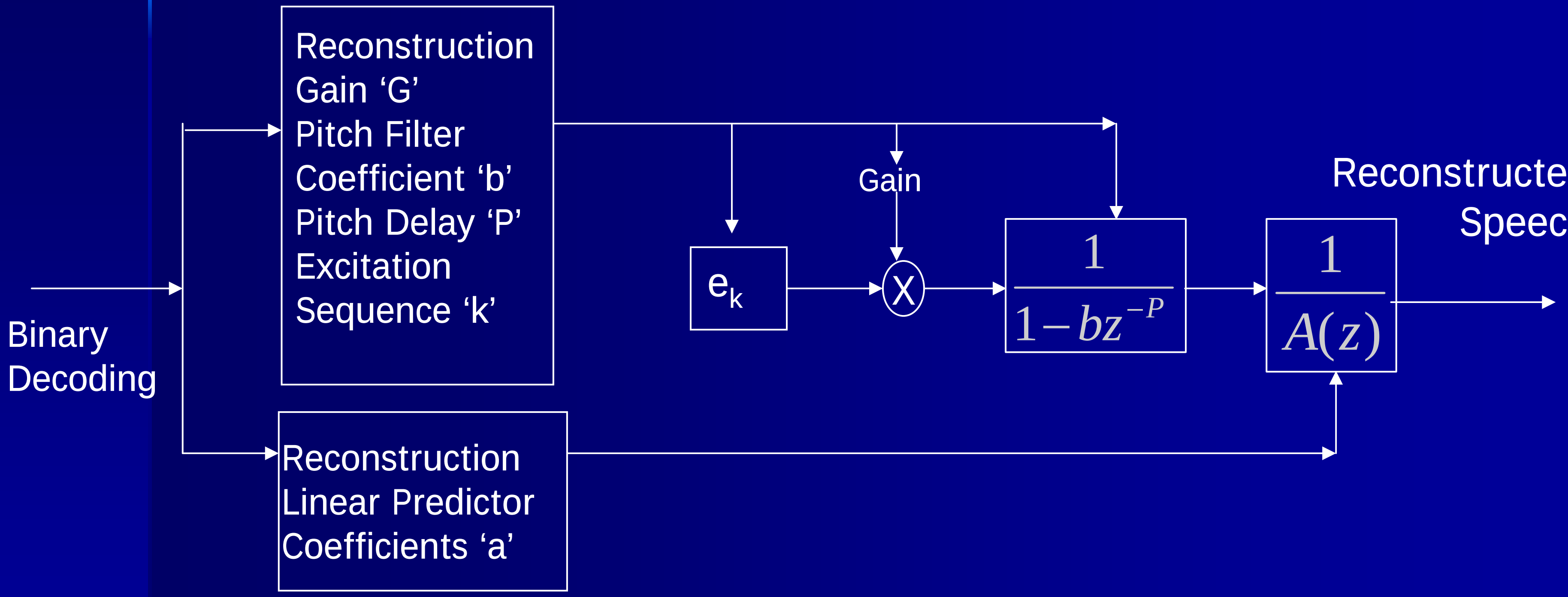
CELP Encoder



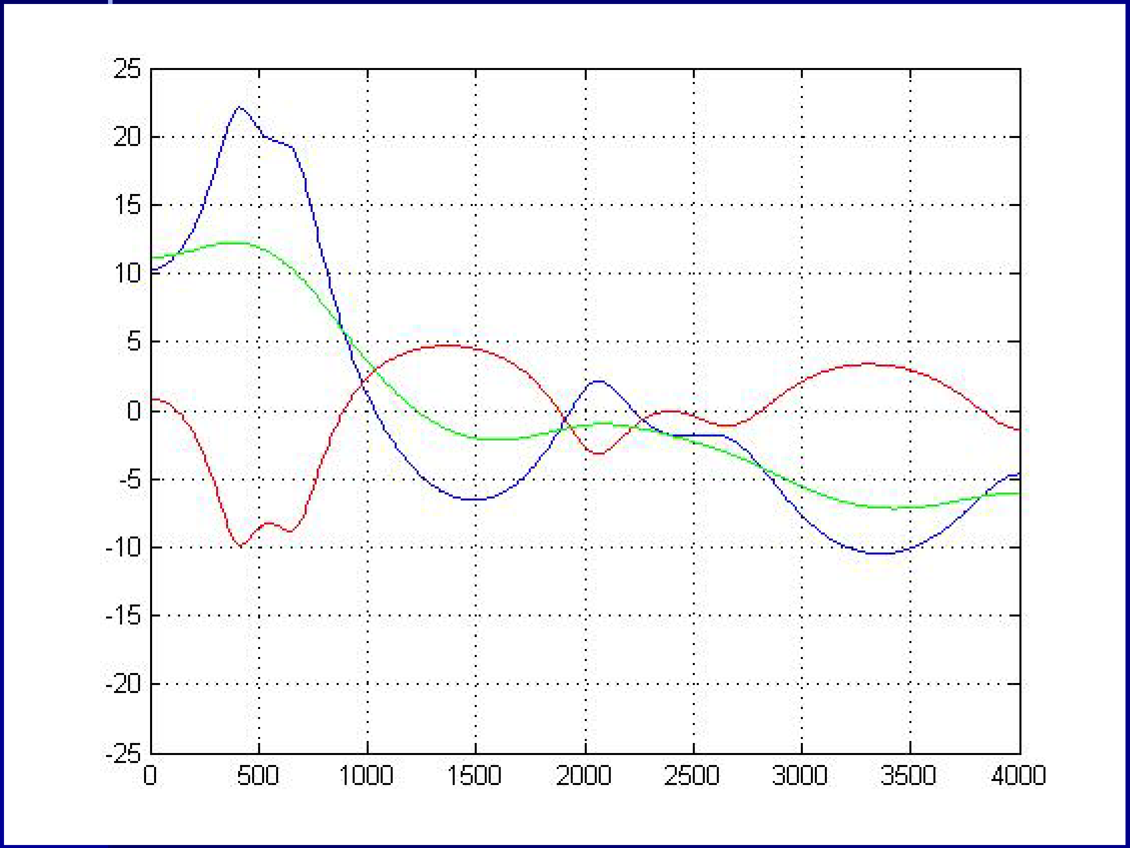
CELP Encoder (Contd.)



CELP Decoder



Perceptual Filtering



Frequency (Hz)

