Play-out Scheduling and Loss-Concealments in VoIP for Optimizing Conversational Voice Communication Quality

Batu Sat, Benjamin W. Wah

09/25/2007

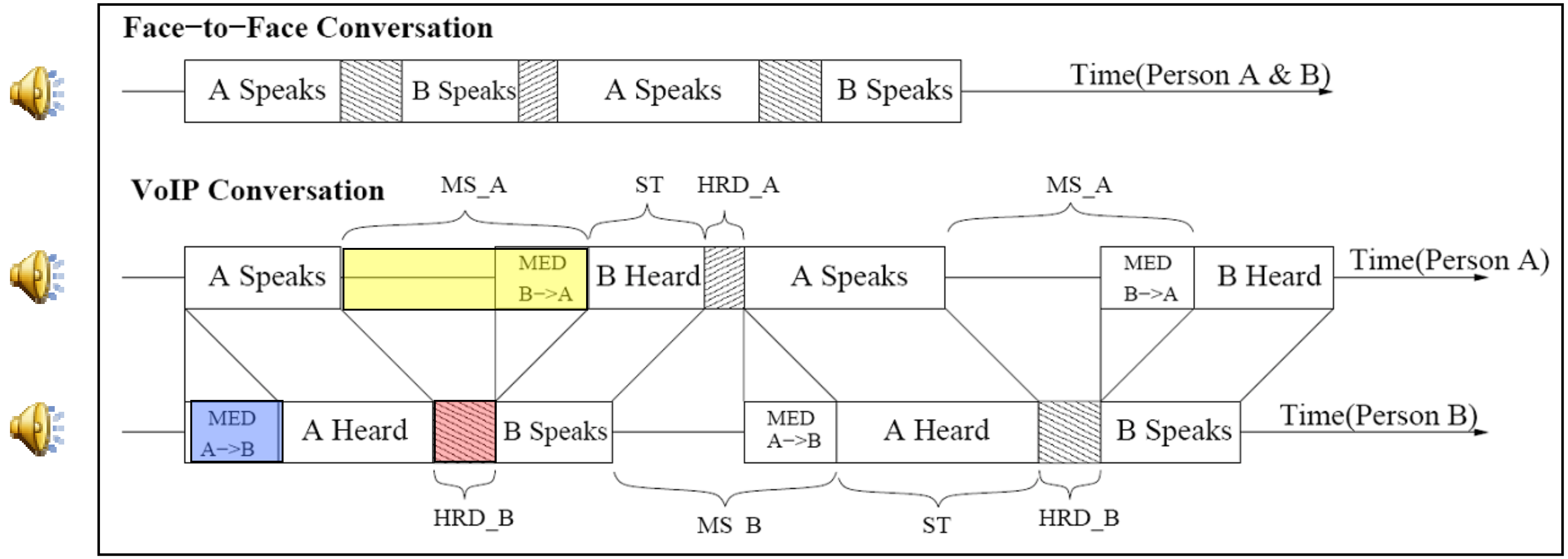
Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign, USA

Outline

- Introduction
  - Conversational dynamics
  - Conversational voice communication quality
  - Perceived effects of delay
- Network environment & network control
- Trade-offs in CVCQ attributes
- Previous work & our contributions
- Subjective tests
- Design of adaptive POS & LC schemes
- Experimental results

