LOSS CONCEALMENTS FOR LOW-BIT-RATE PACKET VOICE IN VOIP

Benjamin W. Wah

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

December 15, 2004

Outline

- Voice-over-IP: status and problems
- Loss concealment problem
  - Previous work
  - IP voice traffic loss characteristics
- Loss concealments for low bit-rate coded speech
  - Parameter-based MDC
  - Improving MDC quality
- Summary
- Future outlook

Application Areas of VoIP (in chronological order)

- Internet telephony
- PC-to-phone services
- Teleconferencing
- Phone-to-phone calling cards
- Telecommunication industry using VoIP
- Consumer Broadband Telephony
- Hosted IP PBX
- Wi-FI VoIP
- 4G Wireless communications

Current Status of VoIP

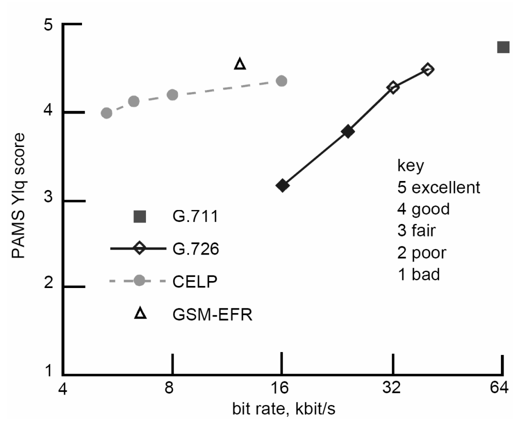
- Size of Business
  - Accounts for 10% of long-distance phone traffic around the world
  - Homes with broadband network using VoIP: 1% in 2004 (17% in 2009)
  - Sustainable expansion of market share is predicted, (1% per year)
  - Large long-distance carriers started using VoIP due to cost efficiency
  - Competition from small companies offering free or inexpensive VoIP
- Business Strategy
  - Initial business strategy, low-cost low-quality alternative to PSTN
  - Current strategy, equivalent quality, inexpensive substitute, additional features
- Requirements for VoIP to be mainstream
  - VoIP technology to be transparent (and easy to use) to users
  - Toll quality, inexpensive
  - No PC should be required, IP phones should be inexpensive
  - Extra features: image transfers, multicasting, broadcasting

VoIP Speech Quality

- Interactive real-time communications:
  - End-to-end delay due to codec, network, and jitter buffer
  - ITU G.114: one way delay, $<$ 150 ms acceptable, $<$ 400 ms noticeable
  - Mobile loop one-way delay about 100 ms; Mobile-VoIP-Mobile: about 300ms
- Acoustic echo: due to PSTN wiring or PC setup
  - Noticeable for delays more than 30ms
- Loss: some degradations on voice samples tolerable
  - Low bandwidth/congestion: due to dial-up connections, other streaming media
  - Long-burst or frequent short-burst intolerable
- Codec
  - Codec in tandem: code conversions at hosts or gateway, causing degraded quality and increased delay
  - Using PC as phone: Speaker and microphone not optimal for phone conversation
  - Standard low bit-rate speech codecs: Error propagation

Voice Codecs

Codec	Kbps	Coding Technique
G.711	64	Pulse code modulation (PCM)
G.726	40-16	Adaptive differential PCM (ADPCM)
G.728	16	Low-delay code excited prediction (LD-CELP)
G.729	8	Algebraic code-excited linear prediction (ACELP)
G.723.1	6.3/5.3	Multi-pulse max likelihood quantization (MP/MLQ)/ACELP
GSM FR	13	Regular pulse-excited long term predictor (RPE-LTP)
GSM EFR	12.2	Algebraic code-excited linear prediction (ACELP)



Ref: Reynolds and Rix: Quality VoIP

Network Environments: Packet Network

- IPv4: best-effort, no real-time support
- Packet size: less than MTU to avoid fragmentation
- Packet rate: 20 - 30 packets per second
- IPv6: best-effort, may support real-time traffic
- Wireless: future IP-based
- Loss unavoidable in packet networks

Network Environments: Transport-Layer Protocol

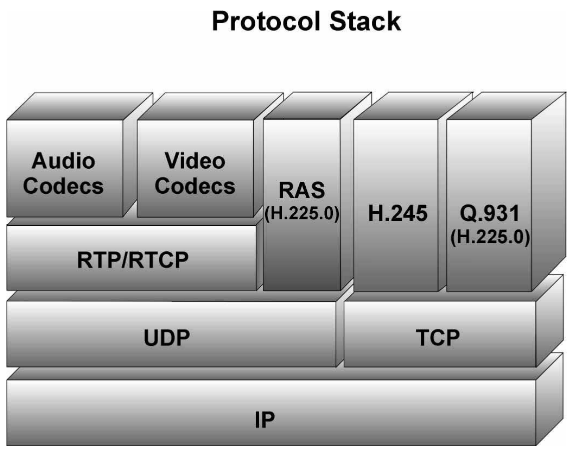
- TCP
  - Reliable but not suitable for real-time
  - Connection oriented, more secure
  - Allowed through firewalls
- UDP
  - Lossy and unreliable
  - No congestion control mechanism to slow the flow
  - Not permitted through firewalls
- TCP in real-time mode

  Provides connection-oriented transmission without congestion avoidance

  Suitable for current VoIP systems for firewall penetrability
- Loss of real-time voice not handled at the transport layer

Network Environments: Application-Layer Protocol

- H.323: umbrella standard for interoperability



- RTP: no loss recovery scheme
- Loss of real-time voice not handled

Packet losses in real-time voice communications left for end-point applications

Solutions for Improving Speech Quality

- Echo cancellation implementation in software for VoIP applications
- Jitter buffer at receiving end
- Easier access to broadband connection
- Both ends agree on a codec while initializing a VoIP session
- Dedicated IP Phones
- Improved Codecs with low delay and lower bit-rate requirements
- New speech coding standards developed for IP networks

