LSP-BASED MULTIPLE-DESCRIPTION CODING FOR REAL-TIME LOW BIT-RATE VOICE TRANSMISSIONS

Dong Lin and Benjamin W. Wah

Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

August 29, 2002

Outline

- Introduction
- Previous work
- Loss characteristics of IP voice traffic
- Loss concealments for low bit-rate coded speech
  - Parameter-based MDC
- Summary and future work

Problem Scope

- Real-time:
  - Interactive communication
  - ITU G.114: end-to-end delay less than 400 msec acceptable
- Loss:
  - Some degradations on voice samples tolerable
  - Long-burst or frequent short-burst intolerable
- Packet network:
  - Packet size: less than MTU to avoid fragmentation
  - Packet rate: 20 - 30 packets per second
- Low bit-rate coded speech:
  - Error propagation

Environments

- Network-layer protocol — Loss unavoidable
  - IPv4: best-effort, no real-time support
  - IPv6: best-effort, may support real-time traffic
  - Wireless: future IP-based
- Transport-layer protocol — Loss of real-time voice not handled
  - TCP: reliable but not suitable for real-time
  - UDP: unreliable
- Application-layer protocol — Loss of real-time voice not handled
  - RTP: no loss recovery scheme
  - H.323: umbrella standard for interoperability
- Packet losses in real-time voice communications left for end-point applications

Problem Addressed

Design, analyze and evaluate robust end-to-end loss-concealment schemes

- Allow reliable and real-time low bit-rate voice transmissions
- For unreliable IP networks, like the Internet and wireless wide area networks
- Without increasing bit rate in transmissions

Applications:

- Internet telephony
- Teleconferencing
- Wireless communications

Previous Work: Coder-Independent Schemes

- Schemes depending priority support from the network
  - Different priorities of different frames, e.g.: voiced, unvoiced [DaSilva 89]
  - Two pass coding: first pass coding original signals, second residue [Yong 92]
- Schemes adding explicit redundancy
  - Send extracted information of a packet in its following packet [Valenzuela 89]
  - Use forward error correction (FEC) [Shacham89]
- Schemes exploiting inherent redundancy in voice streams
  - Replay, pad by silence or white noise (receiver-only) [Tucker 85]
  - Waveform substitution (receiver-only) [Wasem 88]
  - Sample-based MDC (sender-receiver with no redundancy) [Jayant 81]

Previous Work: LP Coder-Dependent Schemes

- Schemes depending priority support from the network
  - assign different priorities of parameters [Yong 92]
- Schemes adding explicit redundancy
  - FEC for the most sensitive parameters [Atungsiri 93]
  - duplicate base information, e.g. LP [Anandakumar 00]
- Schemes exploiting inherent redundancy in voice streams
  - Single description
    - ∗ Parameter reconstruction (receiver-only)[Atungsiri 93]
    - ∗ Parameter re-initialization (sender-receiver) [Montminy 00]
  - No existing non-redundant MDC for LP coders

Outline

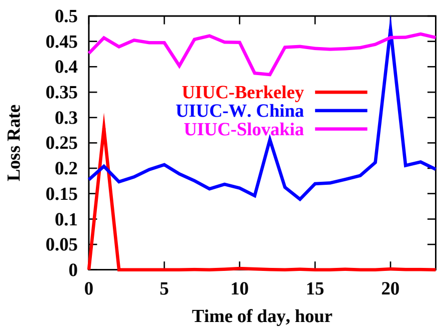
- Introduction
- Previous work
- Loss characteristics of IP voice traffic
- Loss concealments for low bit-rate coded speech
  - Parameter-based MDC
- Summary and future work

Loss Characteristics of IP Voice Traffic

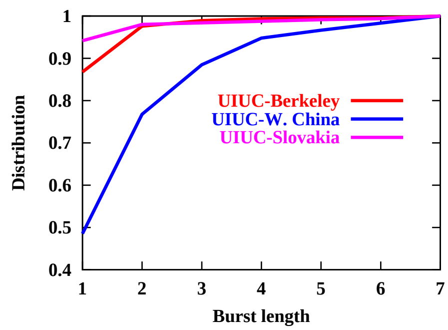
- Example connections

Connection	Loss rate
UIUC-Berkeley	low-medium
UIUC-W. China	medium-high
UIUC-Slovakia	high

