Speech-Adaptive Layered G.729 Coder for Loss Concealments of Real-Time Voice Over IP

Batu Sat and Benjamin W. Wah

Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Email: wah@uiuc.edu
URL: http://manip.crhc.uiuc.edu

Outline

- Background
  - IP network environment
  - Implications on low-bit-rate encoded speech
- Problems Addressed
  - Protection and delivery methods
  - Coding and reconstruction by layers of perceptual importance
- Conclusions

IP Network Environment

- Packet loss rate
  - Connection-dependent and non-stationary
  - International destinations: higher losses (some >50%)
- Bursty losses
  - Three or more bursty losses observed
- End-to-end delays
  - Packet transmission delays highly varying (~100 ms)
  - ITU G.114: <150 ms desirable; >400 ms unacceptable
- Packet rate
  - High IP packet rate (e.g. 100 packets/sec) increases loss rate
- Packet size
  - Does not affect packet loss rate (size within MTU)

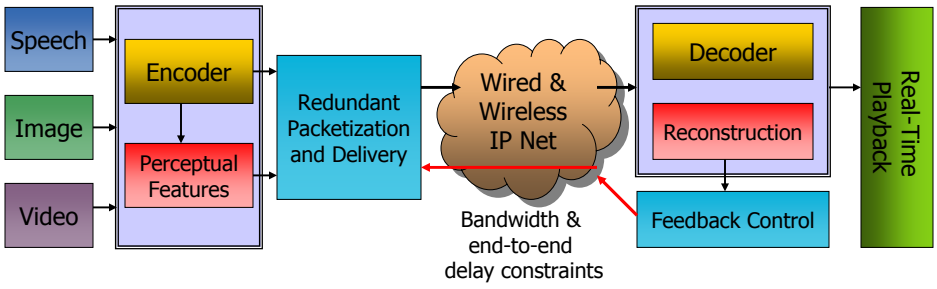
Implications on Low Bit-Rate Speech

- Low bit-rate encoded speech
  - ITU G.729 speech frames at 10 ms each
  - Dependencies across frames (for coding efficiency)
    - Effect of a lost frame propagates to many subsequent frames
- Robust delivery
  - Jitter buffers used to smooth out irregular arrivals
  - Multiple frames (GOF) placed in a packet (to reduce packet rate)
    - Single packet loss → multiple frames in close proximity lost
  - Piggyback copies of GOFs transmitted in the past in the current packet
    - End-to-end delay constraint: only a few replicas are relevant

Goals

- Develop coding, protection, and reconstruction schemes
  - For transmitting low bit-rate speech data with high perceptual quality
  - In lossy IP networks under constraints on end-to-end delay and bandwidth