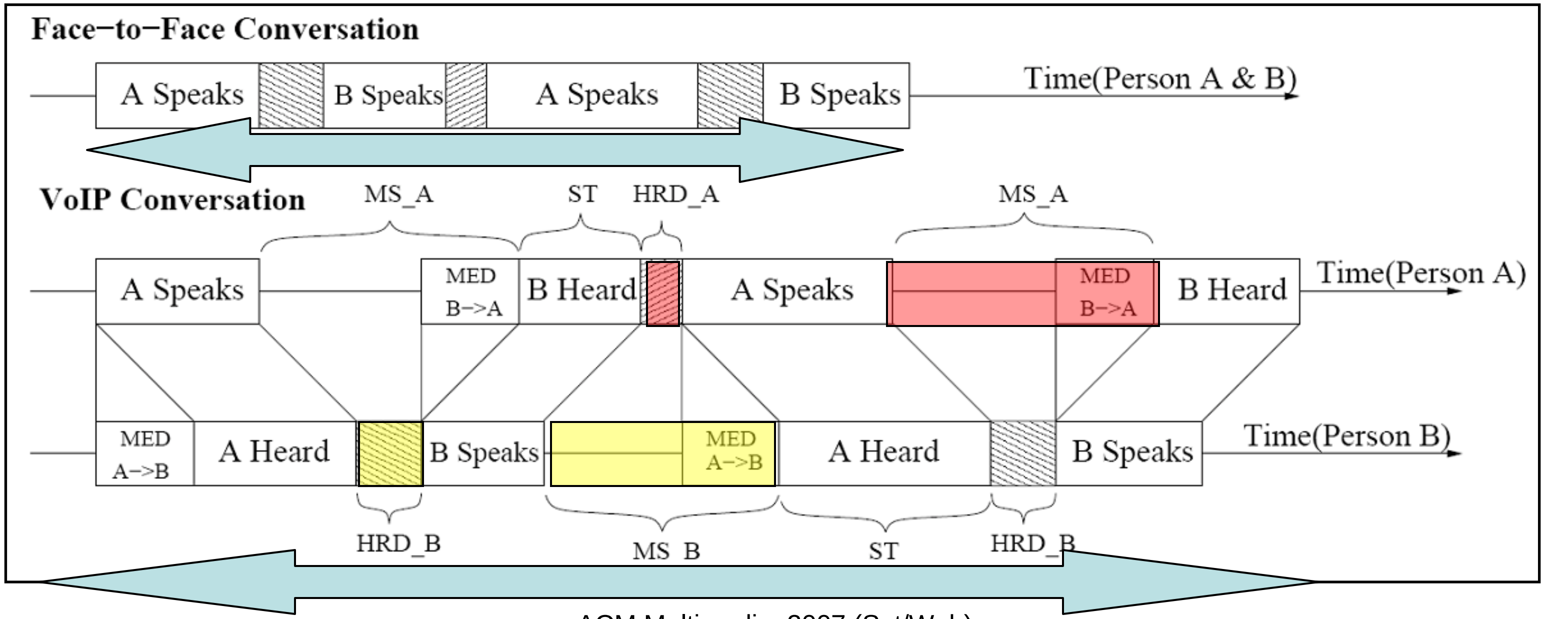
Conversational Dynamics

- Interactive conversation
  - One-way speech segments
  - Switches (turn-taking) between speakers
- Common perception of reality in face-to-face conversation
- Multiple realities in conversation over channel with delays
  - Participants perceive different timing and duration of speech & silence events
    - *Mouth-to-Ear Delay* (MED)
    - *Human Response Delay* (HRD): Duration waited after hearing speech until responding
    - *Mutual Silence* (MS): Perceived duration before hearing response to speech



Conversation over Channel with Delays

- Effects of delays on conversational dynamics
  - *Conversational Interactivity* (**CI**)
    - Ratio of perceived silence durations before and after person's speech.
    - Asymmetric with increased MED
  - *Conversational Efficiency* (**CE**)
    - Ratio of conversation duration over channel with delays vs. face-to-face setting
    - Decreasing with increased MED



Conversational Voice Communication Quality

- Quality of an interactive conversation
  - Listening-only speech quality (LOSQ) of one-way speech
  - Perceived degradations due to delays in communication channel
    - MED cannot be perceived directly
    - CI and CE can be perceived
- CVCQ can be represented by (LOSQ, CI, CE)
- Trade-offs between (LOSQ, CI, CE) depend on MED
  - MED ↑: LOSQ ↑, CI ↓, CE ↓
  - MED ↓: LOSQ ↓, CI ↑, CE ↑

Perceived Delay Effects

- Perceived effects of MED depend on **conversational conditions**
  - CI depends on the **Human Response Delay**
  - CE depends on **Switching Frequency** of conversation

Table 2: Statistics of two face-to-face conversations.

Conversation Type	Avg. single-talk duration	Avg. HRD duration	# of switches	Total Time
Social	3,737 ms.	729 ms.	7	35 sec.
Business	1,670 ms.	552 ms.	15	35 sec.

