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

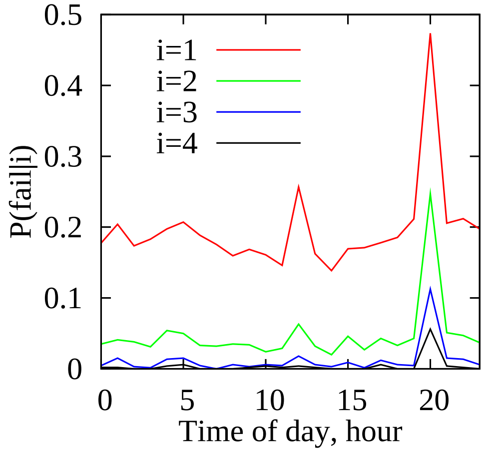
- Design good-quality codec for VoIP applications under
  - Limited bit rate
    - FS-1016 CELP (4.8kbps)
    - G723.1 ACELP (5.3kbps)
    - G723.1 MP-MLQ (6.3kbps)
    - G.729 (8kbps)
  - Non-stationary packet loss rate: low to high

Existing Techniques

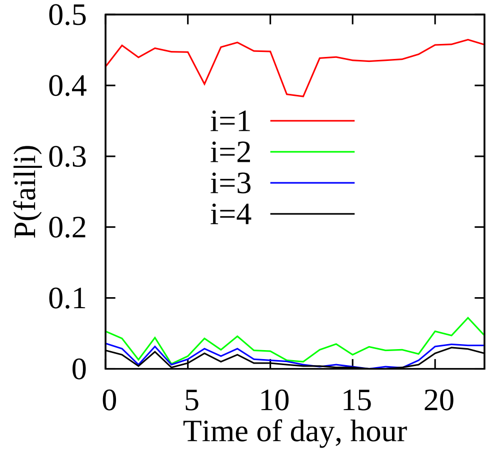
- Accurate channel model is difficult to obtain for IP networks
  - Forward Error Correcting (FEC) code
  - Joint Source Channel Coding (JSCC)
- Multiple Description Coding (MDC)
  - Information is interleaved into multiple descriptions at the source
  - Receiver can recover from any description received
  - Better quality with more descriptions

IP Packet Losses Concealed by MDC

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



UIUC-Western China



UIUC-Slovakia

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: $H(z) = \dfrac{1}{A(z)} = \dfrac{1}{1 + a_1 z^{-1} + \cdots + a_{10} z^{-10}}$
  - Line spectrum pairs (LSP) x_k:

    $$P(z) = A(z) + z^{-11}A(z^{-1}); \qquad Q(z) = A(z) - z^{-11}A(z^{-1})$$

    Stable, less sensitive to quantization errors, and contain redundancy
  - Excitations $E(z)$: random, not much redundancy

Quality Metrics

- Likelihood Ratio $LR = \dfrac{a_r R_o a_r^T}{a_o R_o a_o^T}$

  $\vec{a}_o$: 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]