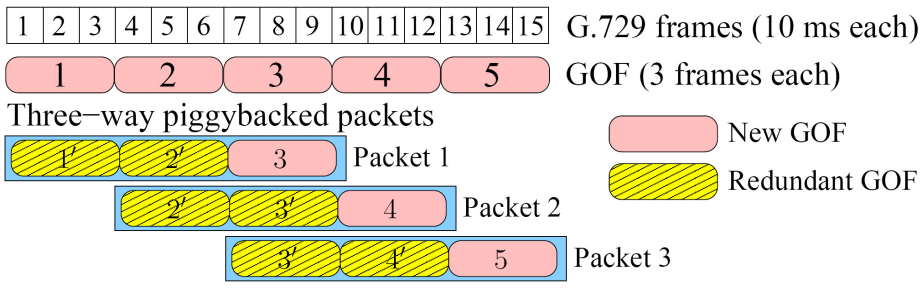
Approach



Issues Addressed

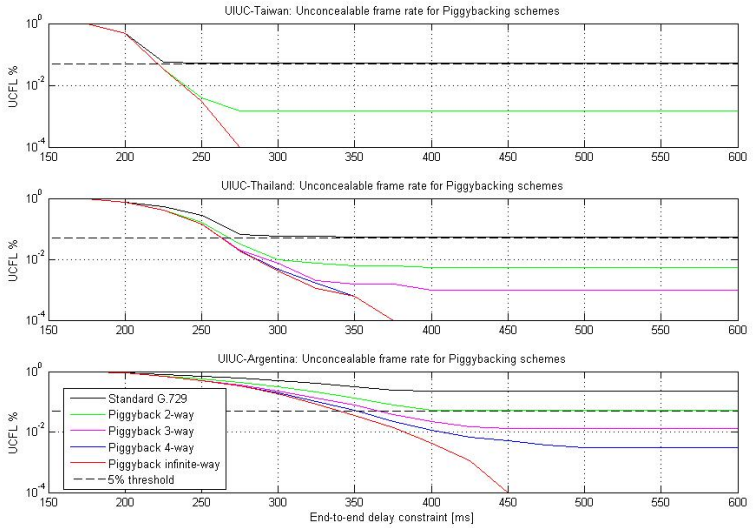
- Protocol with redundant delivery and dynamic adaptation
  - Tradeoffs among redundancy, end-to-end delays, loss concealments, and overhead in feedbacks
- Coding and reconstruction methods by layers of perceptual importance
  - Tradeoffs between perceptual quality and bit budget

End-to-End Delivery with Redundancy



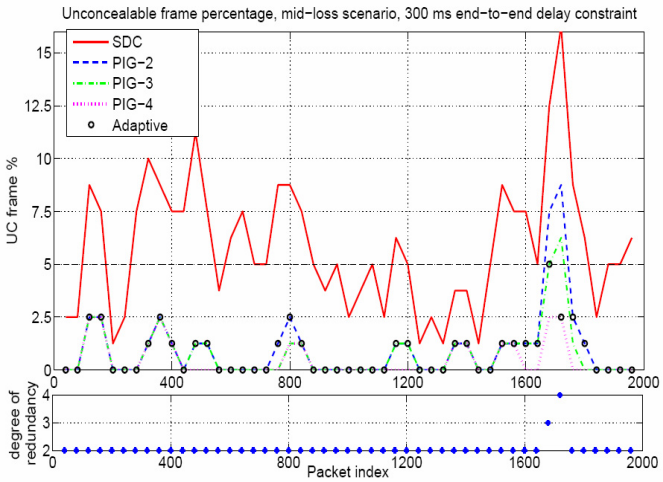
Tradeoffs between Delay and Loss Concealment

Network-adaptation: Pick the piggybacking degree and end-to-end delay with a tolerable unconcealable frame loss rate (UCFLR)



Network-Adaptive Redundant Piggybacking

UIUC ↔ Thailand



Pick the minimum piggybacking degree to achieve an end-to-end delay of 300 ms and an UCFLR of 5%

Infrequent feedbacks

Issues Addressed

- Protocol with redundant delivery and dynamic adaptation
  - Tradeoffs among redundancy, end-to-end delays, loss concealments, and overhead in feedbacks
- Coding and reconstruction methods by layers of perceptual importance
  - Tradeoffs between perceptual quality and bit budget

CELP Speech Coders

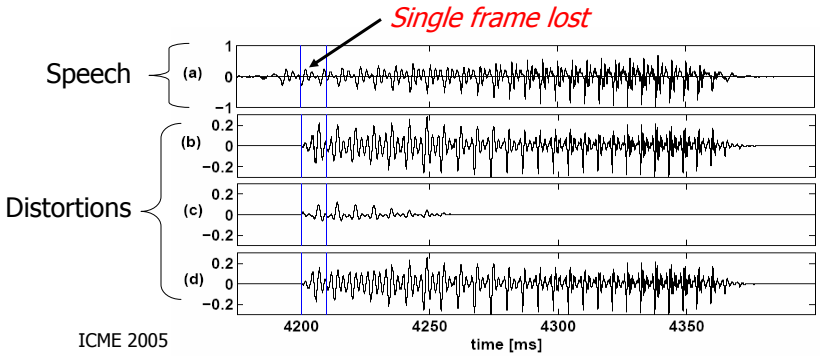
- Two types of encoded parameters
  - LPC
    - Predictive coding used (with 4-frame memory in G.729)
  - Excitation reconstruction
    - Algebraic codebook (ACB pulses and gain)
      - Individual pulses (significant amplitude)
      - Independent across frames
    - Adaptive codebook (pitch period and gain)
      - Extensive use of previous locally generated excitation signals
      - Vulnerable to propagation of errors