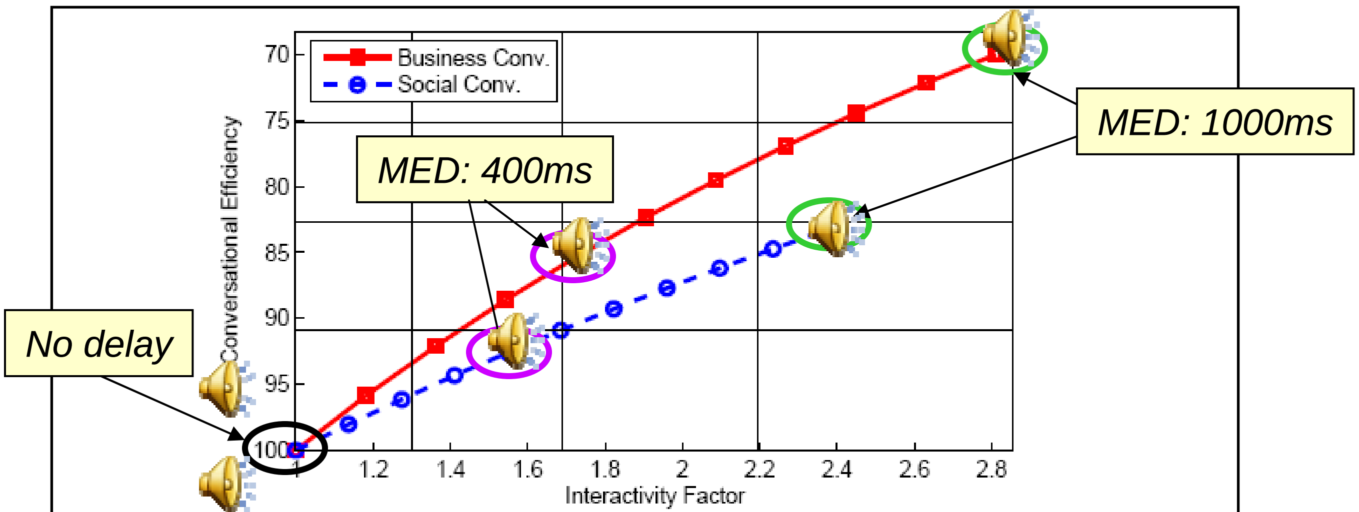


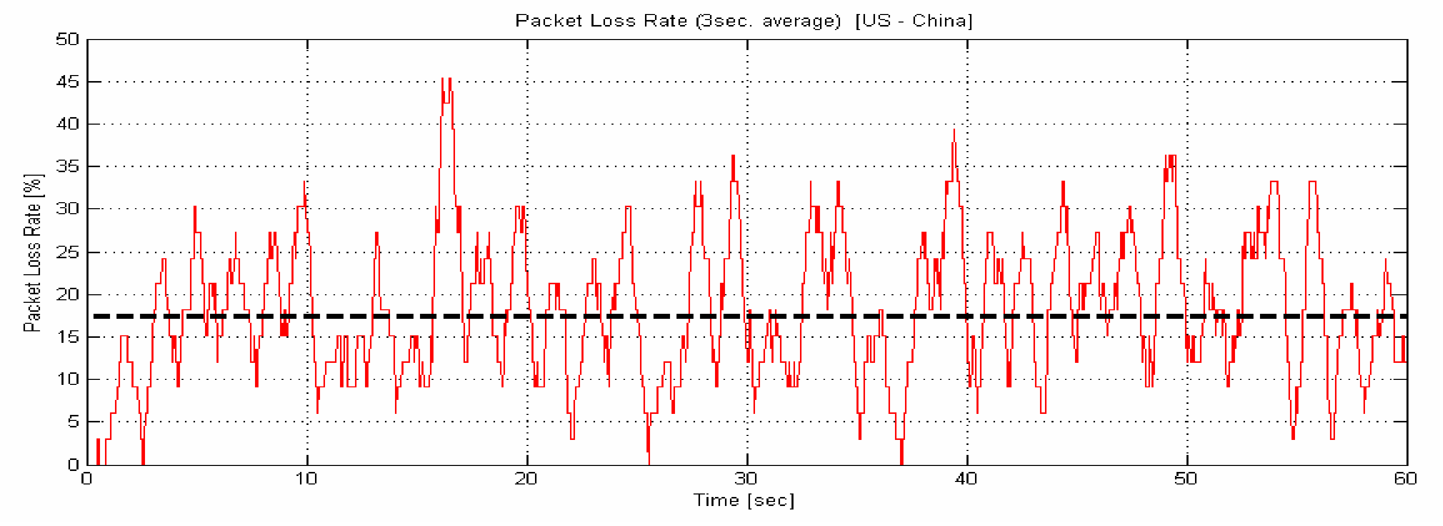
Figure 11: Effect on CI and CE when MED changes for the two conversations in Table 2.