$c = 0.8$

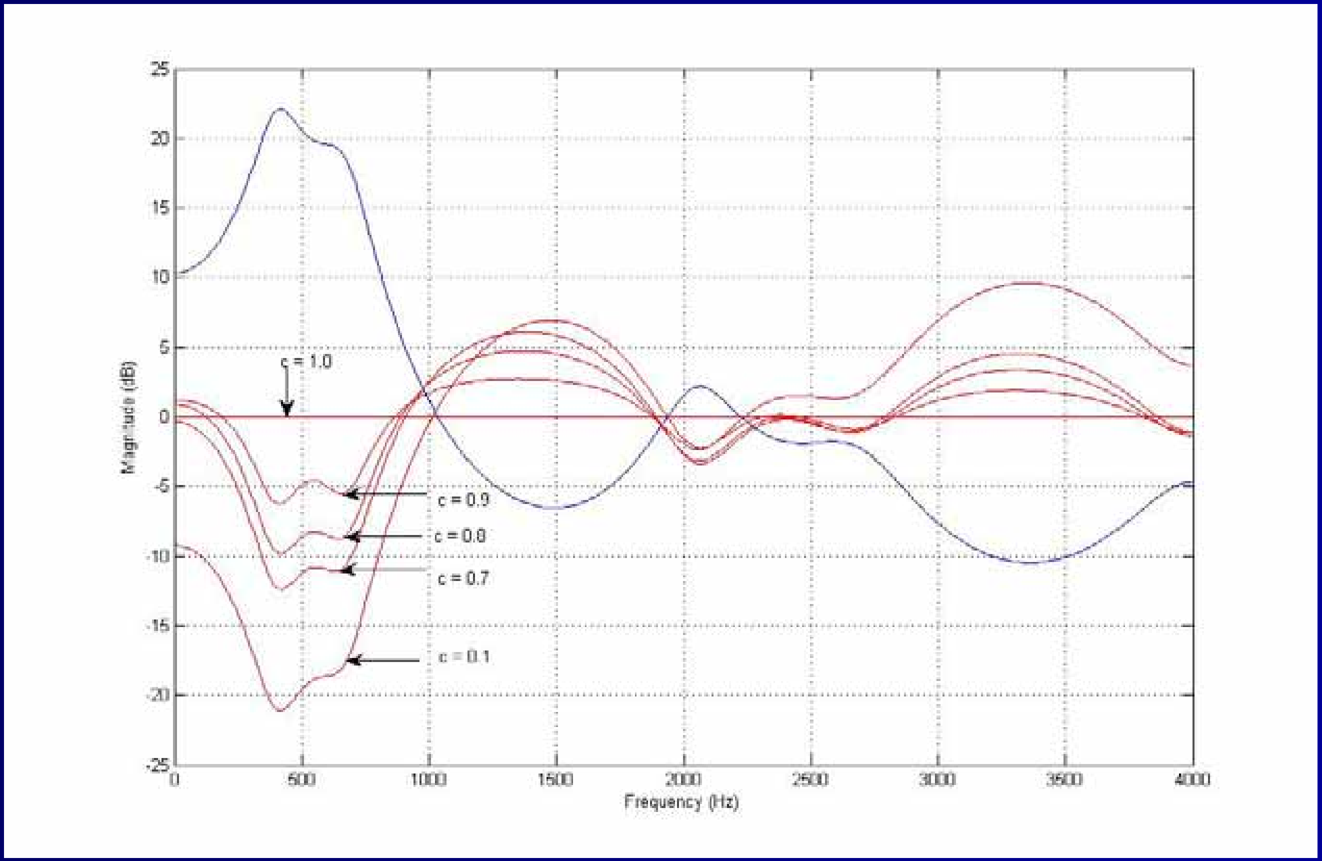
$$red = \frac{A(z)}{A(z/c)}$$

$$green = \frac{1}{A(z/c)}$$

$$blue = \frac{1}{A(z)}$$