Standard Loss Concealment

- When a frame is lost
  - LPC: reconstruct from that of last frame
  - Excitation parameters of voiced frame
    - Reconstruct pitch period from past frame
    - Scale down pitch gain to reduce perception of loss
  - Excitation parameters of unvoiced frame
    - Generate randomly ACB pulses
    - Scale down ACB gain

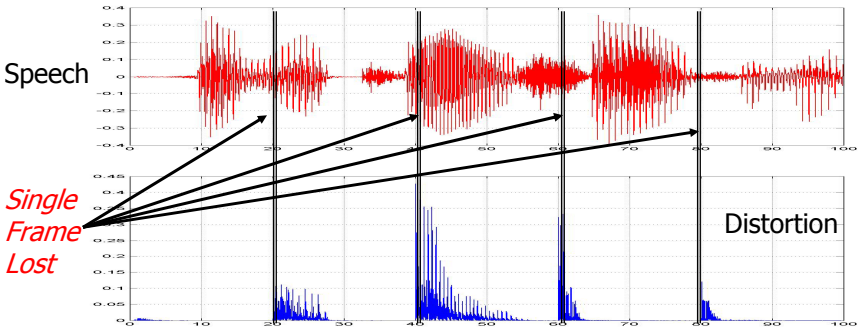
Inadequate Standard Loss Concealment

- Distortion due to single frame loss & standard error concealment
  - (b) All parameters lost: significant distortion till end of voiced region
  - (c) LPC parameters lost: shorter, insignificant distortion
  - (d) EXC parameters lost: similar to all parameters lost
- Conclusion: excitation is more important than LPC



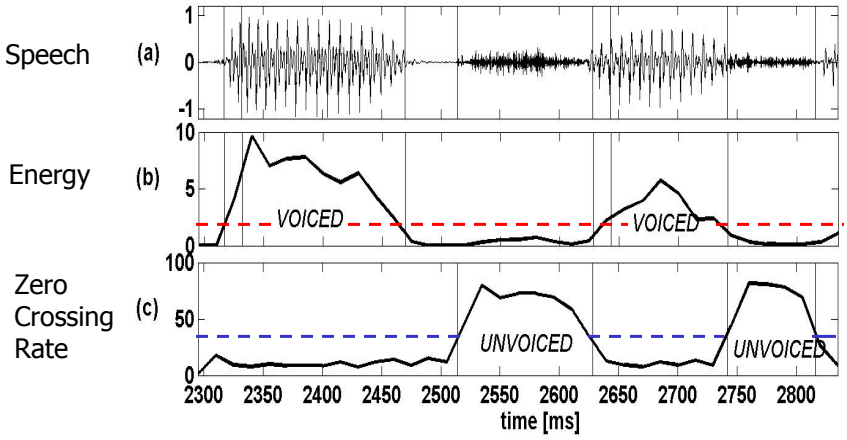
Effects of Frame Loss on Speech Classes

- Effect of frame loss is speech-class dependent
  - Loss in voiced frames: strong, long-lasting impact
  - Loss in unvoiced frames: strong local impact
  - Distortion lasts till voiced/unvoiced region boundary