Outline

- Introduction
- Network environment & network control
  - Network conditions
  - Network control: POS and LC
  - Trade-offs in system-controllable metrics
- Trade-offs in CVCQ attributes
- Previous work & our contributions
- Subjective tests
- Design of adaptive POS & LC schemes
- Experimental results

Network Conditions: Packet Loss

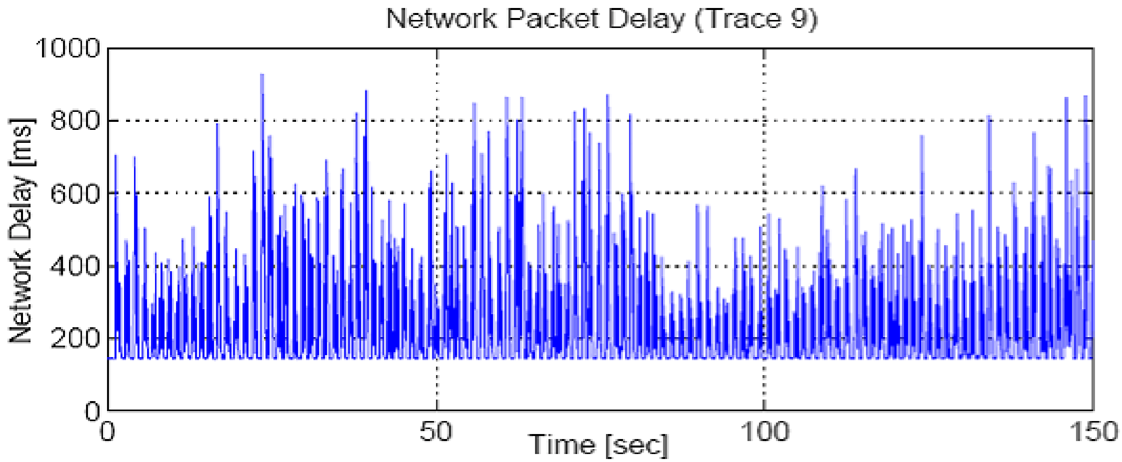
- Network-loss conditions change in a matter of seconds
  - Stationary models cannot track fast changing conditions



- Retransmission of lost speech packets not feasible in real-time VoIP
- Redundancy needed to conceal lost packets at receiver
  - Piggybacking previously sent frame(s) in current packet
  - Require receiver to wait for redundant packet

Network Conditions: Packet Delay

- Packets experience network delays
  - Delays exhibit jitters and high spikes



- IP packets arrive irregularly but speech needs to be played smoothly
  - Employ jitter-buffers and adaptive play-out scheduling (POS)
  - Additional delays

Network Control via POS/LC

- Goal of Network Control: Mitigate network imperfections
- **System-observables**
  - Network-loss rate & burstiness
  - Network delays & jitters
- **System controls**
  - Redundancy rate (degree of piggybacking)
  - Play-out schedule of speech segments
- **Intermediate quality metrics (system-controllables)**
  - Un-concealable Frame Rate (UCFR)
  - Un-concealable Frame Pattern (UCFP)
  - Mouth-to-ear delay (MED)

Trade-offs in System-Controllable Metrics

- Trade-offs between UCFR and MED
  - Depending on network conditions
  - Must be adaptive

<table>
<tr><td colspan="2" rowspan="2">Network Control used under conditions</td><td colspan="2">Network Delay Condition</td></tr>
<tr><td>Low Jitter</td><td>High Jitter</td></tr>
<tr><td rowspan="2">Network Loss Condition</td><td>Low Loss</td><td>No-redundancy
Short & slow changing MED</td><td>No-redundancy
UCFR improves gracefully with MED</td></tr>
<tr><td>High Loss (Bursty)</td><td>Redundant Piggybacking
MED to allow receipt of redundant packets</td><td>Redundant Piggybacking
High MED to reduce UCFR</td></tr>
</table>

Outline

- Introduction
- Network environment & network control
- Trade-offs in CVCQ attributes
  - Trade-offs via system controllable metrics
  - CVCQ representation
- Previous work & our contributions
- Subjective tests
- Design of adaptive POS & LC schemes
- Experimental results

Trade-offs by System Controllable Metrics

- Trade-offs between
  - CVCQ attributes (LOSQ, CI, CE)
  - System-controllable (UCFR,MED)
- LOSQ(UCFP)
  - Codec's intrinsic quality
  - Codec's robustness to losses
- CI(MED) and CE(MED)
- Optimizing CVCQ
  - By controlling UCFR/MED