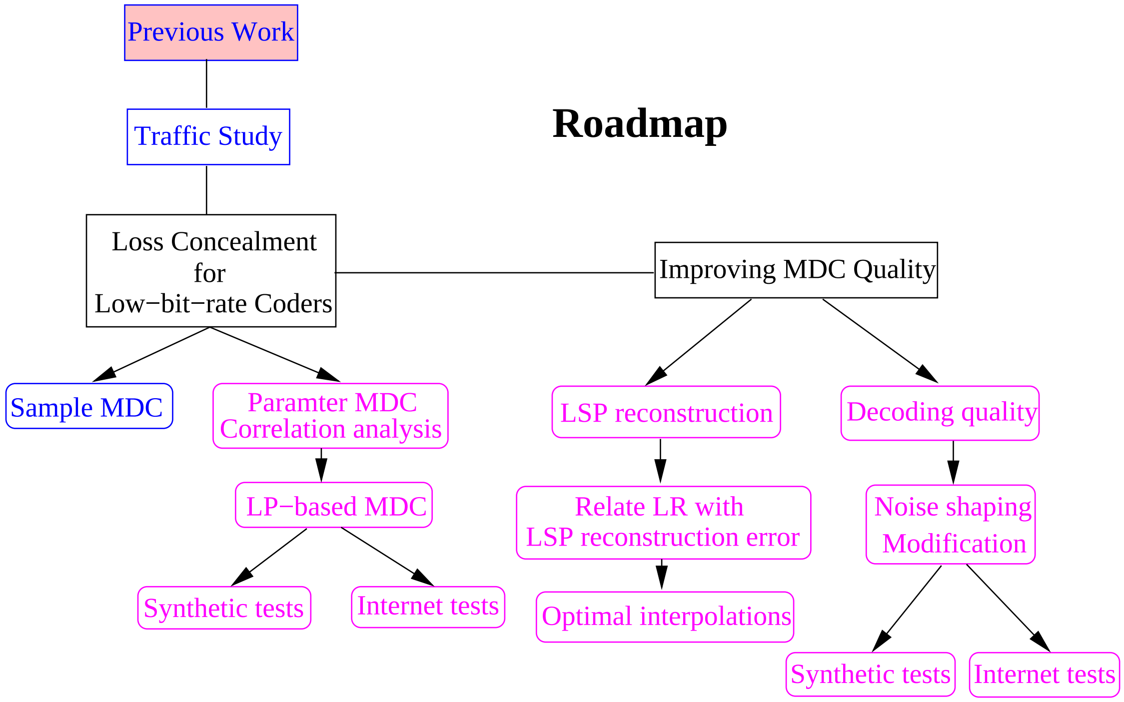
Outline

- Voice-over-IP: status and problems
- Loss concealment problem
  - Previous work
  - IP voice traffic loss characteristics
- Loss concealments for low bit-rate coded speech
  - Parameter-based MDC
  - Improving MDC quality
- Summary
- Future outlook

Loss Concealment Problem

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

- Allow reliable and real-time low bit-rate voice transmissions
- Unreliable IP networks, like the Internet and wireless wide area networks

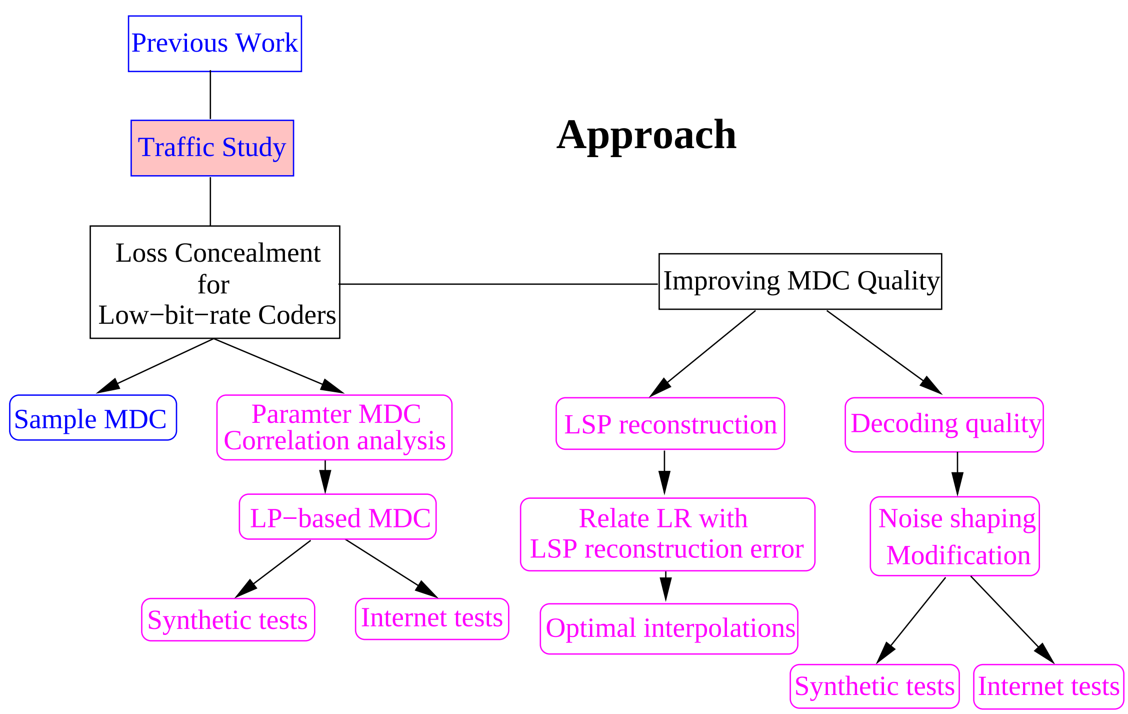


Previous Work: Coder-Independent Schemes

- 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: Coder-Dependent Schemes

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

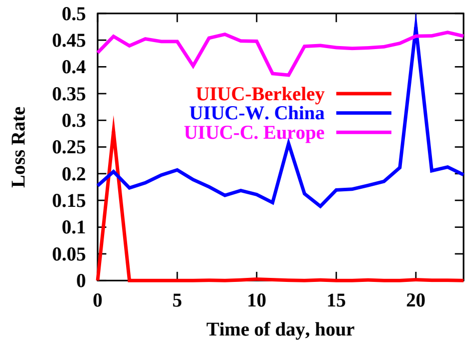


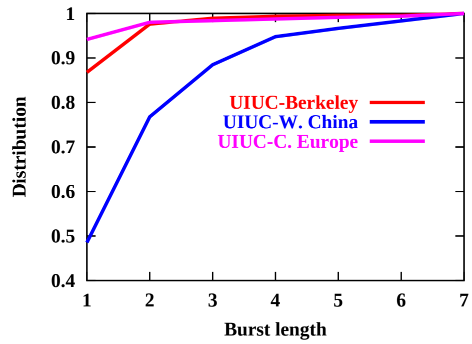
IP Voice Traffic Loss Characteristics

- Example connections

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

- Loss behavior

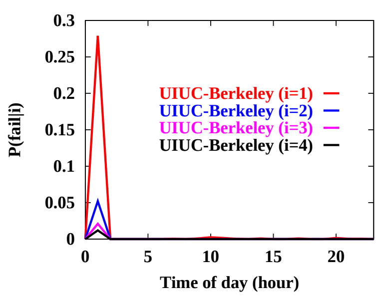


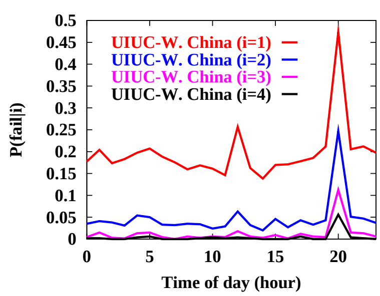


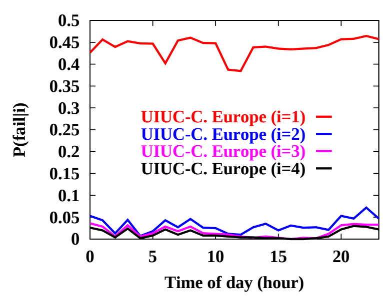
Loss rate can go up to 50% Most losses have short burst lengths