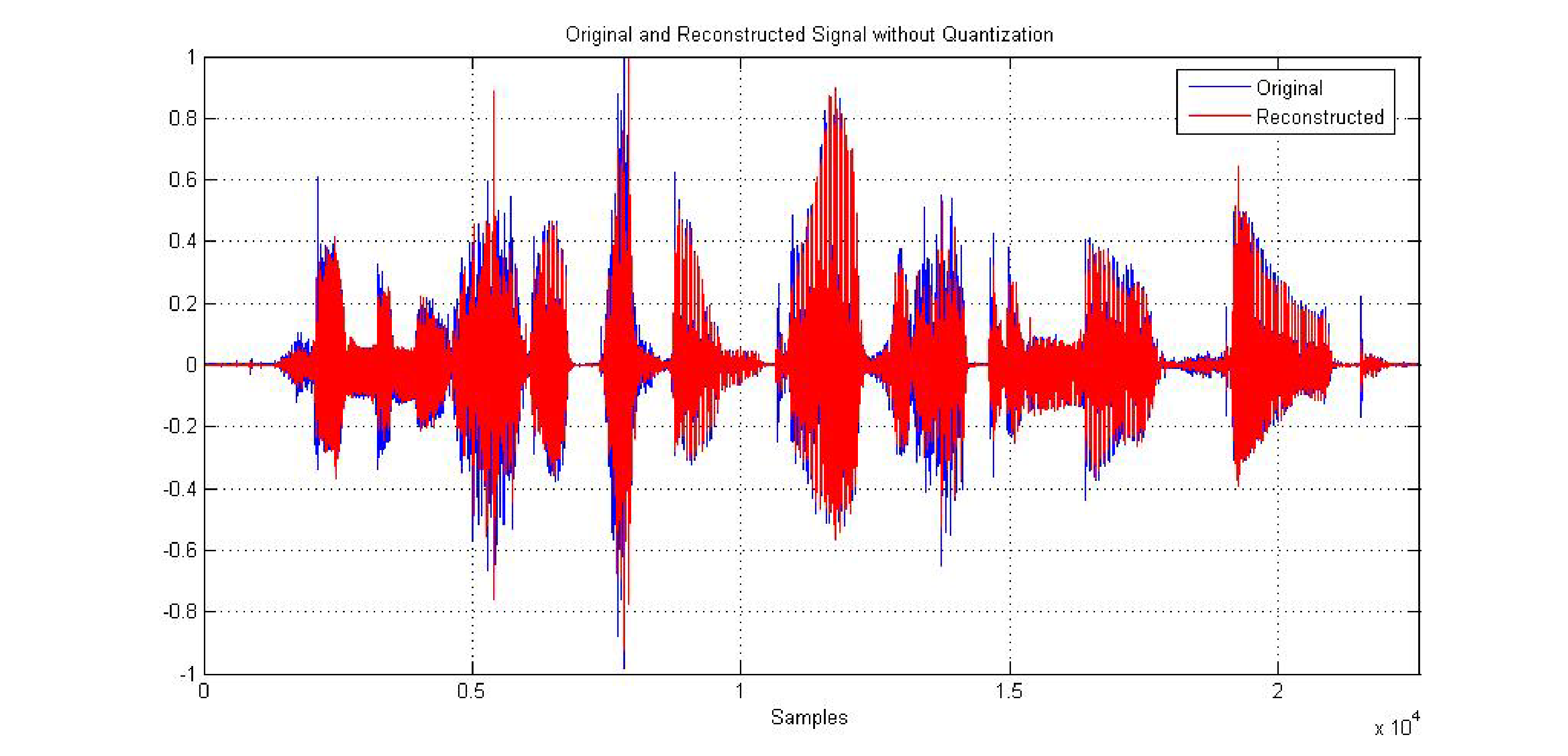
Perceptual Filtering (Contd.)

$$\frac{A(z)}{A(z/c)}$$



Different values of 'c' in Perceptual filter.