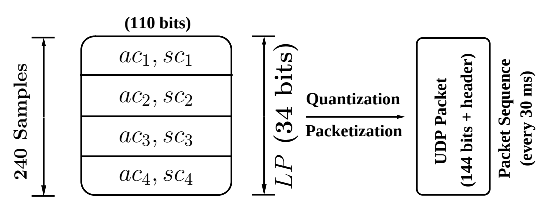
  c_i: cepstra of original samples

  c'_i: cepstra of the reconstructed samples

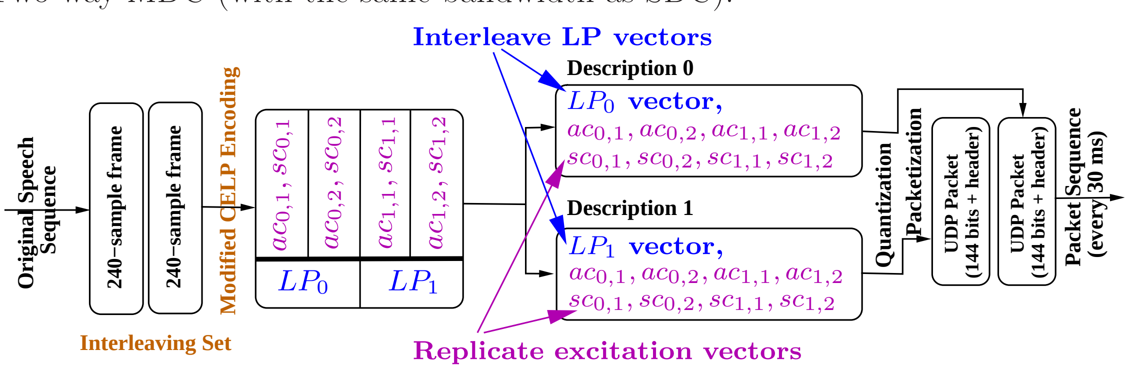
- Perceptual Evaluated Speech Quality (PESQ: ITU P.862)
  - Close correlation to Mean Opinion Score (MOS)

FS CELP SDC and LP-Based Two-Way MDC

- FS CELP SDC:



- Two-way MDC (with the same bandwidth as SDC):

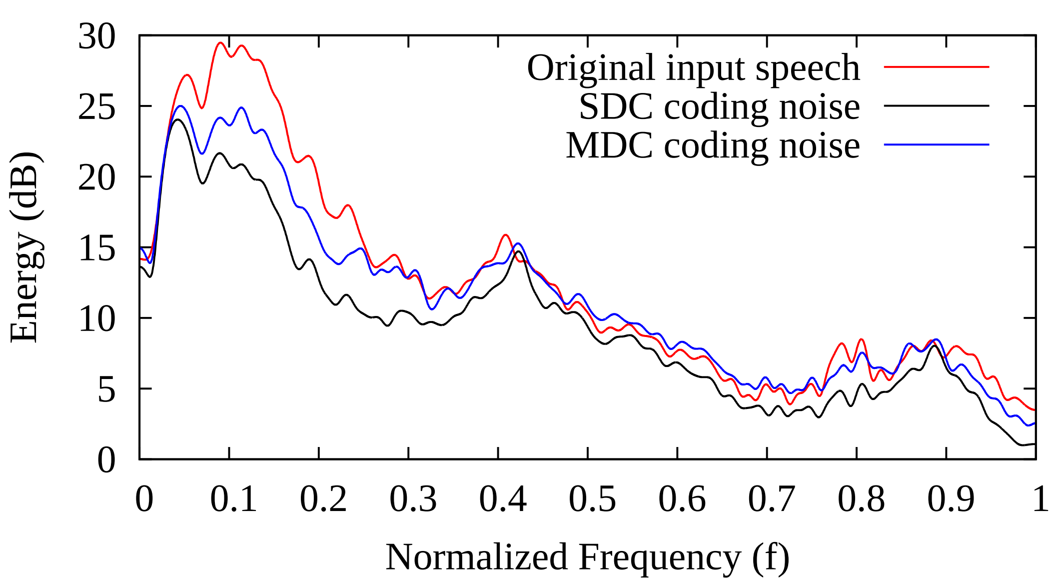


- Decoded speech at receivers has degraded quality due to longer subframes

Outline

- 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

  ℓ: loss scenario

  γ: coder-dependent PWF parameter (explained later)

- Two frequency-domain measures

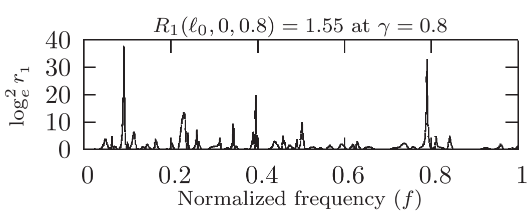
  – Relative coding noise of MDC wrt SDC at f: $r_1(f, \ell_0, v, \gamma) = \frac{|D_{M(\ell_0,v,\gamma)}(j2\pi f)|^2}{|D_{S(\ell_0,v,\gamma)}(j2\pi f)|^2}$

