ELEN 027 Lecture 5

- Review of last class
- Audio Compression basics

MPEG

- Successive frames may have significantly same data
- Can use delta encoding
  - Picture differencing
- Delta encoding leads to better compression
- MPEG exploits motion prediction
  - Image content changes due to motion
  - Camera movement, Object movement, Camera panning, etc
Basic MPEG encoding

- Preprocess data
- Use motion prediction
- Find errors of receiver's prediction and actual values
- Encode the differences using DCT
- Quantization
- Apply variable-length entropy coding

IPB frames

- B frames require future frames for decoding
- Decoding of B frames requires future P frames
- Displaying and transmitting order may change
- MPEG decoders may have to buffer extra frames of data besides the frame being displayed
Speech Coding

• Waveform Coders
  – Try to reproduce the original waveform
  – Mostly work in time domain

• Vocoders
  – Don’t try to reproduce the original waveform
  – Mostly work in frequency domain

Vocoders

• Example
  – Break up the signal energy into different frequency domains
  – Send the energy as coded signal
  – Excite a proper filter at the receiver to generate the original speech

• Found to lack warmth/naturalness of original speech

• Synthesized speech couldn’t identify the speaker
**Linear Predictive Coding**

- Predict the next sample as a linear combination of previous samples and a gain term
- The gain term depends on Voiced/Unvoiced signal
- Can be seen as waveform encoding where prediction error is not transmitted
- Found to be difficult to classify the signal as voiced/unvoiced
- Found to sound synthetic

**Waveform coders**

- Delta Modulation -DM
  - Predict the current sample to be same as previous sample
  - Send the difference as encoded signal
- One form of delta modulation uses only 2 quantization steps
  - Up or down
- Quantization steps can be chosen adaptively
Waveform Coders

- Differential Pulse Code Modulation - DPCM
  - More general prediction than DM
  - Quantize the prediction error and send it as signal

- Waveform coders = Predictive coders

Embedded DPCM

- Like to drop bits at the time of congestion
- Speech coding is not very forgiving of lost bits
  - Quality shown to decrease rapidly
  - Quality can be lower than lower-bitrate signal

- Embedded DPCM uses a lower resolution predictor model for receiver
  - Allows more flexibility in handling network congestion
  - Dropping bits results in more graceful loss of quality
Analysis-by-synthesis coders

- Synthesize all possibilities of speech for a given coder
- Use the best model that matches the perceptual qualities
- Send this model as coded data
- Allow perceptual distortion measures to be used in encoding

CELP Coding

- Code Excited Linear Predictive Coding
- Analysis-by-synthesis coders
- From a speech sample, remove the long-term and short-term redundancy
  - What is left is called an excitation sequence
- Use a finite number of excitation sequences
- Number these excitation sequences - Codebook
- Find the excitation sequence that best matches the current sample
- Transmit the codeword for this sequence from the codebook
- Can achieve substantial compression in bitrate
**CELP example**

- 2 kbps for excitation
- Original signal 8000 samples/second
- Consider a 5ms block or frame
  - 40 samples to be coded in 10 bits
  - Resulting in 0.25 bits/sample
- With a 10-bit excitation codeword
  - Can have 1024 excitation sequences
  - Choose a codebook offline
- Apply these 1024 excitation sequences to predictors
  - Find the sequence that minimizes the perceptual error with the sample
  - Transmit that codeword

**Complexity of CELP**

- Searching for the best matching excitation sequence
- Can this be done in reasonable time?
- Lot of research going on
- Found its way into standards
- Earlier example: 40 samples
  - 1 bit per sample, block = $2^{40}$ possible sequences
  - Can code with 1024 sequences!!
**CELP Quality**

- Found to be pretty good at such low bit rates
- Quality highly dependent perceptual weighting
- MSE measures don’t work very well
- Noise spectral shaping used to improve quality
- Noise spectral shaping
  - Tries to reshape noise energy
  - Make signal energy uniformly higher than noise energy in frequency band of interest