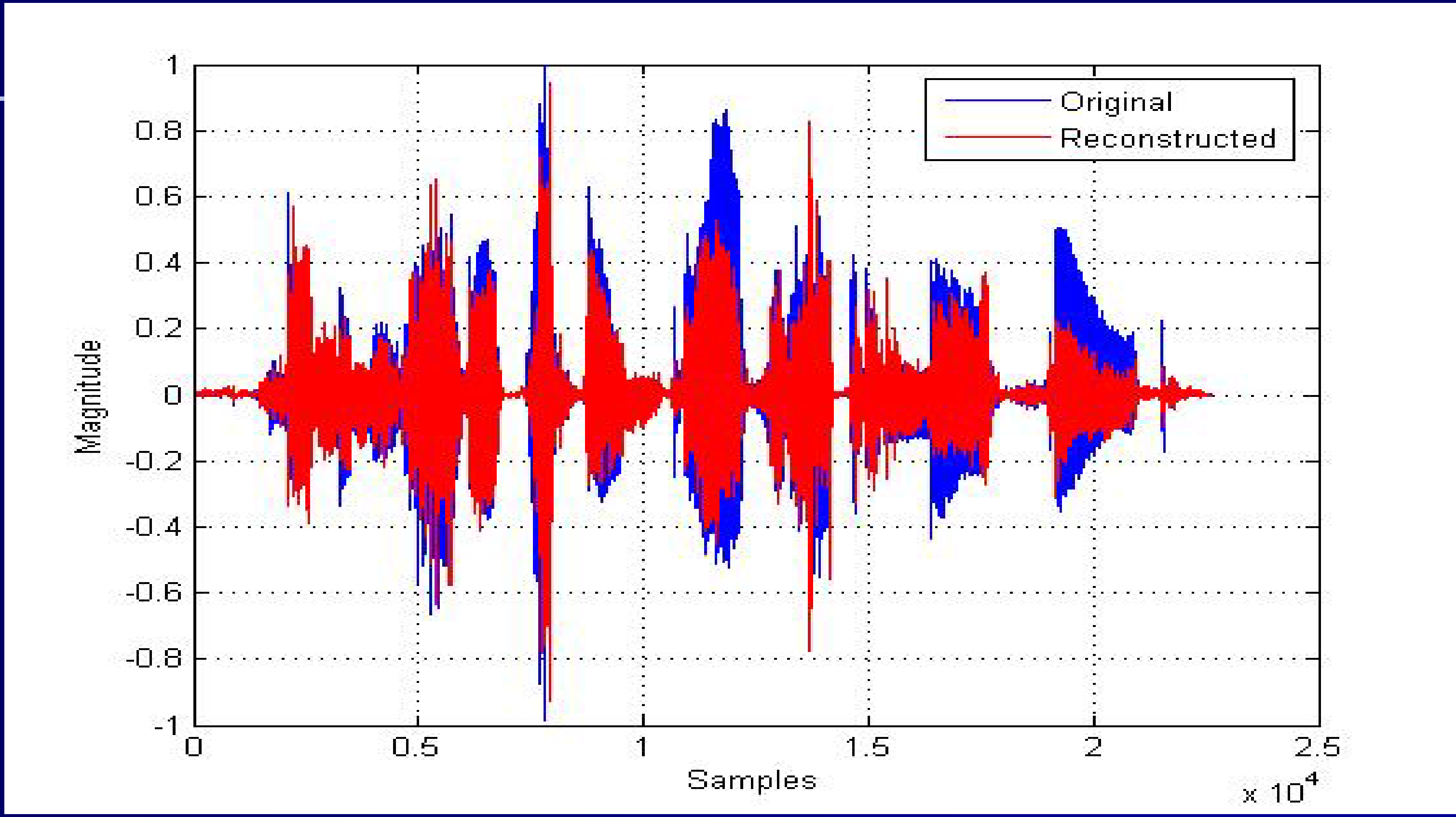
Performance of CELP (Unquantized) mse = 0.0041