  $R_1{}^2(\ell_0, v, \gamma) = \int_0^1 \log_e^2 r_1(f, \ell_0, v, \gamma) df$ (over the entire spectrum)

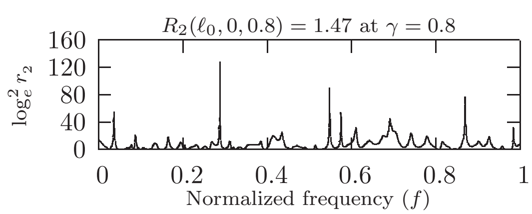
  – Relative energy of MDC wrt SDC at f: $r_2(f, \ell_0, v, \gamma) = \frac{|\hat{S}_{M(\ell_0,v,\gamma)}(j2\pi f)|^2}{|S(j2\pi f)|^2}$

  $R_2{}^2(\ell_0, v, \gamma) = \int_0^1 \log_e^2 r_2(f, \ell_0, v, \gamma) df$ (over the entire spectrum)

Illustration of the Cause of Degradation

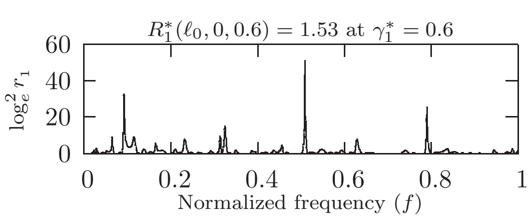


a) original PWF with $\gamma = 0.8$

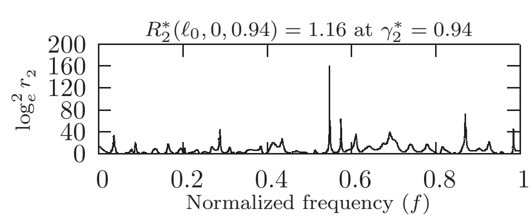


b) original PWF with $\gamma = 0.8$

MDC has much higher relative coding noise in formant regions



c) modified PWF with $\gamma = 0.6$

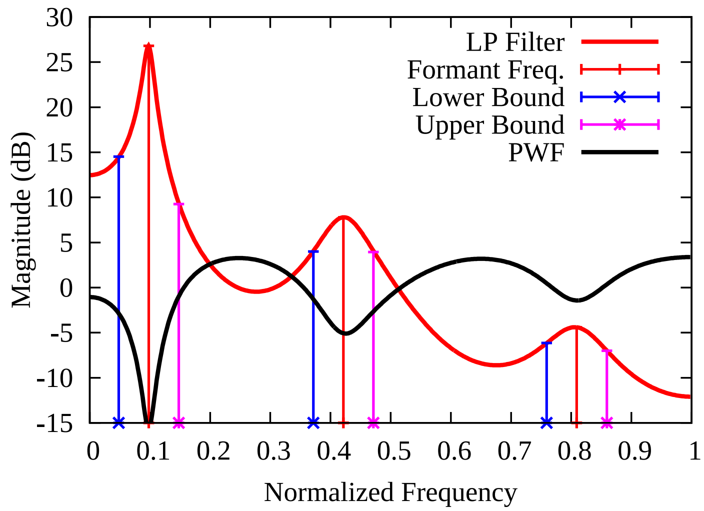


d) modified PWF with $\gamma = 0.94$

Using modified PWF reduces the relative coding noise of MDC in formant regions

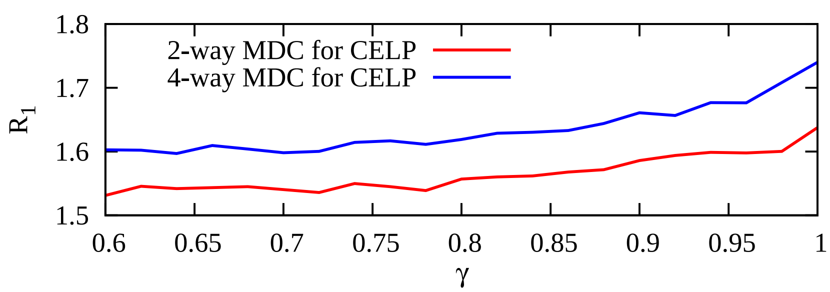
Noise Shaping using Perceptual Weighting Filter