Reducing Unrecoverable Loss by Interleaving

- Bursty losses are difficult to handle
- Interleaving: disperse bursty losses to isolated losses
- $P(fail \mid i)$: prob. of losses that cannot be recovered under interleaving factor i





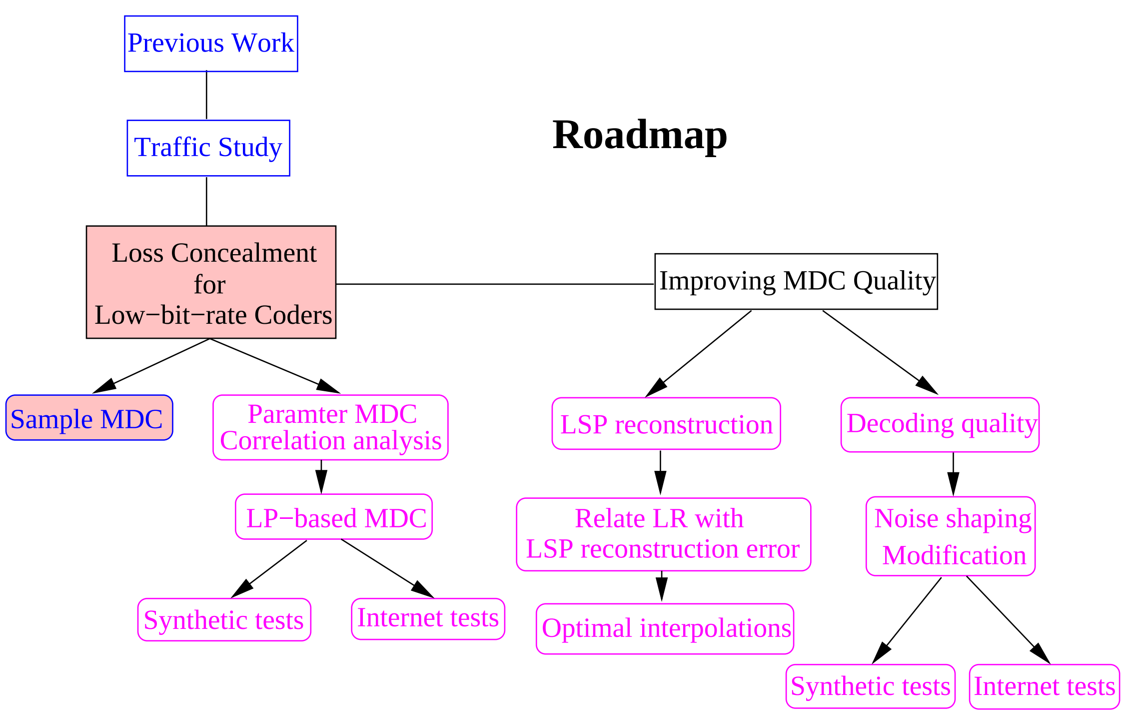


- Small interleaving factor 2 – 4 is enough
- Multiple-description coding is promising

Outline

- Voice-over-IP: status and problems
- Loss concealment problem
  - Previous work
  - IP voice traffic loss characteristics
- Loss concealments for low bit-rate coded speech
  - Parameter-based MDC
  - Improving MDC quality
- Summary
- Future outlook

Roadmap



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

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

  - a_r: vector of LP coefficients of reconstructed speech
  - a_0: vector of LP coefficients of original speech
  - R_0: correlation matrix derived from original speech

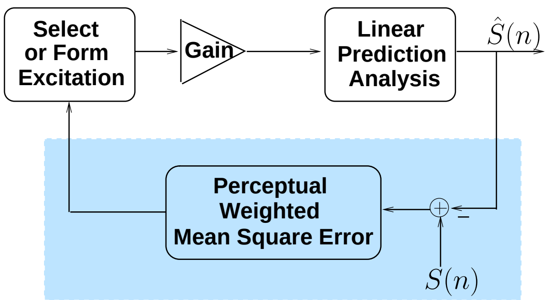
- Cepstral distance:

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

  - $c_{o,0}$: cepstra of original samples
  - $c_{r,0}$: cepstra of reconstructed samples

Typical Linear Predictive Coder

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

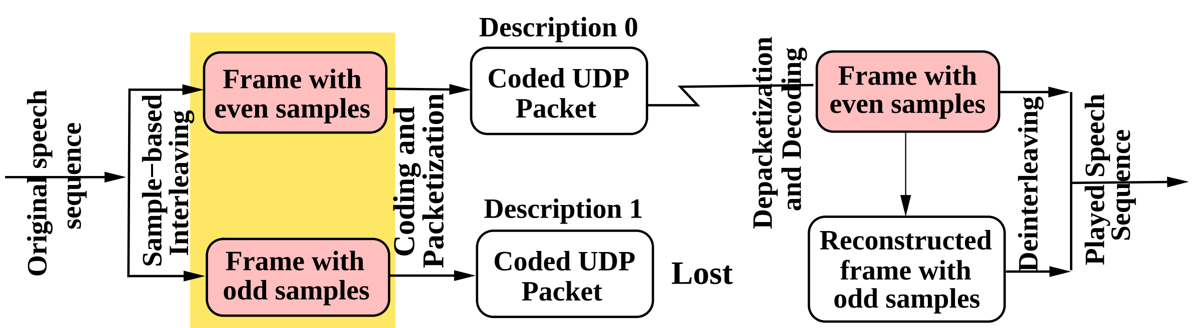


Major techniques:

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

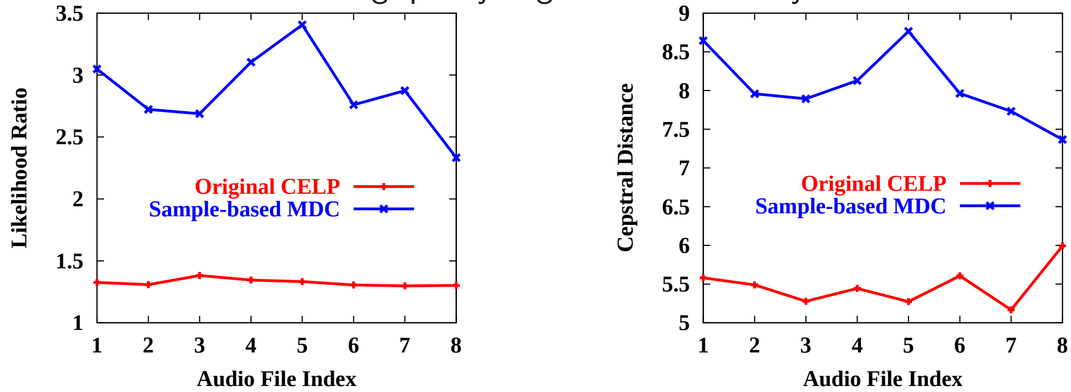
- FS CELP⋆, ITU G.723.1 ACELP, ITU G.723.1 MP-MLQ, and MELP