Original

Unquantized

Performance of CELP (Quantized) mse = 0.0120



LP Coefficients: Unquantized
Other Parameters: Quantized

Quantization Methods Used

- Scalar Quantization
- DPCM
- Vector Quantization
- TSVQ

Scalar Quantization

- Quantize one sample at a time
- The simplest quantization scheme
- Design quantizers with sizes M = 2, 4 , 8, 16, 32, 64, 128, 256

Scalar Quantizer Design

- Lloyd algorithm

- Initial guess:
  a uniform codebook

Scalar Quantizer Design

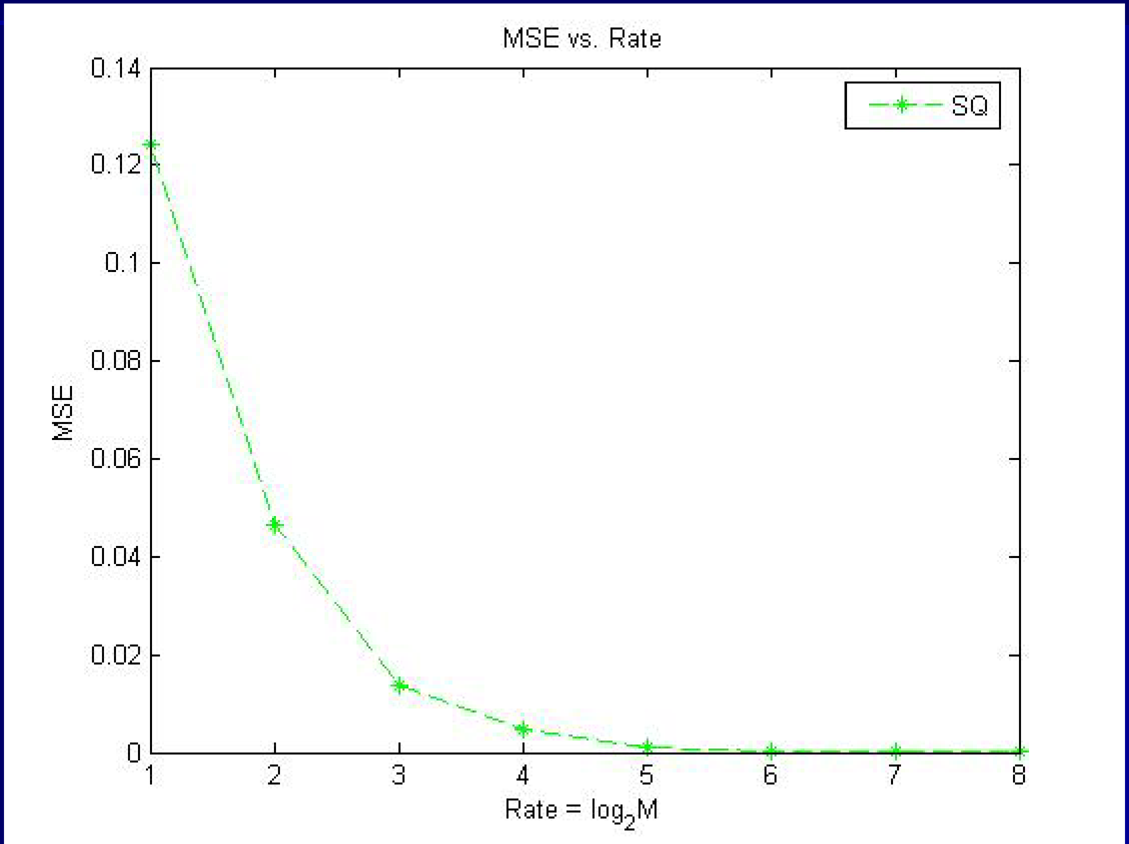
- Training data:

  15000 samples of LP coefficients generated from different speech sources

  15000/256 = 58 points/cell for M=256

  15000/2 = 7500 points/cell for M=2

Performance of the SQ



DPCM

- Quantizing the prediction error, once at a time
- Essentially a scalar quantizer
- Good for slowly varying sources
- Need a model for the source to design the linear predictor

DPCM Design – Predictor

- Assume a source model
- First-order AR, zero-mean Gaussian

DPCM Design – Predictor

- Gaussian?

  Many different kinds of speech, and LP coefficients

- Zero-mean?

  Empirical mean is near to zero

DPCM Design – Predictor

- First-order AR?