Speech Classifier

- Heuristic speech classification algorithm



Voiced Segment: Important Parameters

- Both pitch period and gain are important and should be protected

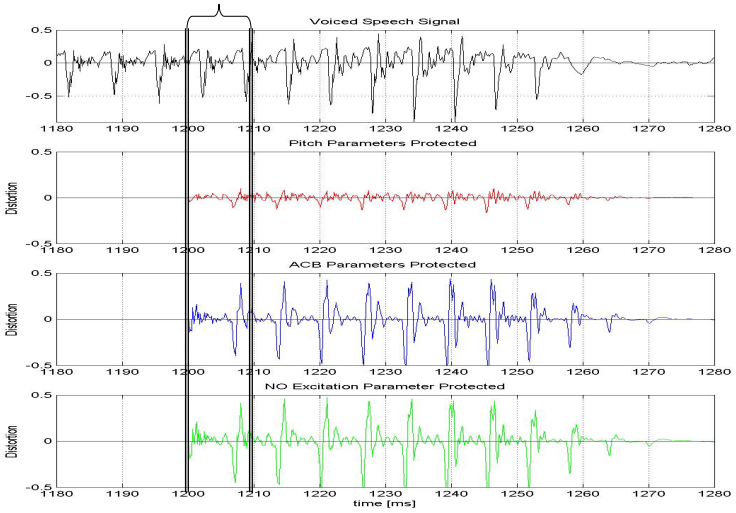
Single frame lost

Speech

Pitch protected

ACB protected

Nothing protected



Unvoiced Segment: Important Parameters

- Both ACB pulses and gain are important and should be protected

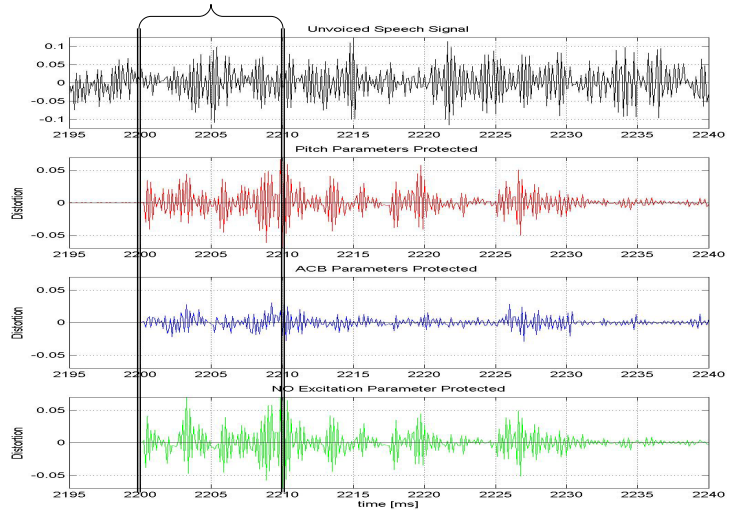
Single frame lost

Speech

Pitch protected

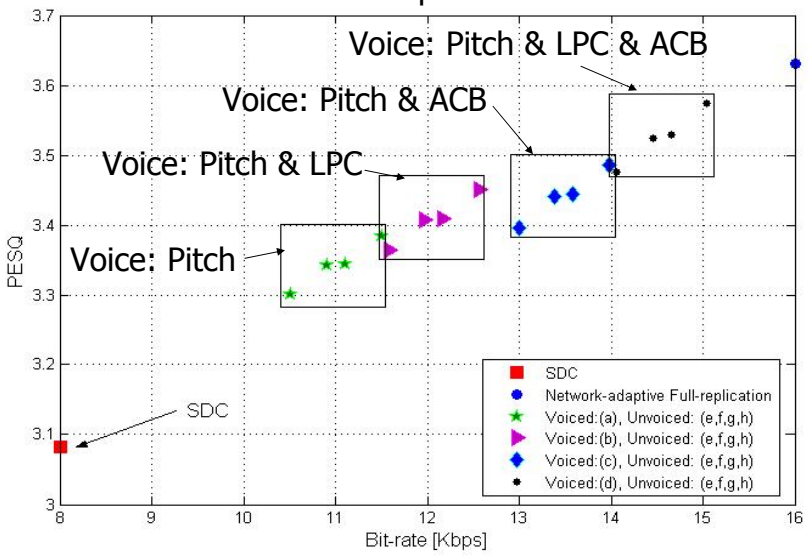
ACB protected

Nothing protected



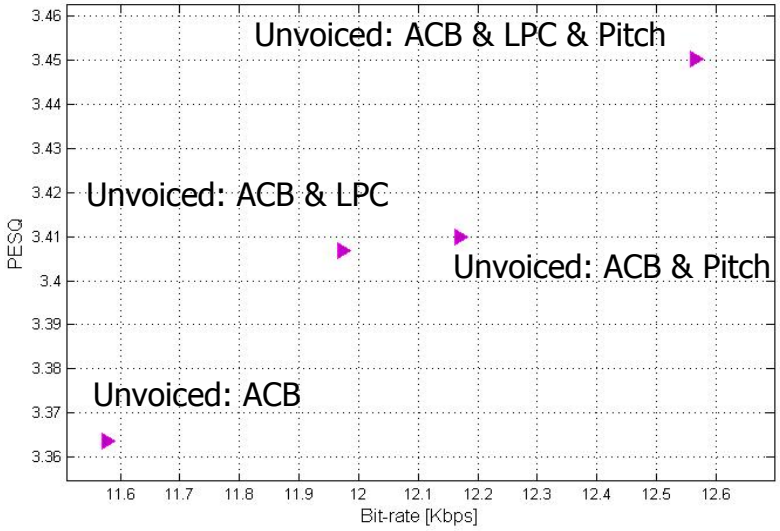
Perceptual Importance of Parameters (Voiced)