Coder-Independent Sample-Based MDC



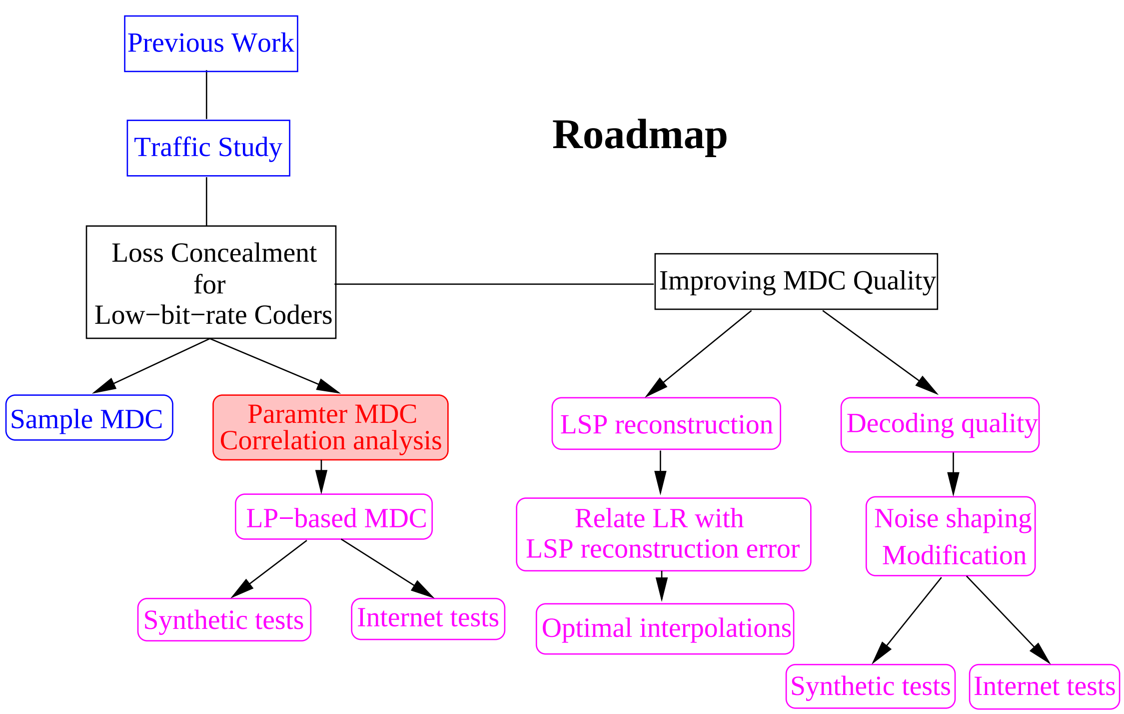
Performance of Coder-Independent Sample-Based MDC

Coding quality degrades dramatically

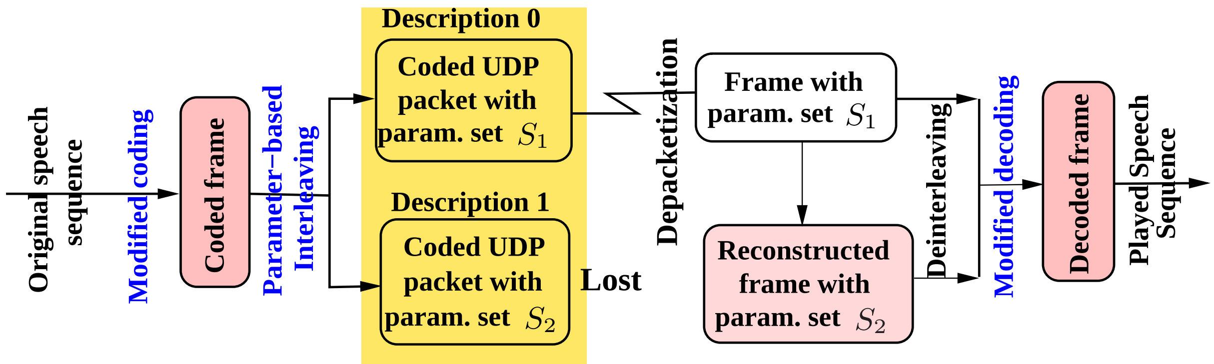


Drawbacks:

- 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

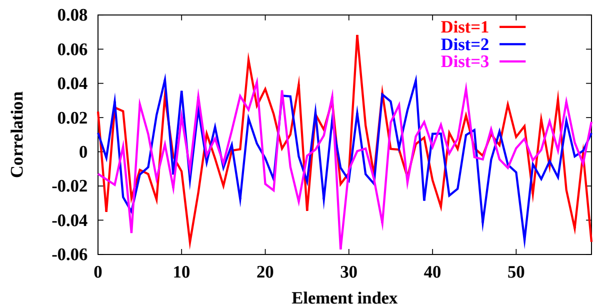
Sample Dist.	1	2	3
Correlation	0.83	0.60	0.35