- 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 γ
  - G723.1: $W(z) = \frac{A(z/\alpha)}{A(z/\beta)}$, shape controlled by β

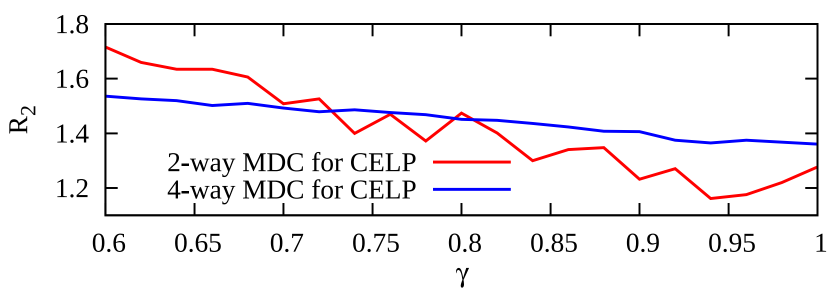


Perceptual Weighting in LSP-Based Multi-Description Coding for Real-time Low-Bit-Rate VoIP Proposed Approach

Effect of PWF on Relative Coding Noise and Energy (FS CELP)



$R_1(\ell_0, 0, \gamma)$ for CELP (default $\gamma = 0.8$)



$R_2(\ell_0, 0, \gamma)$ for CELP (default $\gamma = 0.8$)

Generalization to Different Voice Files and Loss Scenarios

- The best PWF parameter (γ) is dependent on voice file and loss scenarios
- Generalization procedure
  - Select a common γ to minimize the deviation from the optimal R_1 (or R_2) over all voice files and loss scenarios

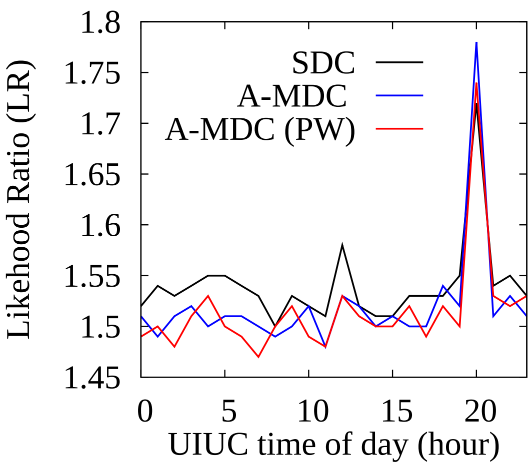
Coder	FS-1016 CELP				G723.1 ACELP				G723.1 MP-MLQ			
MDC	2-way		4-way		2-way		4-way		2-way		4-way	
Metric	R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2
γ_j^a	0.62	0.94	0.60	0.98	0.85	0.85	0.45	0.85	0.30	0.90	0.20	0.90
$\Delta R_j^{\min}$	0.02	0.04	0.03	0.03	0.06	0.07	0.04	0.04	0.07	0.05	0.02	0.04

Outline

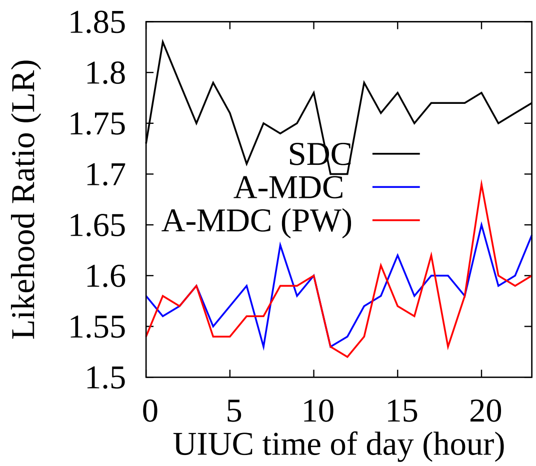
- Introductions
  - Goal
  - Quality metrics
  - LSP-based MDC
  - Problem statement
- Proposed Approach
  - Identifying the cause of degradations
  - PWF tuning
- Experimental Results