  Correlation analysis indicates a large first-order correlation coefficient, near 0.8, and small higher-order coefficients, smaller than 0.01

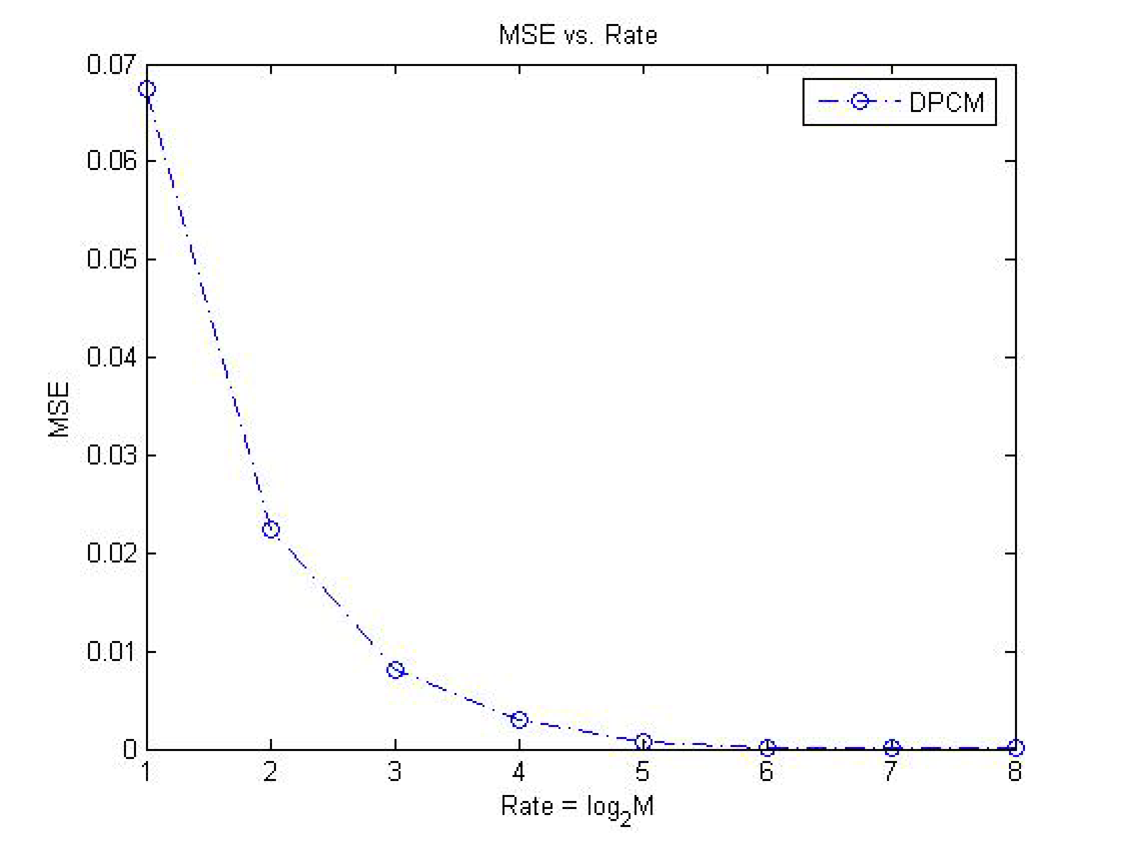
DPCM Design – Quantizer

- Designed to be optimal for the random variables

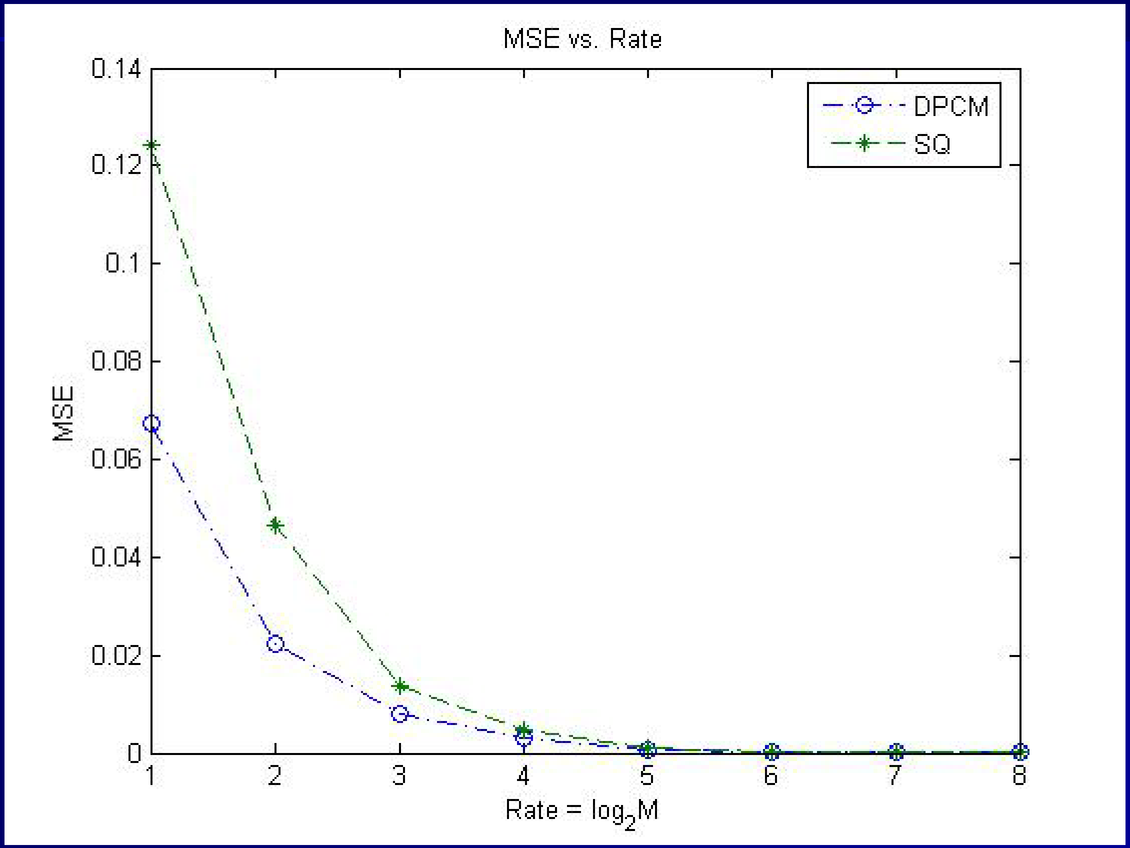
$$V_i = X_i - a_1 X_{i-1}$$

- Extract a_1 from correlation analysis, like solving the Yule-Walker equation
- Avoid calculating the limiting density of the prediction error