Correlations of FS CELP Excitation Parameters

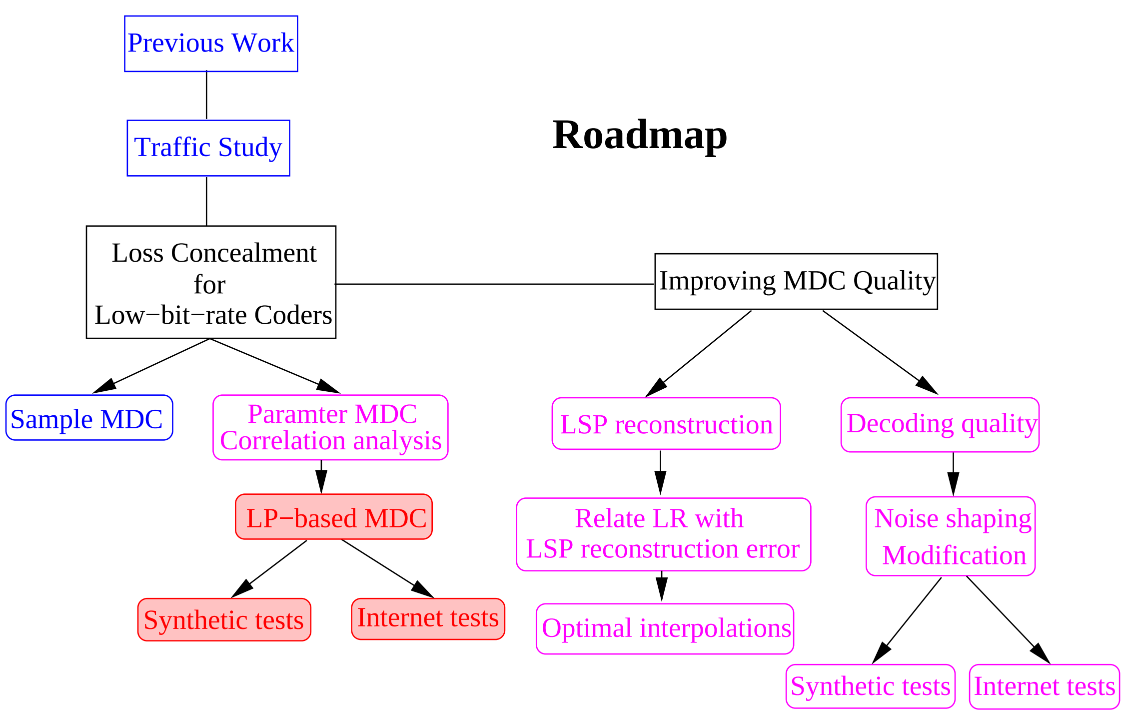
- Adaptive codewords: 2-element vector

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

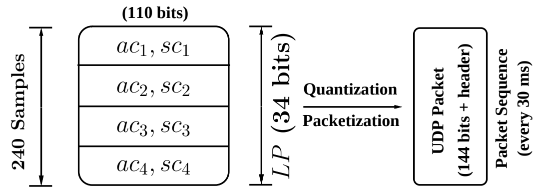


- Very low or no correlation for excitation parameters

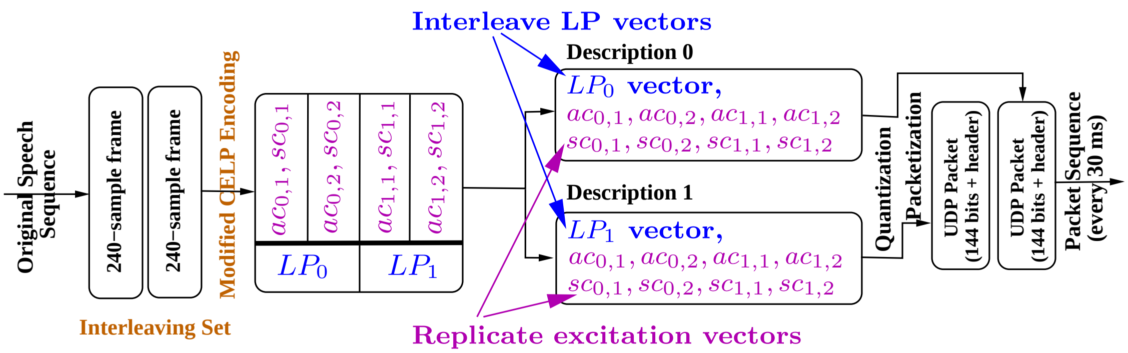


FS CELP SDC and LP-Based Two-Way MDC (US Patent 6754203 B2)

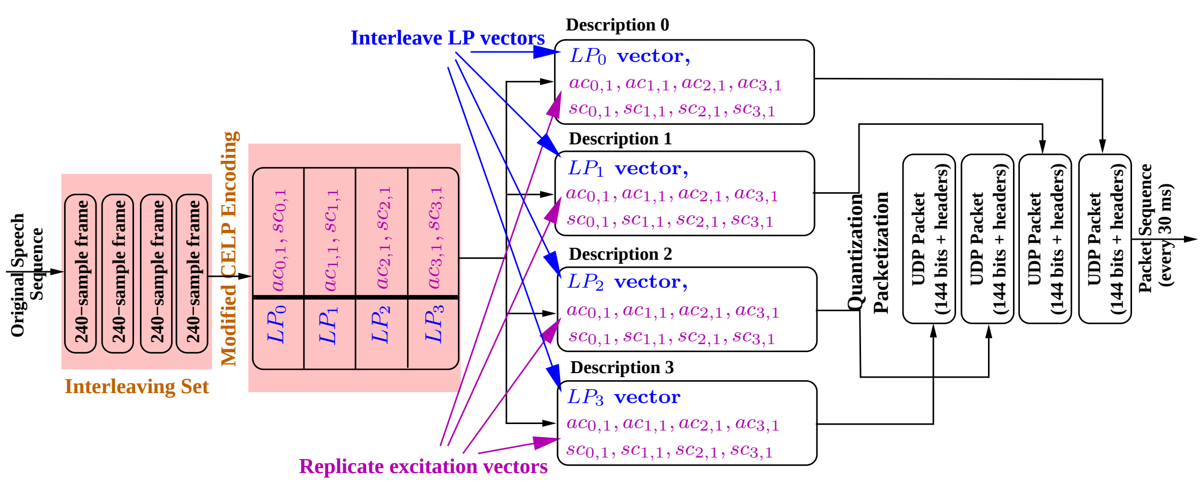
- FS CELP SDC:



- Construction of two-way MDC (with the same bandwidth as SDC):



LP-Based Four-Way MDC

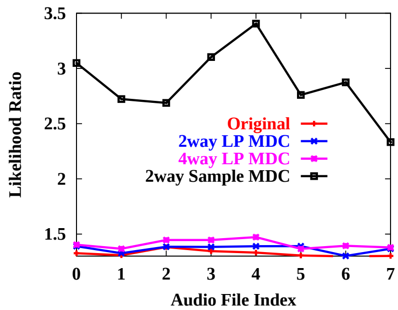


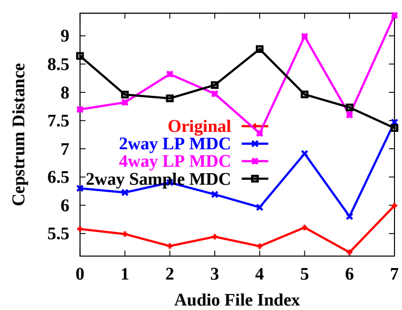
- Further quality degradation with longer subframes and the same packet size
- No quality degradation if 60-bit subframes but larger packet size are used

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





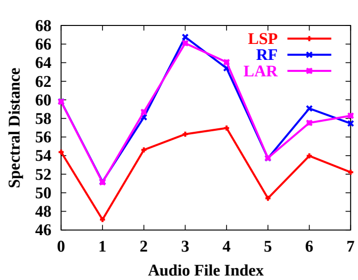
- Much better than sample-based MDC method

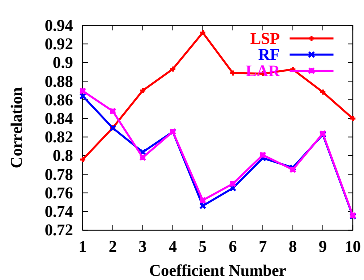
Results of Two-Way MDC With One Description received

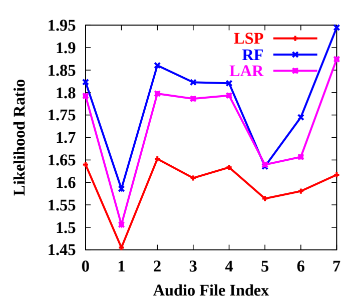
- Reconstruction of lost LP vectors based on one of the three representations
- Comparison using two extra measures:
  - Spectral distortion

