SPEECH- AND NETWORK-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
Urbana, IL 61801, USA
{batusat, wah}@uiuc.edu

Contributions

- Layered CELP speech coding scheme that adapts dynamically to the speech characteristics and the network conditions
- Redundant piggybacking to combat bursty losses and jitters while maintaining acceptable end-to-end delay
- Protecting only perceptually important parameters

IP Network Environment

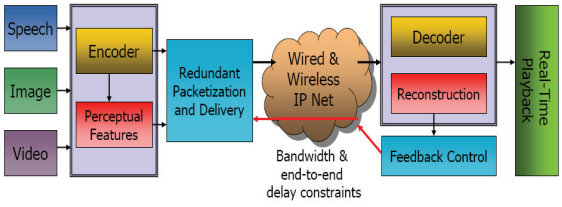
- Packet Loss Rate
  - Connection-dependent and non-stationary
  - International destinations: higher losses (some >50%)
- Bursty Packet Losses
  - Three or more bursty losses observed
- End-to-end Delays
  - Highly varying packet transmission delays (~100ms)
  - ITU G.114: <150 ms desirable; >400 ms unacceptable
- Packet Rate
  - High loss rate under high packet rate (e.g.100 pack/sec)
- Packet Size (within MTU)
  - No effect on packet loss rate

Implications on Low Bit-rate Speech

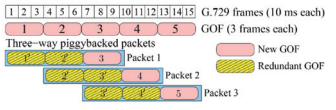
- Low bit-rate encoded speech
  - ITU G.729 speech frames at 10 ms each
  - Dependencies across frames (for coding efficiency)
- Robust delivery
  - Jitter buffers for smoothing out irregular arrivals
  - Multiple frames (GOF) placed in a packet (to reduce packet rate)
  - Past copies of GOFs piggybacked in the current packet
    - End-to-end delay constraint: only a few relevant replicas

Problem Statement & Approach

- Design of speech coding, protection and reconstruction scheme with a relaxed bit-rate requirement
  - For transmitting low-bit-rate speech data with high perceptual quality (ITU P.862: PESQ)
  - In lossy and non-stationary IP networks under constraints on end-to-end delay and bandwidth.

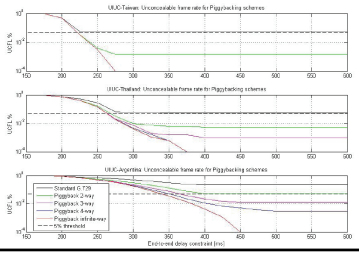


End-to-End Delivery with Redundancy



Piggyback Redundant Packing Scheme

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



Network-Adaptive Redundant Piggybacking

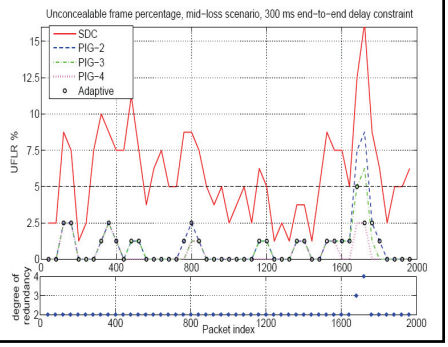
UIUC ↔ Thailand

Algorithm:

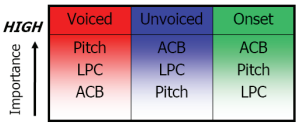
Using receiver feedbacks, pick the minimum piggy-back degree to achieve an end-to-end delay of 300ms and an UCFLR of 5%

Observation:

Infrequent feedbacks needed



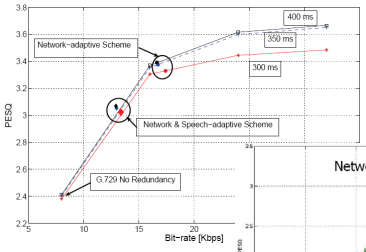
Perceptual Importance of Parameters



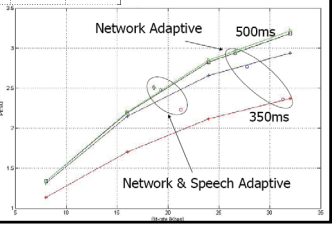
Our speech-adaptive layered-coding scheme

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

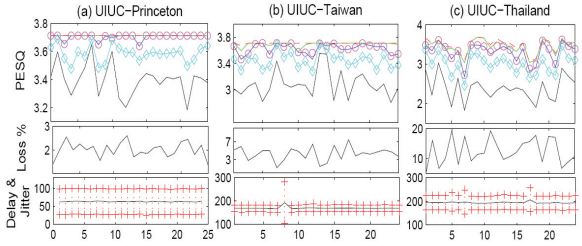


UIUC ↔ Thailand
300ms, 350ms, 400ms



UIUC ↔ Argentina
350ms, 400ms, 450ms, 500ms

Experimental Results



Conclusions & Future Work

- Layered coding scheme to adapt dynamically to speech characteristics and network loss and delay conditions
- Need finer classification of speech
  - Protect parameters that are both important and unpredictable at the receiver
- Better choice of unconcealable loss rate
- Compression of frames within a single GOF
- Uneven bit-budget for different GOFs in a packet
  - Older GOFs are rarely used → less budget

References

[1] B. Sat and B. W. Wah
Speech-adaptive layered G.729 coder for loss concealments of real-time voice over IP,. in *Proc. IEEE Int'l Conf. on Multimedia and Expo*, July 2005.

[2] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra
Perceptual evaluation of speech quality (PESQ)-a. new method for speech quality assessment of telephone networks and codecs,. *J. of the Audio Engineering Society*, vol. 50, no. 10, pp. 755.764, Oct. 2002.

[3] D. Lin and B. W. Wah
LSP-based multiple-description coding for real-time low bit-rate voice over IP,. *IEEE Trans. on Multimedia*, vol. 7, no. 1, pp. 167.178, Feb. 2005.

[4] ITU G.114

[5] ITU P.862 Perceptual evaluation of speech quality (PESQ)

[6] ITU G.729