DPCM Performance

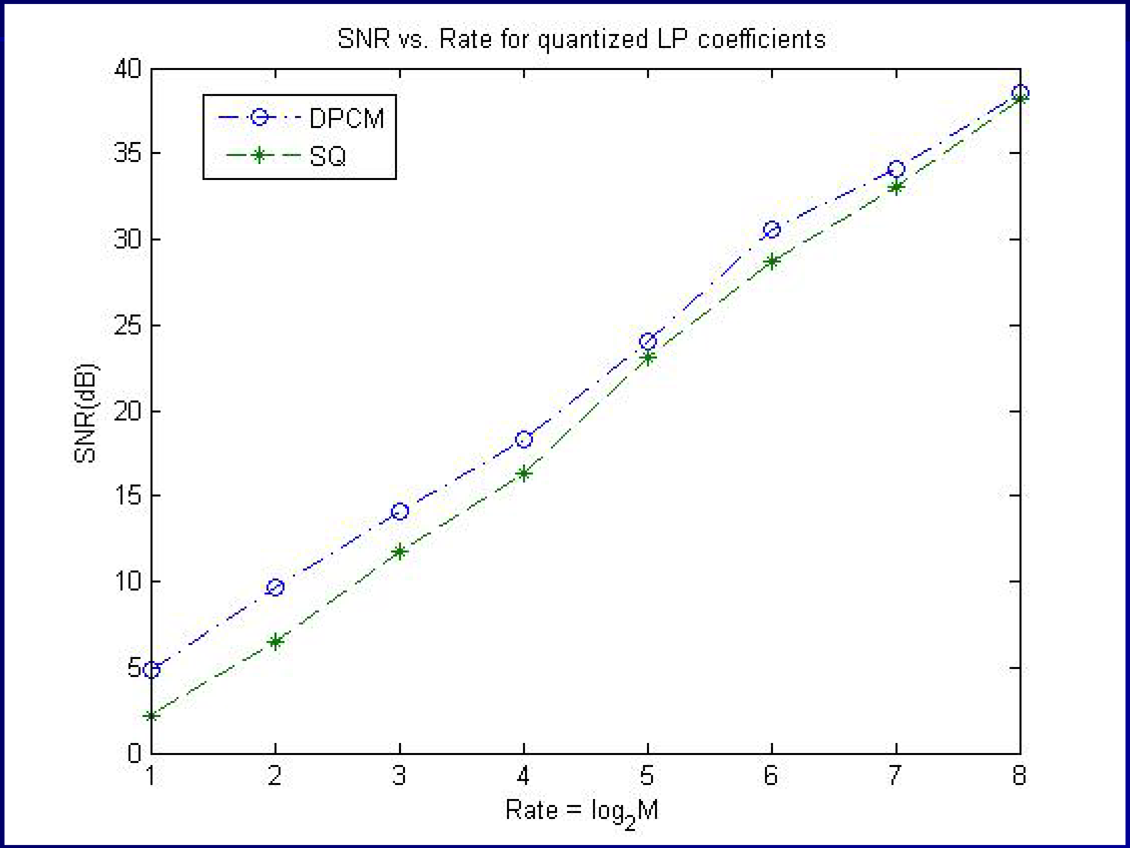


SQ vs. DPCM



MSE vs. Rate
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
MSE
1
2
3
4
5
6
7
8
Rate = log2M
DPCM
SQ

SQ vs. DPCM



SNR vs. Rate for quantized LP coefficients
DPCM
SQ
SNR(dB)
Rate = log2M
0
5
10
15
20
25
30
35
40
1
2
3
4
5
6
7
8

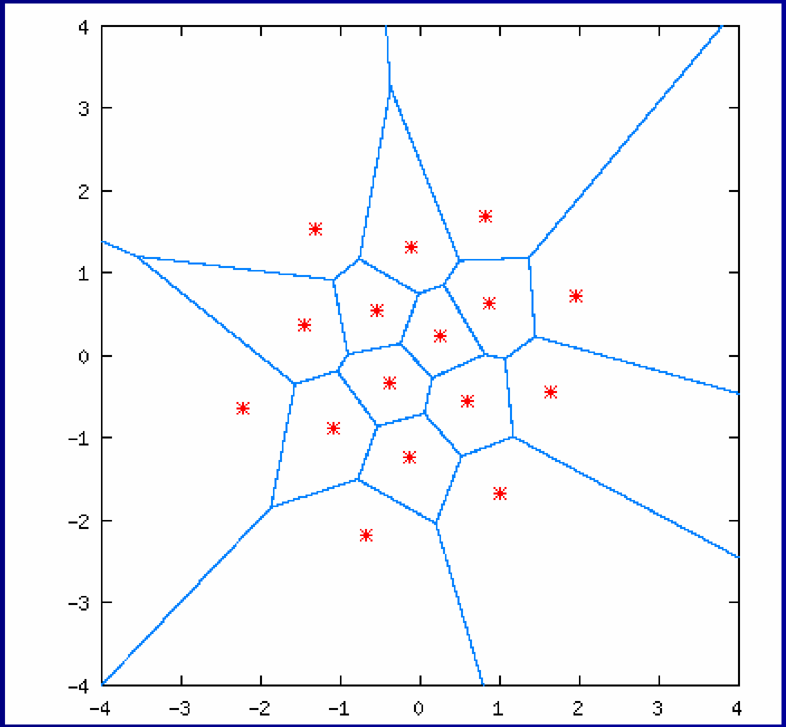
SQ vs. DPCM

For DPCM:

- Significant improvement for lower rate than SQ
- The simple models for sources and quantizer input are effective

Vector Quantization

- Key challenge
  - Given a source distribution, how to select codebook (*) and partitions (---) to result in smallest average distortion



VQ Design

- LBG algorithm was designed and implemented in Matlab
- Computes a codebook of a desired size given a training sequence

Performance of the CELP coder

- MOS, Mean Opinion Score
  - A sample of 20 people
  - Listen to reconstructed speech sample and rate the intelligibility
    - Excellent – 5
    - Good – 4
    - Fair – 3
    - Poor – 2
    - Bad – 1

Performance of Coder with DPCM

M = 2	MOS = 1
M = 4	MOS =1
M = 8	MOS =1
M = 16	MOS =1
M = 32	MOS =2.3
M = 64	MOS =3.1
M = 128	MOS =3.9
M = 256	MOS =4.5

Original

Performance of Coder with SQ

M = 2	MOS =1
M = 4	MOS =1
M = 8	MOS =1
M = 16	MOS =1
M = 32	MOS =1.8
M = 64	MOS =2.9
M = 128	MOS =3.6
M = 256	MOS =4.1

Original

Performance of Coder with VQ

M = 2 MOS =1.7

M = 4 MOS = 1.9

M = 8 MOS = 2.5

M = 16 MOS =2.9

M = 32 MOS =3.1

M = 64 MOS =3.1

M = 128 MOS =2.9

M = 256 MOS =3.0

Original

Conclusions

- Improvement in the quantization of LP coefficients improves the performance of the coder
- For a given codebook size, VQ performed better in terms of MSE
- DPCM performed better in terms of perceptual MSE

Questions

??????????

THANK YOU