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

  - Correlation
- LSP gives best reconstruction quality





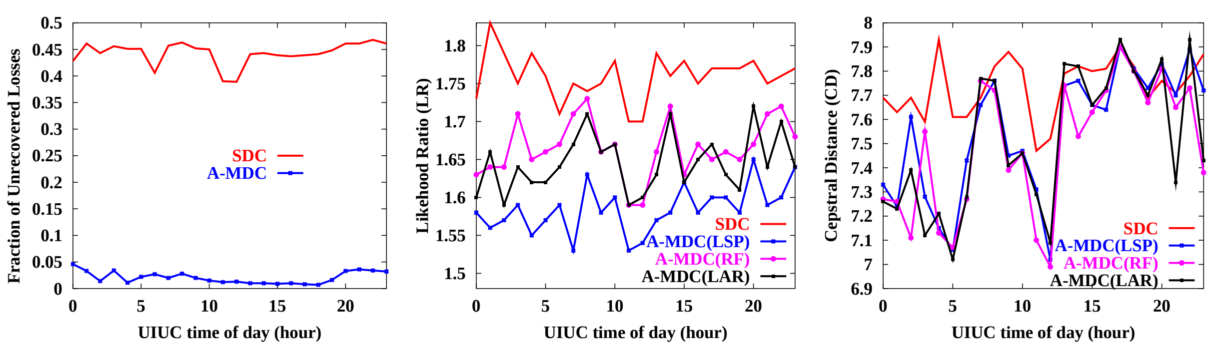


Internet Test Setup

- 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

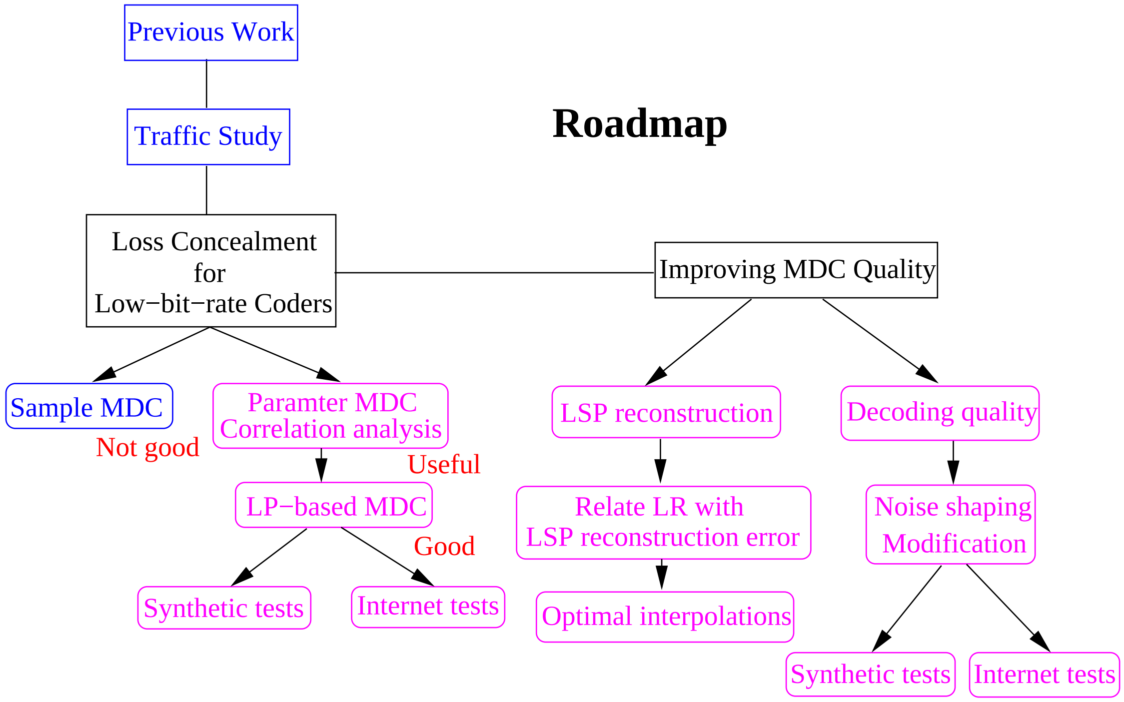
UIUC-Central Europe



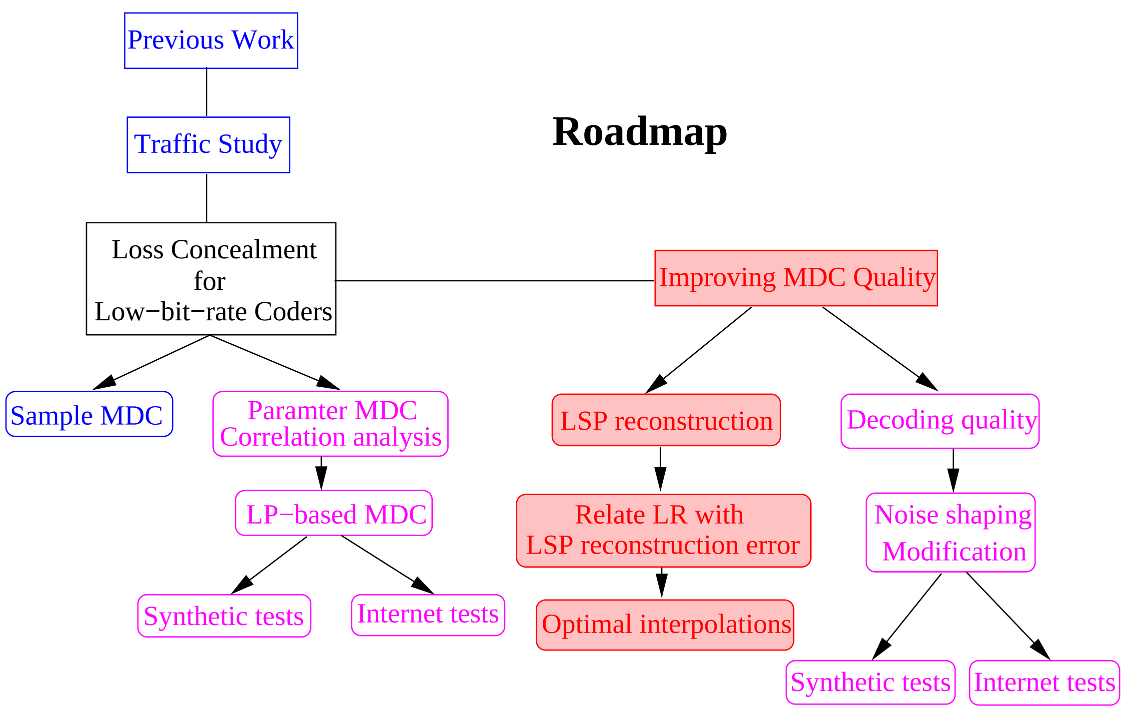
Summary of adaptive MDC:

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

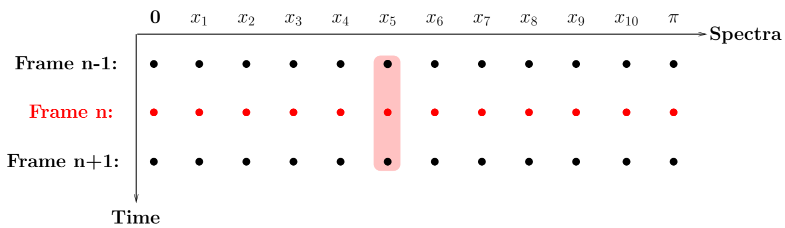
Roadmap



Roadmap

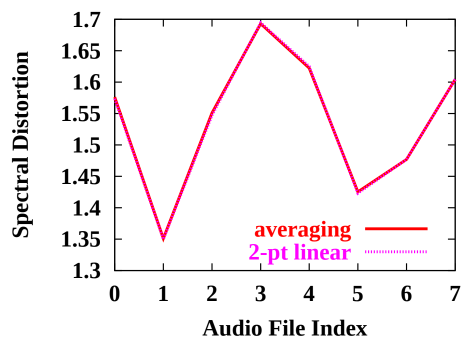


Optimized Two-Point Linear Interpolation



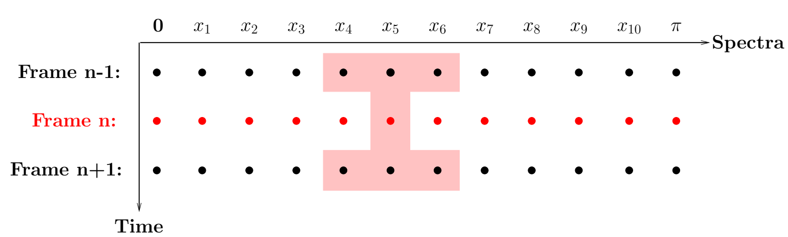
$$\vec{x}_r^n = \alpha \vec{x}^{n-1} + \beta \vec{x}^{n+1}$$

Averaging: $\alpha = \beta = 0.5$

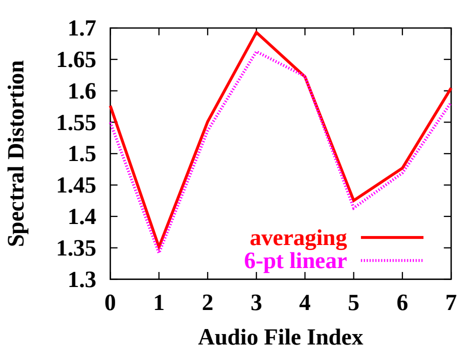


Averaging is almost optimal for two-point linear interpolation

Optimized Six-Point Interpolations

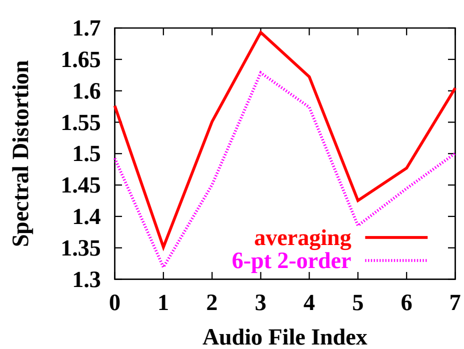


Six-point linear interp.



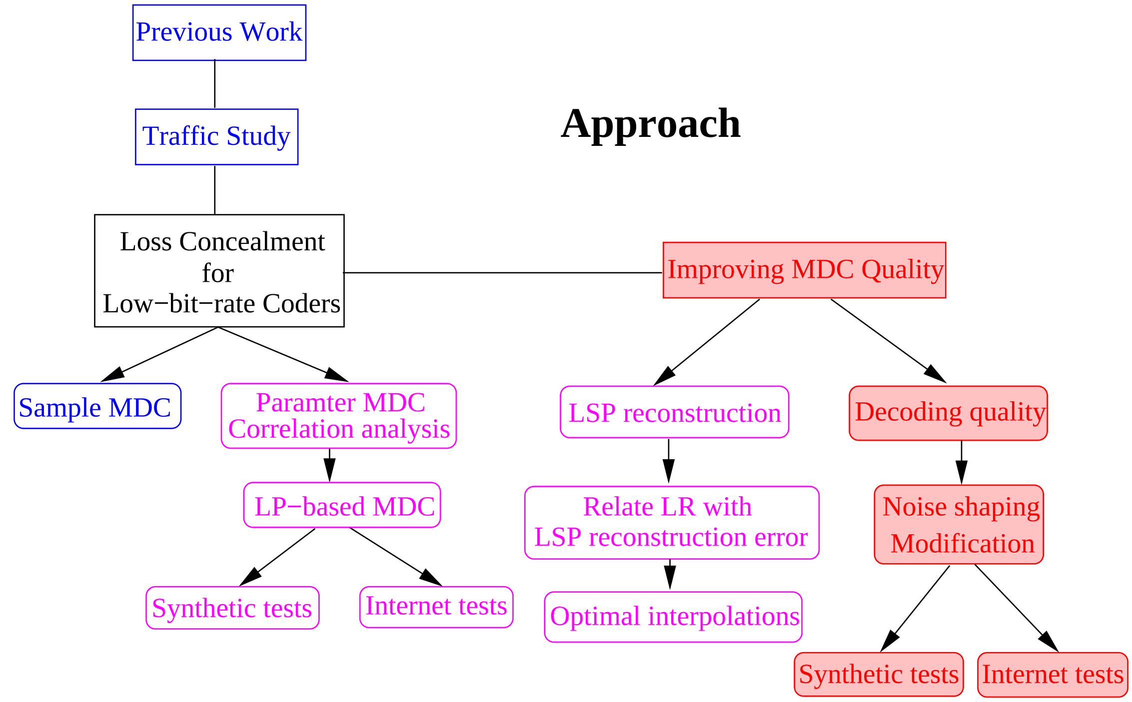
Averaging is near optimal

Six-point 2-order interp.