Experimental Results: LR

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

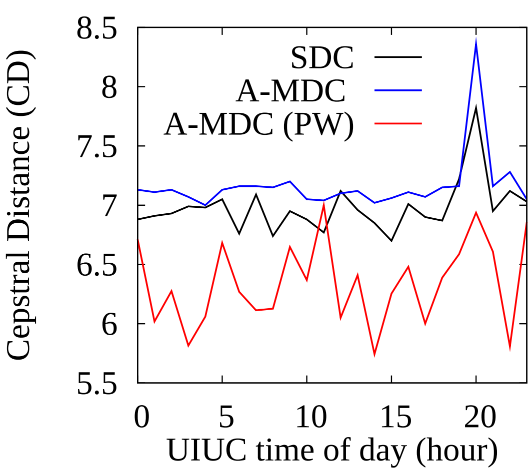


UIUC-China (20% loss rate)

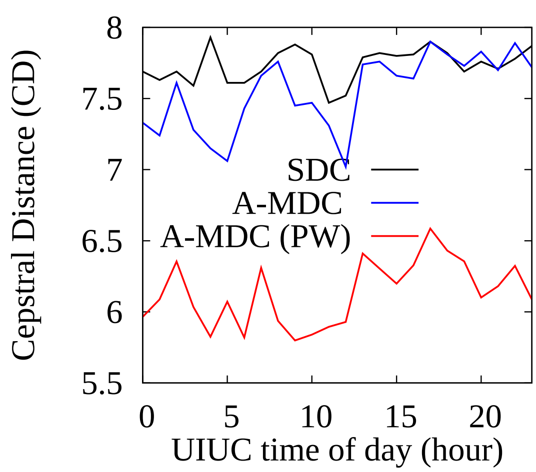


UIUC-Slovakia (45% loss rate)

Experimental Results: CD

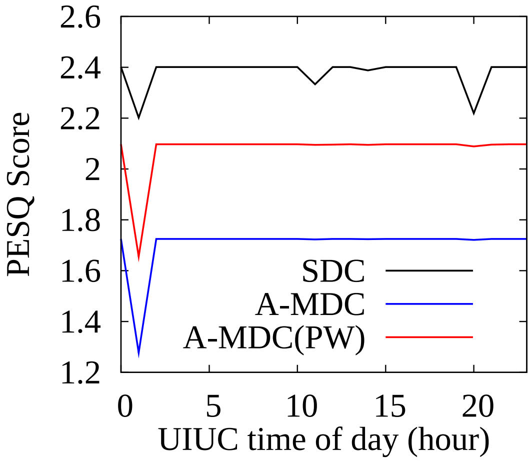


UIUC-China (20% loss rate)

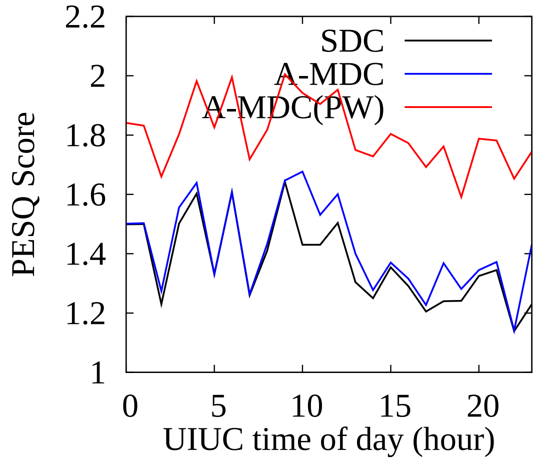


UIUC-Slovakia (45% loss rate)

Experimental Results: PESQ



UIUC-China (20% loss rate)



UIUC-Slovakia (45% loss rate)

Conclusions

- 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)