Multimedia Systems

Speech II

Mahdi Amiri

February 2014
Sharif University of Technology
Speech Compression

Road Map

Based on Time Domain analysis
- Differential Pulse-Code Modulation (DPCM)
- Adaptive DPCM (ADPCM)

Based on Frequency Domain analysis
- Linear Predictive Coding (LPC)
- Code Excited Linear Prediction (CELP)
Differential PCM (DPCM)

Idea

Take advantage of data redundancy

Or histogram of PCM samples in a chunk of digitized audio.
Typ. Chunk length: 50 ms
Differential PCM (DPCM)

Prediction

- Simplest prediction: The difference between the current sample value and previous sample is quantized. It means we predicted current sample by assuming it to be equal to the previous sample.

- Another better prediction: The difference between the current sample value and the average of e.g. 2 previous samples.

- Much better prediction: The difference between the current sample value and the weighted average of e.g. 10 previous samples.
Differential PCM (DPCM)

Basic Scheme

General Predictive Coding

Delta Modulation (DM): $\sum a_i x_{n-i} \Rightarrow z^{-1}$

Problem?
Differential PCM (DPCM)

### Error Propagation

The output of the dequantizer in the decoder is not equal with the input of the quantizer in the encoder.

- The input of the predictor in the decoder is not the same as the input values of the predictor in the encoder.
- This is the source of error propagation.

### General Predictive Coding

<table>
<thead>
<tr>
<th>$x_n$</th>
<th>0</th>
<th>0.6</th>
<th>1.2</th>
<th>1.8</th>
<th>2.4</th>
<th>3.0</th>
<th>3.6</th>
<th>4.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_n$</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>1.2</td>
<td>1.8</td>
<td>2.4</td>
<td>3.0</td>
<td>3.6</td>
</tr>
<tr>
<td>$d_n$</td>
<td>0</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>$\hat{d}_n$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\hat{p}_n$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\sum a_i x_{n-i}$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

$\text{Err}$: 0 0.4 0.8 1.2 1.6 2.0 2.4 2.8

Error propagation example in above structure when input signal increases constantly by +0.6.
Quantization step size is 1; Therefore, e.g. -0.5 thr. 0.5 quantized as 0 and 0.5 thr. 1.5 quantized as 1.
Differential PCM (DPCM)
Adaptive DPCM (ADPCM)

Idea

Problem?
Adaptive DPCM (ADPCM)

Size of Quantization Step

Delta Modulation (DM)

1 bit quantizer: 0 means $+\Delta$ and 1 means $-\Delta$

Adaptive Delta Modulation (ADM)

ADM: $\Delta[n] = M \Delta[n-1]$

$M = P > 1$ if $c[n] = c[n-1]$

$M = Q < 1$ if $c[n] \neq c[n-1]$

$P = 2, \quad Q = \frac{1}{2}$
Speech Compression Concepts

FFT, No Time Localization

Speech Signal

Joseph Fourier, 1768-1830

FFT (is only localized in frequency)
Speech Compression Concepts

FFT, No Time Localization, Ex. 1

See Power Spectral Density (PSD) examples in MATLAB
Speech Compression Concepts

FFT, No Time Localization, Ex. 2

See Power Spectral Density (PSD) examples in MATLAB

Multimedia Systems, Speech II
Speech Compression Concepts

STFT (fixed time and frequency localization)

Dennis Gabor, 1900-1979

Speech Signal

Multimedia Systems, Speech II
Speech Compression Concepts

3D surface spectrogram of a part from a music piece.
Speech Compression Concepts

Spectrogram of a male voice saying ‘nineteenth century’.
Speech Compression Concepts

Spectrogram Display in AudaCity

Waveform

Spectrogram
Speech Compression Concepts

Spectrogram Display in AudaCity

FFT Window size: 128

FFT Window size: 1024

AudaCity | Edit | Preferences | Spectrograms | FFT Window | Window size
Speech Compression Concepts

Spectrogram, Demonstration

Bat Echolocation Call

Flute by Jean Pierre Rampal

Singing Voice

Face!
The time and frequency domain presentation of vowels /a/, /i/, and /u/.
Speech Compression Concepts

Sample Application

A computing system to answer questions posed in natural language.

Design a reliable speech recognition unit for IBM Watson project.

Jeopardy! champions Ken Jennings (left) and Brad Rutter (right) versus the IBM computer Watson

www-943.ibm.com/innovation/us/watson/

Dr. David Ferrucci, Watson Principal Investigator
Linear Predictive Coding (LPC) Modeling

Simplified
Linear Predictive Coding (LPC) Modeling (Hiss or Buzz)

Buzzer $\rightarrow$ Filter

Chunks: 30 thr. 50 frames/sec.