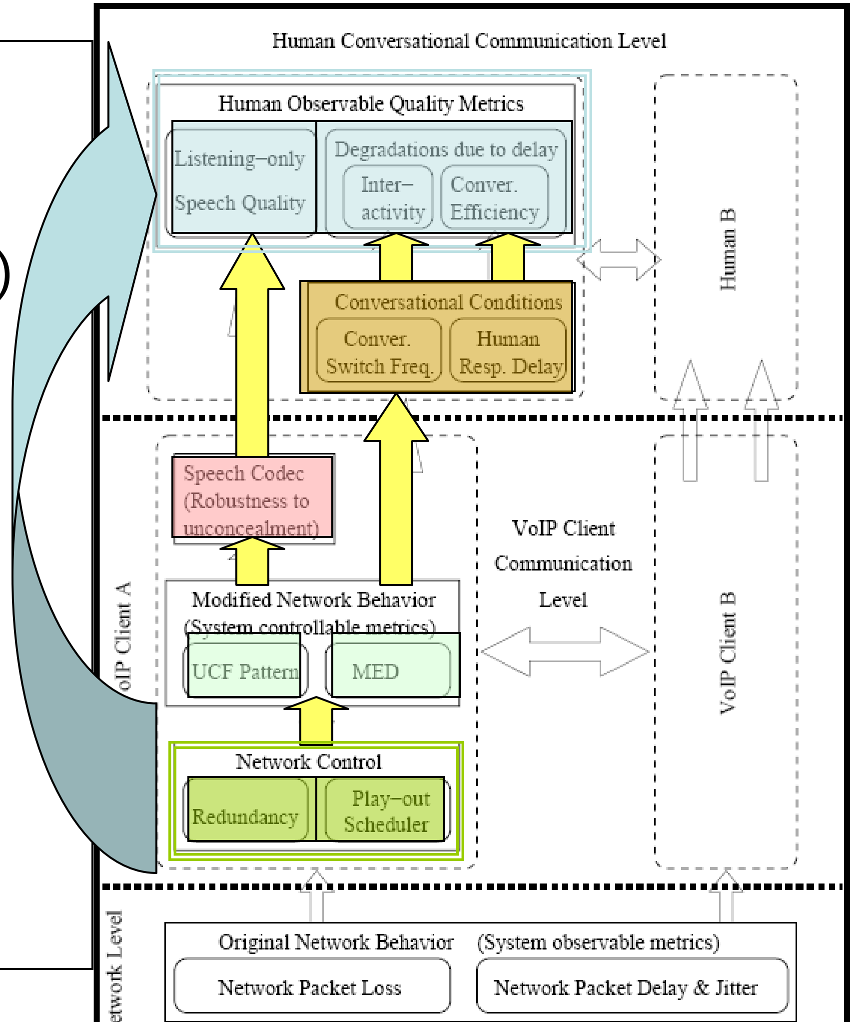
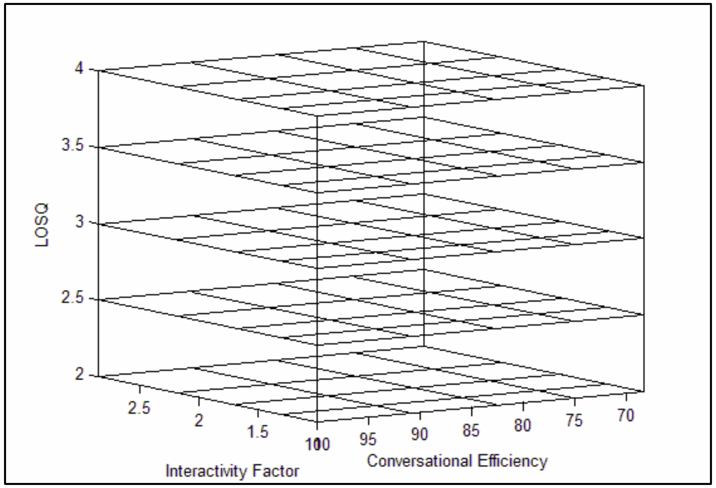


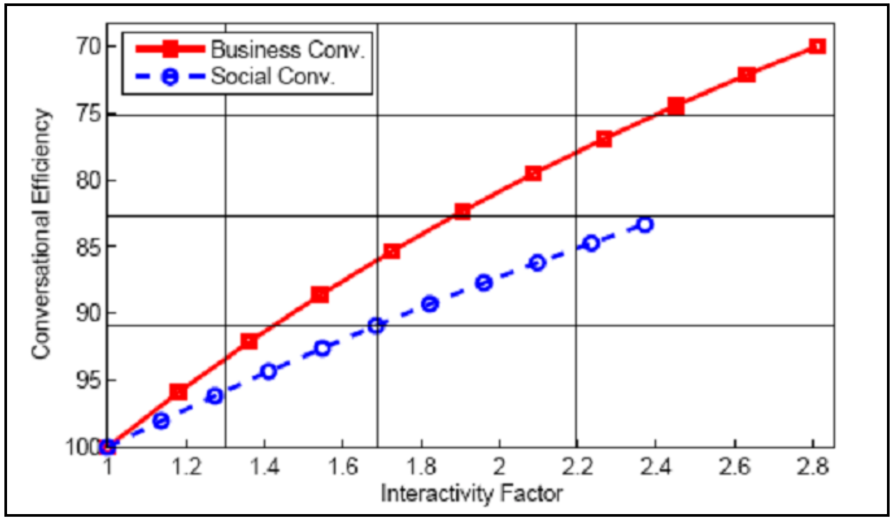
Figure 1: Architecture of a VoIP system.