- Loss behavior



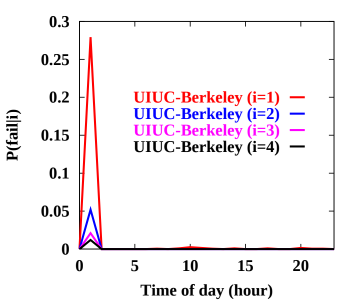
Loss rate can go up to 50%

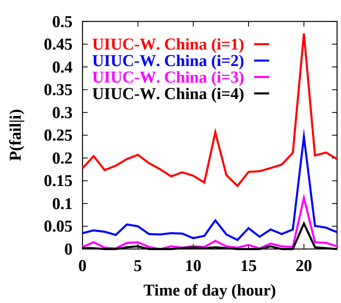


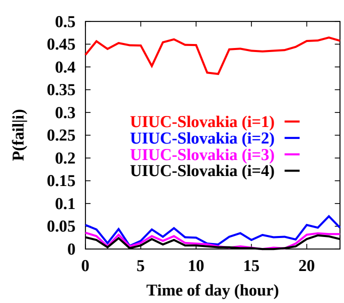
Most losses have short burst lengths

Reducing Unrecoverable Loss by Interleaving

- Bursty losses are difficult to handle
- Interleaving: disperse burst losses to isolated losses
- *P(fail|i)*: prob. of losses that cannot be recovered under interleaving factor i







- Small interleaving factor 2 – 4 is enough
- Multiple-description coding is promising

Outline

- Introduction
- Previous work
- Loss characteristics of IP voice traffic
- Loss concealments for low bit-rate coded speech
  - Parameter-based MDC
- Summary and future work

Testing Coders and Streams

- Coders

	Bit rate (bps)	Quantization of LSP	Excitation
FS CELP	4800	scalar	stochastic code/adaptive code
ITU G.723.1 (I)	5300	predictive-split VQ	algebraic code/adaptive code
ITU G.723.1 (II)	6300	predictive-split VQ	multi pulse/adaptive code
FS MELP	2400	multi-stage VQ	mixed pulse- and noise-like

- Streams

Index	Length (ms)	Speakers	Index	Length (ms)	Speakers
1	21432	2 male, 1 female	5	4160	1 male
2	22560	2 male, 1 female	6	4082	1 male
3	4424	1 female	7	4867	1 male, 1 female
4	5091	1 female	8	73615	1 male, 1 female

Objective Measures

- Itakura-Saito likelihood ratio

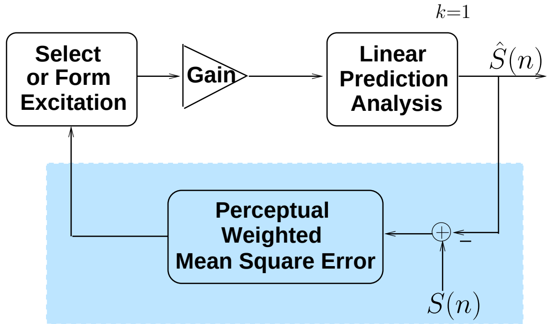
$$LR = \frac{a_r R_o a_r^T}{a_o R_o a_o^T}$$

- Cepstral distance:

$$CD = 4.34[(c_{o,\,0} - c_{r,\,0})^2 + 2\sum_{i=1}^{\infty}(c_{o,\,i} - c_{r,\,i})^2]^{\frac{1}{2}} \text{ [dB]}$$

Typical Linear Predictive Coder

$$H(w) = \frac{1}{A(w)} = \frac{1}{1 - \sum_{k=1}^{10} a_k e^{jwk}}$$

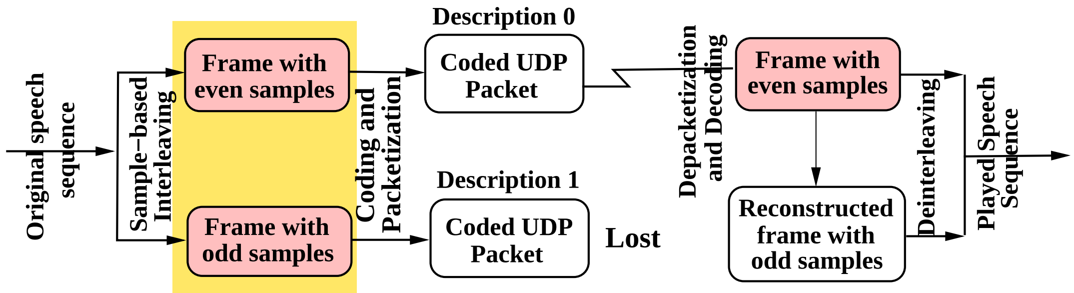


Major techniques:

- Frame-oriented
- Linear prediction analysis, coefficients generally represented by LSP
- Excitations: pitch information and random noise
- Can be open-loop or closed-loop

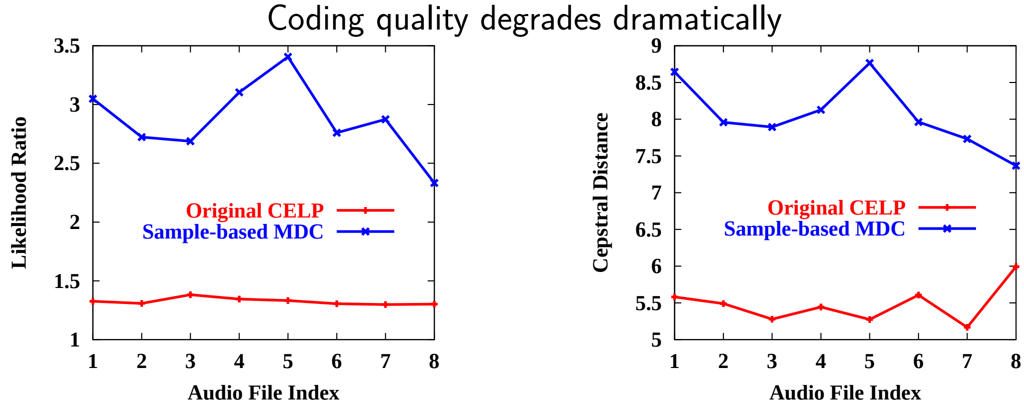
- FS CELP⋆, ITU G.723.1 ACELP, ITU G.723.1 MP-MLQ, and MELP

Coder-Independent Sample-Based MDC



Performance of Coder-Independent Sample-Based MDC

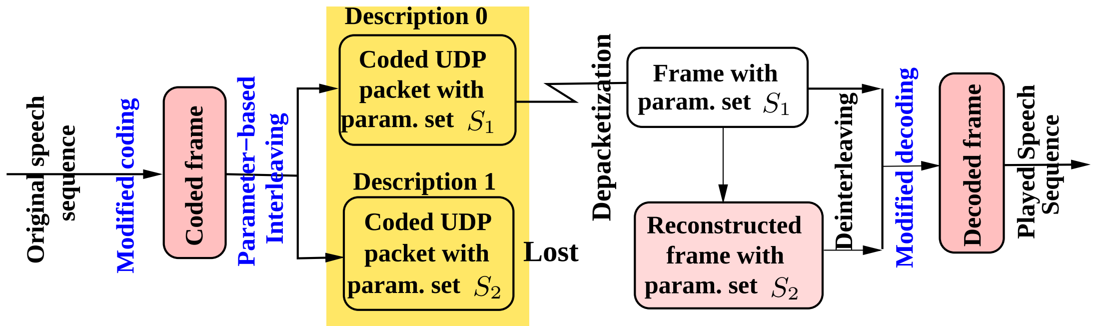
Coding quality degrades dramatically



Drawbacks:

- Aliasing: caused by down sampling
- Coding-frame time span lengthened

Coder-Dependent Parameter-Based MDC



- Parameters of linear predictive coders:
  - Linear predictor equivalent representations:
    - $*$ Reflection coefficient (RF), Log area ratio (LAR), LSP
  - Excitation
- MDC design by correlation analysis

Correlations of Linear Predictor Representations

- Correlations of LSP

Frame Distance	LSP x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	0.82	0.81	0.75	0.72	0.81	0.76	0.74	0.73	0.73	0.74
2	0.61	0.64	0.50	0.45	0.59	0.46	0.43	0.43	0.45	0.55
3	0.46	0.52	0.35	0.26	0.40	0.24	0.21	0.24	0.26	0.42

- Correlations of RF, LAR are similar
- Comparable to voice-sample correlations

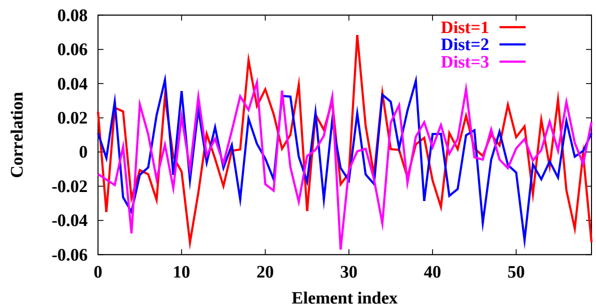
Sample Dist.	1	2	3
Correlation	0.83	0.60	0.35