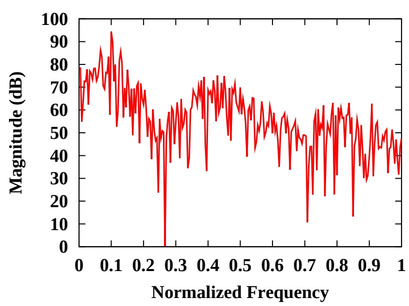


Not too much improvement

Approach



Causes for Quality Degradation



Speech perception:

- Valley noise more noticeable
- Formant important

- Significant higher coding-noise inside formant regions due to MDC

a) Two-way MDC

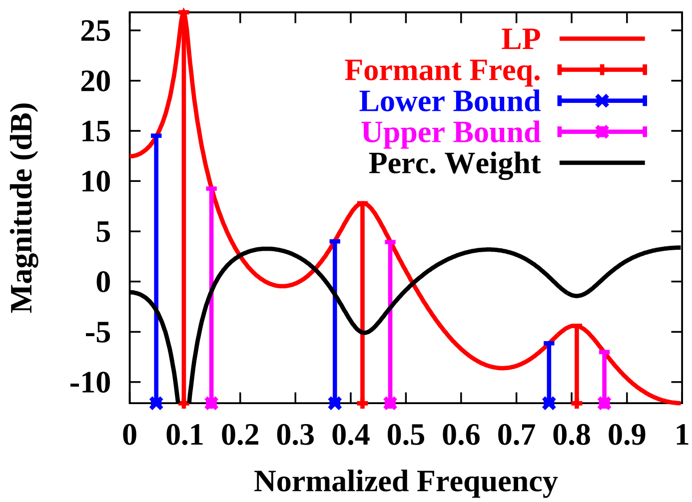
	E_{R_F}	$E_{\overline{R_F}}$
SDC	1.1591e+8	2.3976e+7
Two-way MDC	2.3049e+8	4.4340e+7
Ratio	1.99	1.85

b) Four-way MDC

	E_{R_F}	$E_{\overline{R_F}}$
SDC	2.3507e+8	4.4775e+7
Four-way MDC	8.6771e+8	1.2685e+8
Ratio	4.69	3.83

Perceptual Weighting Filter

- Goal: noise shaping
  - De-emphasize coding noise inside formant regions



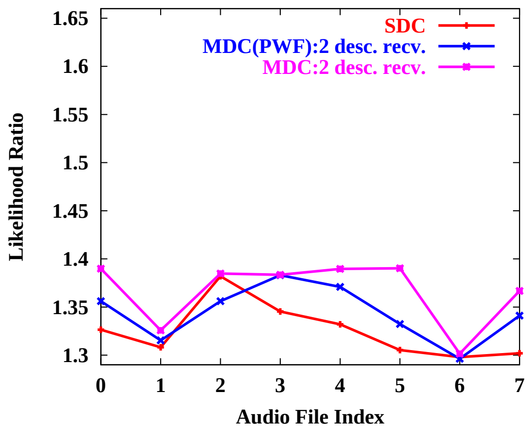
Modification to Perceptual-Weighting Filter

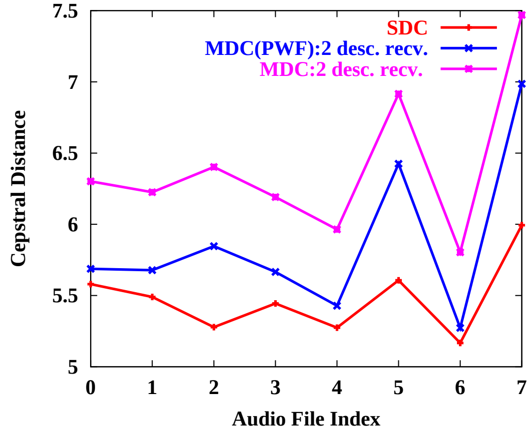
- Modification:
  - Decrease suppression of noises inside formant regions
  - Shift noise outside formant regions
  - Use SDC noise balancing as reference
- Choosing suitable perceptual-weighting filter (γ is the filter parameter)

	E_{R_F}		$E_{\overline{R_F}}$	
	Mag.	Ratio	Mag.	Ratio
SDC	1.1591e+8		2.3976e+7	
two-way MDC ($\gamma = 0.8$)	2.3049e+8	1.99	4.4340e+7	1.85
two-way MDC ($\gamma = 0.82$)	2.2368e+8	1.93	4.3659e+7	1.82
two-way MDC ($\gamma = 0.84$)	2.1501e+8	1.85	4.4372e+7	1.85
two-way MDC ($\gamma = 0.86$)	2.1810e+8	1.88	4.5733e+7	1.91
two-way MDC ($\gamma = 0.88$)	2.0096e+8	1.73	4.6161e+7	1.92
two-way MDC ($\gamma = 0.9$)	2.0098e+8	1.73	4.7287e+7	1.97

Synthetic Tests for Improved Perceptual-Weighting Filter

Two-way MDC when both descriptions received

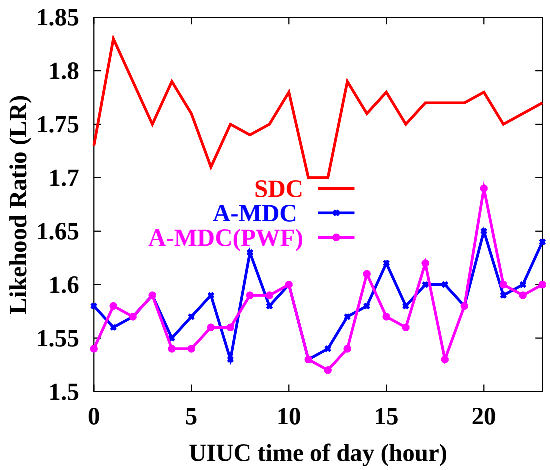


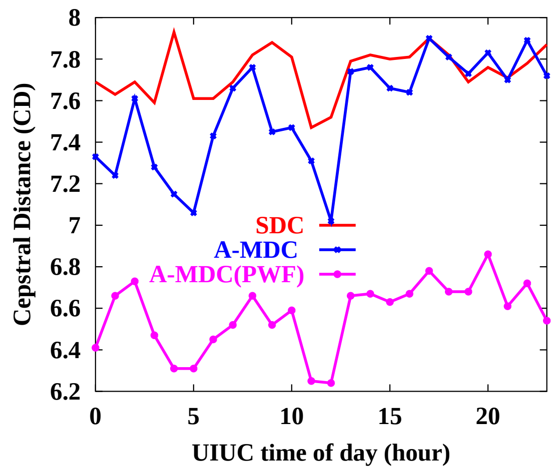


- LR similar
- Noticeable improvements in CD

Internet Tests for Improving Perceptual-Weighting Filter

UIUC-Central Europe





- SDC with no loss: $LR = 1.33$, $CD = 5.55$
- Improved CD

Summary

- Summary
  - MDC design by correlation analysis
  - LP-based MDC for low bit-rate linear predictive speech coders
  - Optimizing LSP reconstruction
  - Improve MDC excitation quality
- Future work
  - Further improve MDC quality
  - Bandwidth and quality tradeoff
  - Rate adaptation

Future Outlook

- Commercial products and services available
  - VoIP solutions for dial-up with poor quality and broadband with acceptable quality
  - Net2Phone, Skype, Netmeeting
  - Broadband connection and low delay $\Rightarrow$ success
- Future research directions
  - Mobile endpoint increases delay
  - Wireless broadband delay problem not solved
- New Application Areas
  - The use of VoIP technology combined with other multimedia for a complete virtual meeting
  - Microsoft announced Project Istanbul for Summer 2005