Speech = Formants + Residue

Predictor for each frame:

$$\tilde{x}[n] = \sum_{i=1}^{P} a_i x[n - i]$$
Linear Predictive Coding (LPC)

Modeling (Hiss or Buzz)

The human vocal tract as an infinite impulse response (IIR) system

Vowel /a/

Vocal Tract $\Leftrightarrow H(z)$ (LPC Filter)
Air $\Leftrightarrow u(n)$ (Innovations)
Vocal Cord Vibration $\Leftrightarrow V$ (voiced)
Vocal Cord Vibration Period $\Leftrightarrow T$ (pitch period)
Fricatives and Plosives $\Leftrightarrow UV$ (unvoiced)
Air Volume $\Leftrightarrow G$ (gain)

LPC Block Diagram

Multimedia Systems, Speech II
Comparison of wide-band sound spectrograms for synthetic and original speech signal for the utterance "It's time we rounded up that herd of Asian cattle," spoken by a male speaker.
Linear Predictive Coding (LPC)

Voiced Frame Example

Original

Synthetic

Time Domain

Frequency Domain

180 samples, Pitch period: 75
Linear Predictive Coding (LPC)

Unvoiced Frame Example

Original

Time Domain

180 samples

Synthetic: White noise with uniform distribution

Frequency Domain

Multimedia Systems, Speech II
Code Excited Linear Prediction

- **Problem of LPC**
  - Where there is both Hiss and Buzz

- **Solution**
  - Encode residue

- **Method**
  - Vector Quantization (Codebook)
Vector Quantization

Block Diagram
Sample scalar quantizer
We have 3 possible colors for each square; so we can quantize each square with 2 bits $\rightarrow (28 \times 2 = 56 \text{ bits})$ for all 28 (7*4) squares.

Sample vector quantizer
We have 8 forms in the codebook; so we can quantize each form with 3 bits $\rightarrow (7 \times 3 = 21 \text{ bits})$ for all 28 (7*4) squares.
Vector Quantization

Codebook Design

[Diagram showing vector quantization and codebook design for Speaker 1 and Speaker 2 with centroids and samples.]
Comparison of Speech Coders

Sample Speech

A lathe is a big tool. Grab every dish of sugar.

Multimedia Systems, Speech II
Comparison of Speech Coders

Demonstration

Original

ADPCM

LPC

CELP
Speech Coding

ITU-T Standards

- **G.711**
  - PCM
  - u-law, a-law
  - 64, 80 and 96 kbps

- **G.722**
  - ADPCM
  - 48, 56 and 64 kbps

- **G.728**
  - A form of CELP
  - 16 kbps

Check out a complete list at http://en.wikipedia.org/wiki/List_of_codecs#Audio_codecs


Vocoders
**Speech Coding**

Free and Open Source Code

**HawkVoice**

http://hawksoft.com/hawkvoice/

<table>
<thead>
<tr>
<th>Kbps</th>
<th>codec name</th>
<th>license</th>
<th>code type</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>G.711 u-law</td>
<td>LGPL</td>
<td>fixed point</td>
</tr>
<tr>
<td>32</td>
<td>Intel/DVI ADPCM</td>
<td>Free *</td>
<td>fixed point</td>
</tr>
<tr>
<td>13.2</td>
<td>GSM</td>
<td>LGPL</td>
<td>fixed point</td>
</tr>
<tr>
<td>4.8</td>
<td>LPC</td>
<td>LGPL</td>
<td>floating point</td>
</tr>
<tr>
<td>4.5-2.3</td>
<td>CELP</td>
<td>LGPL</td>
<td>floating point</td>
</tr>
<tr>
<td>2.4</td>
<td>LPC10</td>
<td>LGPL</td>
<td>floating point</td>
</tr>
<tr>
<td>2.4</td>
<td>LPC10</td>
<td>$</td>
<td>fixed point</td>
</tr>
<tr>
<td>Up to 2.4</td>
<td>VBR-LPC10</td>
<td>LGPL</td>
<td>floating point</td>
</tr>
<tr>
<td>Up to 2.4</td>
<td>VBR-LPC10</td>
<td>$</td>
<td>fixed point</td>
</tr>
<tr>
<td>1.8-1.4</td>
<td>OpenLPC</td>
<td>Free *</td>
<td>floating point</td>
</tr>
<tr>
<td>1.8-1.4</td>
<td>OpenLPC</td>
<td>LGPL, $</td>
<td>fixed point</td>
</tr>
</tbody>
</table>

Check out voice samples of HawkVoice™ codecs at http://hawksoft.com/hawkvoice/codecs.shtml
Thank You

Next Session: Entropy Coding

FIND OUT MORE AT...

1. http://ce.sharif.edu/~m_amiri/