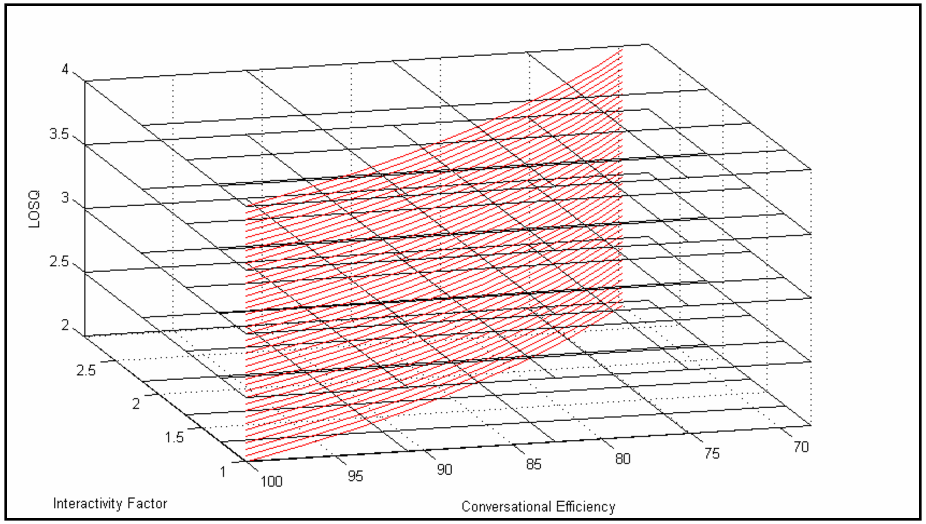
CVCQ Representation

- CVCQ = (LOSQ,CI,CE)
  - Point in 3-D space



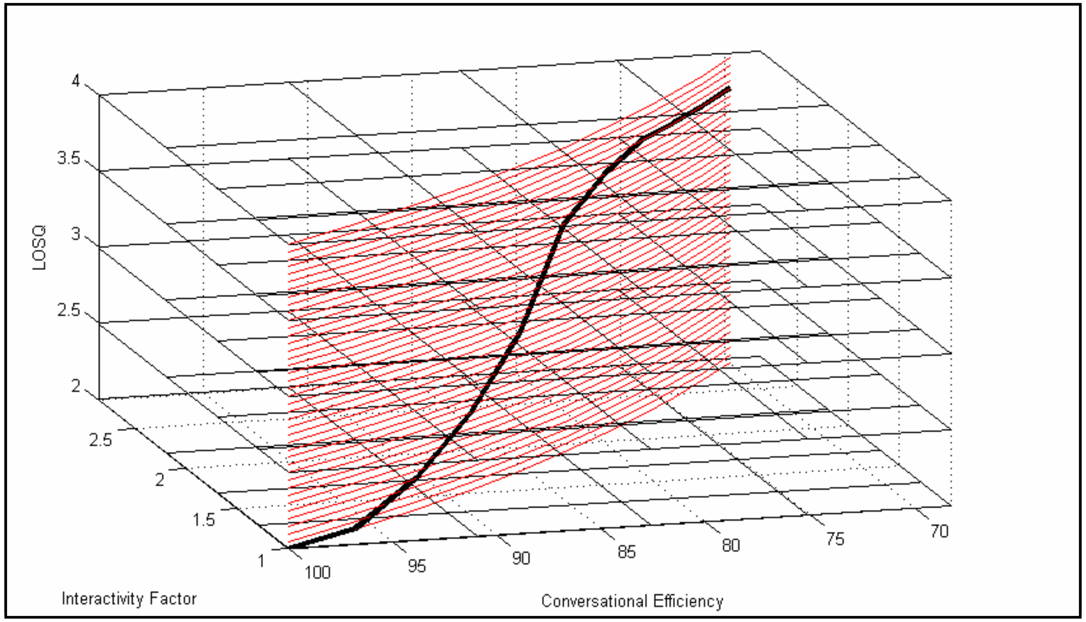
- CI and CE depend on MED and conversational conditions
  - Given conversational condition
    - Restricted to curve (e.g. C_business, C_social) on (CI,CE) plane
    - Restricted to plane (e.g. P_business) on (LOSQ,CI,CE) space