Correlations of FS CELP Excitation Parameters

- Adaptive codewords: 2-element vector

Parameter Distance	*ac* delay Corr.	*ac* gain Corr.
1	0.57	0.004
2	0.22	0.007
3	0.21	0.006

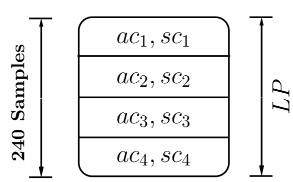
- Stochastic codewords: 60-element vector



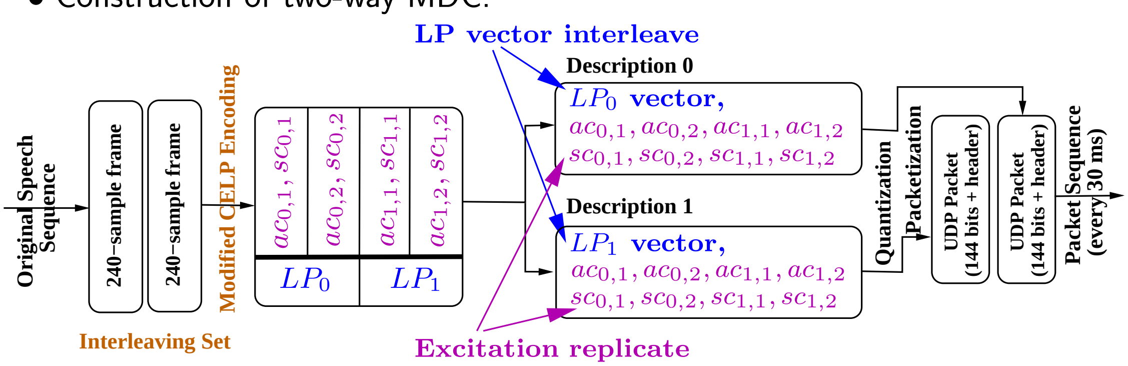
- Very low or no correlation for excitation parameters

FS CELP SDC and LP-Based Two-Way MDC

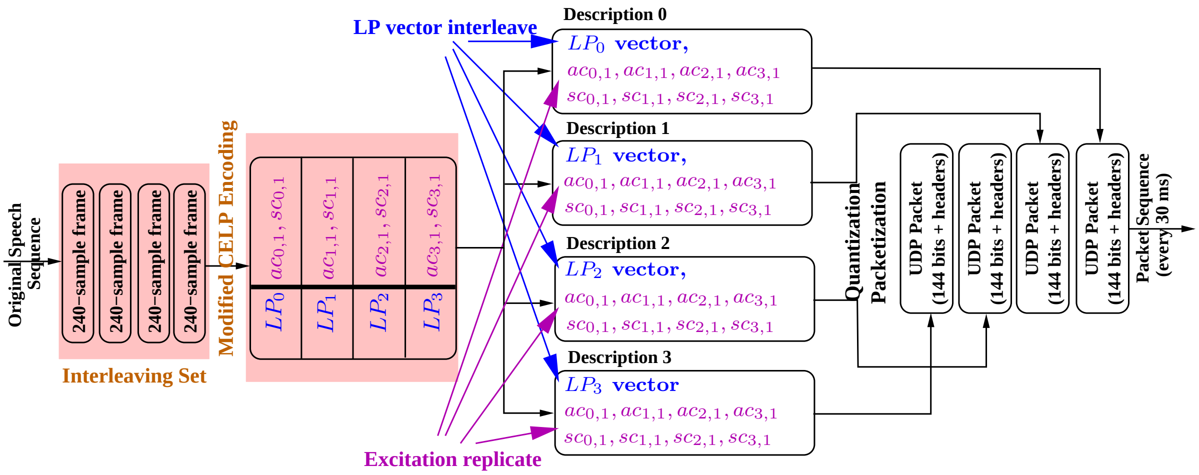
- FS CELP SDC:



- Construction of two-way MDC:



LP-Based Four-Way MDC

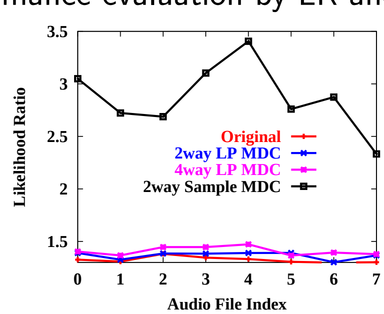


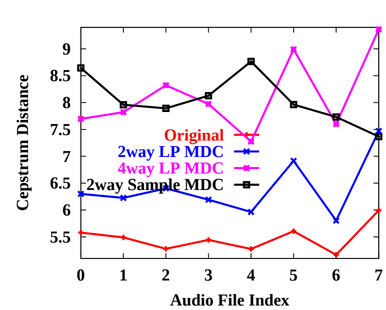
Further quality degradation for excitation parameters

Synthetic Tests — Without Loss

All descriptions are received

- No aliasing
- Linear prediction precision same as SDC
- Excitation quality degraded due to extended subframe size
- Performance evaluation by LR and CD





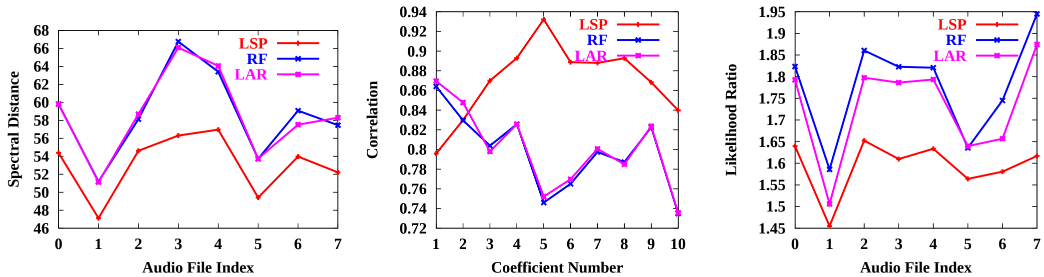
- Much better than sample-based MDC method

Results of Two-Way MDC With One Description Received

- Reconstruction of lost LP vectors based on one of the three representations
- Comparison using two extra measures:
  - Spectral distortion

$$SD = E\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|10\log_{10}|H_{o,\,n}(\omega)|^2 - 10\log_{10}|H_{r,\,n}(\omega)|^2\right|d\omega\right] \quad \text{dB}$$

  - Correlation
- LSP gives best reconstruction quality

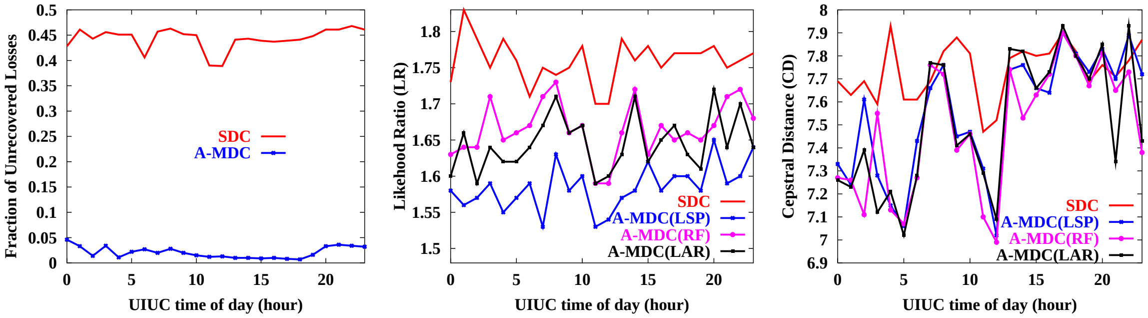


Internet Test Setup

- Components:
  - Sender
  - Receiver: 200 msec jitter buffer, start clock when first packet arrives
  - Internet simulator: delay and drop packet according to traffic traces
- Comparison between:
  - SDC
  - Adaptive MDC: dynamically switch between two-way and four-way MDC depending on loss conditions
- Comparison metrics:
  - Quality in LR and CD
  - Fractions of unrecoverable losses

Internet Tests

UIUC-Slovakia



Summary of adaptive MDC:

- Recovering the decoding state
- LSP best overall
- Effective in reducing unrecovered losses

Conclusions

- MDC design by correlation analysis
- LP-based MDC for low bit-rate linear predictive speech coders
- Optimizing LSP reconstruction
- Improving MDC excitation quality
  - Improving LSP reconstruction using second-order approximation
  - Modifying perceptive weighted filters to de-emphasize coding noise in format regions