Perceptual Weighting in LSP-Based Multi-Description Coding for Real-time Low-Bit-Rate VoIP

Dong Lin, Benjamin W. Wah, and Hang Yu
Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
http://manip.crhc.uiuc.edu

2005 IEEE Workshop on Multimedia Signal Processing

Outline

• Introductions
  – Goal
  – Quality metrics
  – LSP-based MDC
  – Problem statement

• Proposed Approach
  – Identifying the cause of degradations
  – PWF tuning

• Experimental Results
Goal

- Design good-quality codec for VoIP applications under
  - Limited bit rate
    - FS-1016 CELP (4.8kbps)
    - G723.1 ACELP (5.3kbps)
    - G723.1 MP-MLQ (6.3kbps)
    - G.729 (8kbps)
  - Non-stationary packet loss rate: low to high

Existing Techniques

- Accurate channel model is difficult to obtain for IP networks
  - Forward Error Correcting (FEC) code
  - Joint Source Channel Coding (JSCC)

- Multiple Description Coding (MDC)
  - Information is interleaved into multiple descriptions at the source
  - Receiver can recover from any description received
  - Better quality with more descriptions
IP Packet Losses Concealed by MDC

- Target concealed loss rate: 5% or less
- Maximum number of descriptions required: 4

![Graph showing P(fail|i) for different times of day and hour for UIUC-Western China and UIUC-Slovakia](image)

LPC Speech Coding

- LPC coding $S(z) = A(z)E(z)$
  - Decompose frame into LP coefficients $a(n)$ and excitations $e(n)$
  - LP coefficients $a_i$: $H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + \cdots + a_{10} z^{-10}}$
  - Line spectrum pairs (LSP) $x_k$:
    \[ P(z) = A(z) + z^{-11}A(z^{-1}); \quad Q(z) = A(z) - z^{-11}A(z^{-1}) \]
    Stable, less sensitive to quantization errors, and contain redundancy
  - Excitations $E(z)$: random, not much redundancy
Quality Metrics

- Likelihood Ratio \( LR = \frac{a_r R_o a_r^T}{a_o R_o a_o^T} \)
  \( \vec{a}_o \): vector of linear prediction coefficients of original speech
  \( \vec{a}_r \): vector of linear prediction coefficients of reconstructed speech
  \( R_o \): correlation matrix derived from original speech

- Cepstral Distance \( CD = 4.34[(c_0 - c'_0)^2 + 2\sum_{i=1}^{\infty}(c_i - c'_i)^2]^{\frac{1}{2}} \) [dB]
  \( c_i \): cepstra of original samples
  \( c'_i \): cepstra of the reconstructed samples

- Perceptual Evaluated Speech Quality (PESQ: ITU P.862)
  – Close correlation to Mean Opinion Score (MOS)

Dong Lin, Benjamin W. Wah, and Hang Yu
• Introductions
  – Goal
  – Quality metrics
  – LSP-based MDC
  – Problem statement

• Proposed Approach
  – Identifying the cause of degradations
  – PWF tuning

• Experimental Results

Uneven Coding Noise in MDC across Frequencies

• Formants (spectral peaks) have greater perceptual importance than valleys
• Noise energies of MDC in formant regions are excessive
Quantifying the Causes of Degradations

- **Notations**
  - \( f \): normalized frequency, \([0,1]\)
  - \( v \): audio file tested
  - \( \ell \): loss scenario
  - \( \gamma \): coder-dependent PWF parameter (explained later)

- **Two frequency-domain measures**
  - Relative coding noise of MDC wrt SDC at \( f \):
    \[
    R_1^2(\ell_0, v, \gamma) = \int_0^1 \log_e r_1(f, \ell_0, v, \gamma) df
    \]
    (over the entire spectrum)
  - Relative energy of MDC wrt SDC at \( f \):
    \[
    R_2^2(\ell_0, v, \gamma) = \int_0^1 \log_e r_2(f, \ell_0, v, \gamma) df
    \]
    (over the entire spectrum)

Illustration of the Cause of Degradation

- a) original PWF with \( \gamma = 0.8 \)
  - MDC has much higher relative coding noise in formant regions

- c) modified PWF with \( \gamma = 0.6 \)
  - Using modified PWF reduces the relative coding noise of MDC in formant regions

- b) original PWF with \( \gamma = 0.8 \)

- d) modified PWF with \( \gamma = 0.94 \)
Noise Shaping using Perceptual Weighting Filter

- PWF is inversely related to the LP filter response and speech’s spectrum
  - FS-1616 CELP: \( W(z) = \frac{A(z)}{A(z/\gamma)} \), shape controlled by \( \gamma \)
  - G723.1: \( W(z) = \frac{A(z/\alpha)}{A(z/\beta)} \), shape controlled by \( \beta \)

![Magnitude vs Normalized Frequency Graph]

Dong Lin, Benjamin W. Wah, and Hang Yu
Generalization to Different Voice Files and Loss Scenarios

- The best PWF parameter ($\gamma$) is dependent on voice file and loss scenarios

- Generalization procedure

  - Select a common $\gamma$ to minimize the deviation from the optimal $R_1$ (or $R_2$) over all voice files and loss scenarios

<table>
<thead>
<tr>
<th>Coder</th>
<th>FS-1016 CELP</th>
<th>G723.1 ACELP</th>
<th>G723.1 MP-MLQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDC</td>
<td>2-way</td>
<td>2-way</td>
<td>2-way</td>
</tr>
<tr>
<td>Metric</td>
<td>$\gamma^a$</td>
<td>$\Delta R_{\gamma}^m$</td>
<td>$\gamma^a$</td>
</tr>
<tr>
<td></td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_1$</td>
</tr>
<tr>
<td></td>
<td>0.62</td>
<td>0.94</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Dong Lin, Benjamin W. Wah, and Hang Yu
Experimental Results: LR

- Trace-driven simulations with periodic 1-bit feedback to switch between 2-way and 4-way MDC

![Graph showing LR for different hours of the day in UIUC-China (20% loss rate) and UIUC-Slovakia (45% loss rate).]

Experimental Results: CD

![Graph showing CD for different hours of the day in UIUC-China (20% loss rate) and UIUC-Slovakia (45% loss rate).]
Experimental Results: PESQ

- Tuning PWF can reduce quality degradations caused by MDC and fixed bit rate

- Current work
  - Identification of specific voice patterns causing degradation (ICME’05)
  - Study of rate-distortion trade-offs to increase bit rate and eliminate degradations over SDC (MMSP’05b)