Trade-offs in CVCQ Attributes

- LOSQ depends on MED, redundancy, codec, and network conditions
  - Given codec and network conditions
    - Restricted to a curve on the P_business plane in (LOSQ,CI,CE) space



- Different planes for different conversational conditions
- Different curves for different network conditions

Outline

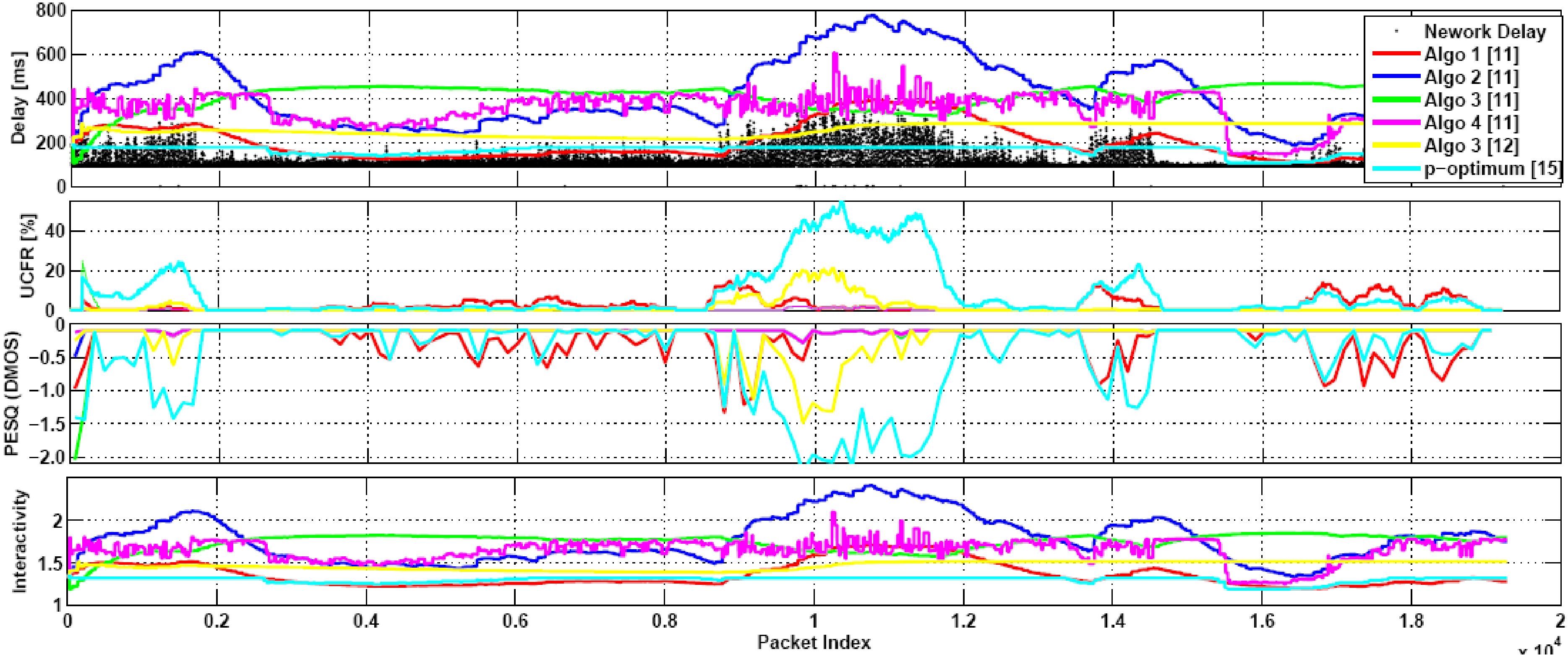
- Introduction
- Network environment & network control
- Trade-offs in CVCQ attributes
- Previous work & our contributions
- Subjective tests
- Design of adaptive POS & LC schemes
- Experimental results

Previous Work: Adaptive POS Schemes

- Open-loop schemes
  - Heuristics to adapt MED, no explicit control, no explicit target
    - PRO: Simple
    - CON: Not robust against all conditions
- Closed-loop schemes with intermediate metrics
  - Adapt MED to control intermediate metrics towards target
    - PRO: Guide control towards well-defined direction
    - CON: Hard to relate to good end-to-end metrics
- Closed-loop schemes with end-to-end metrics
  - Adapt MED to control user-observable metrics → target
    - PRO: Guide control towards target user-observable metrics
    - CON: Hard to estimate at run-time