LPC is more important than ACB



Perceptual Importance of Parameters (Unvoiced)

LPC is more important than pitch



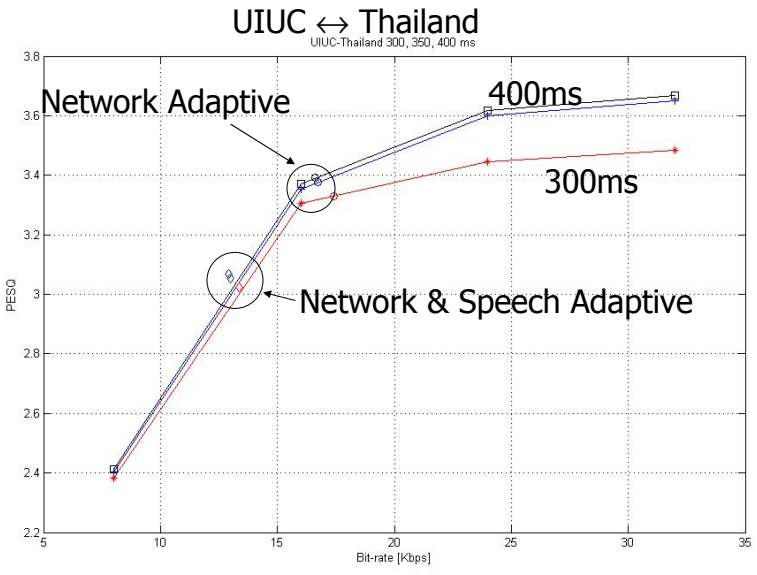
Perceptual Importance of Parameters

HIGH Importance	Voiced	Unvoiced	Onset
	Pitch LPC ACB	ACB LPC Pitch	ACB Pitch LPC

Our speech-adaptive layered-coding scheme

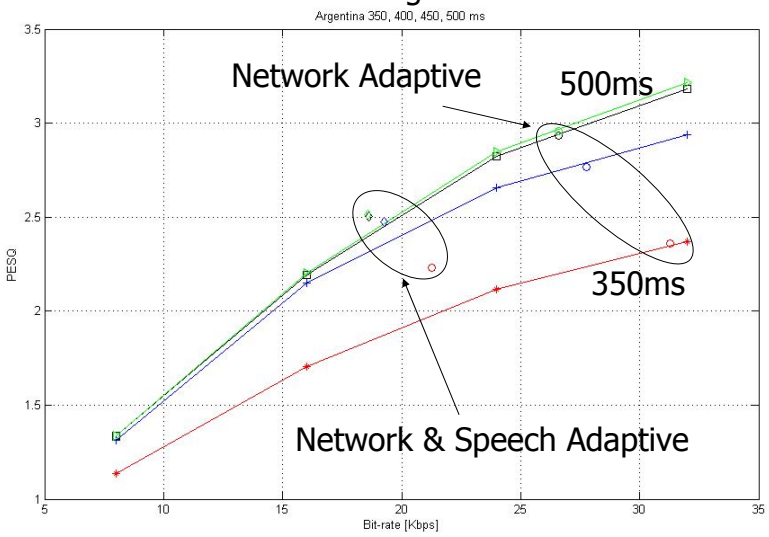
Layered Coding	Voiced	Unvoiced	Onset	Silence
Base Layer	Pitch, LPC	ACB, LPC, Pitch	ACB, Pitch, LPC	---
Enhancement Layer	ACB	---	---	ACB, LPC Pitch

Perceptual Quality vs. Bit-Rate Tradeoffs



Perceptual Quality vs. Bit-Rate Tradeoffs

UIUC ↔ Argentina



Conclusions

- Layered coding scheme to adapt dynamically to speech characteristics and network loss conditions
- Need finer classification of speech
  - Protect important & unpredictable parameters
  - Better choice of unconcealable frame loss rate