Previous Work: Adaptive POS Schemes

- None of previous schemes provides consistent balance between CVCQ attributes under changing network conditions



Contributions of Paper

- Develop network-control POS & LC schemes
  - Optimize trade-offs among LOSQ, CI, and CE observable metrics
    - Trading system-controllable metrics
    - Utilizing knowledge on codec performance, conversational and network conditions
  - Deliver high and consistent CVCQ

Outline

- Introduction
- Network environment & network control
- Trade-offs in CVCQ attributes
- Previous work & our contributions
- Subjective tests
- Design of adaptive POS & LC schemes
- Experimental results

Subjective Tests: Comparative MOS

- Comparing perceived quality of two conversations
  - Subjects asked to compare A against B
- Illustration of user preference in 2-D
  - Direction of arrow represents preference
  - Length of arrow represents strength of preference

Table 3: Comparison MOS tests: User responses.

User response	CMOS score
A is strongly preferred against B	−2
A is preferred against B	−1
A and B are preferred equally	0
B is preferred against A	1
B is strongly preferred against A	2

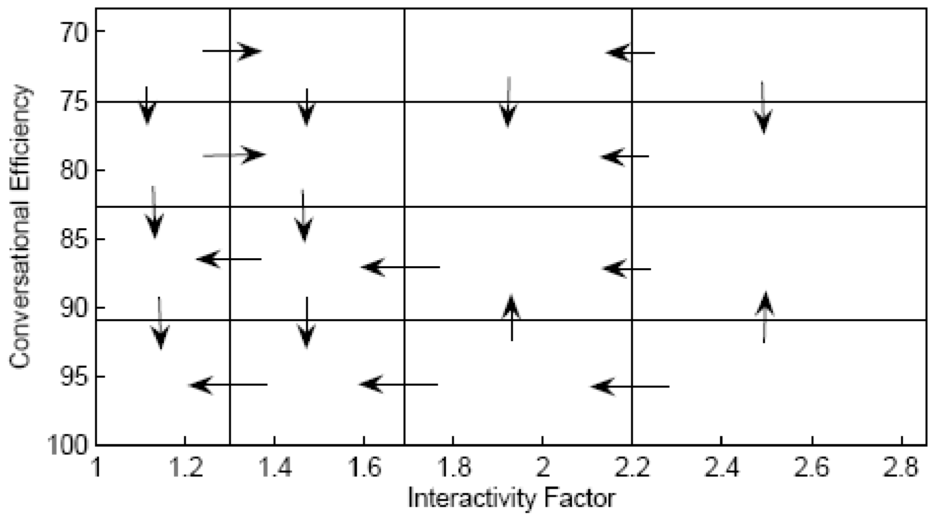


Figure 13: Relative user opinion on the CI-CE plane. The size of an arrow indicates the degree of preference of one alternative over another.

Outline

- Introduction
- Network environment & network control
- Trade-offs in CVCQ attributes
- Previous work & our contributions
- Subjective tests
- Design of adaptive POS & LC schemes
- Experimental results

Proposed POS/LC Schemes

- Loss concealment
  - Control redundancy degree

$$R_{i+FBD} = \min\{R \mid UCFR_i^W(\bar{p}, \bar{R}) \leq 2\%\}.$$

- POS
  - Estimate CVCQ curve by conversational and network conditions
  - Adjust system-controllable metrics to maximize user preference along curve

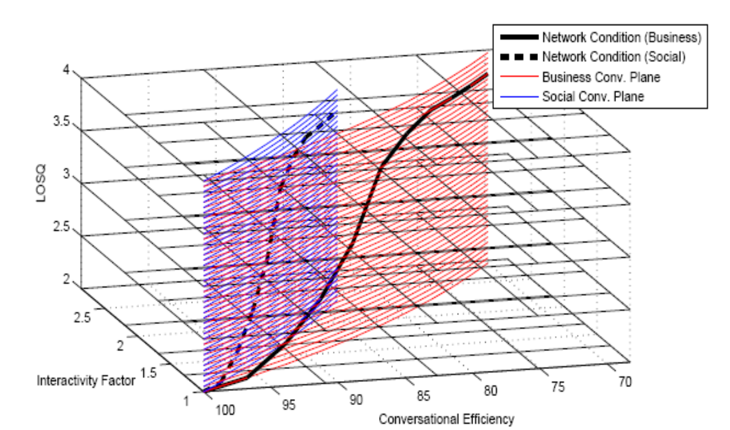


Figure 12: The planes represent the conditions due to the conversational type. The curve on each plane represents the conditions imposed by the network.

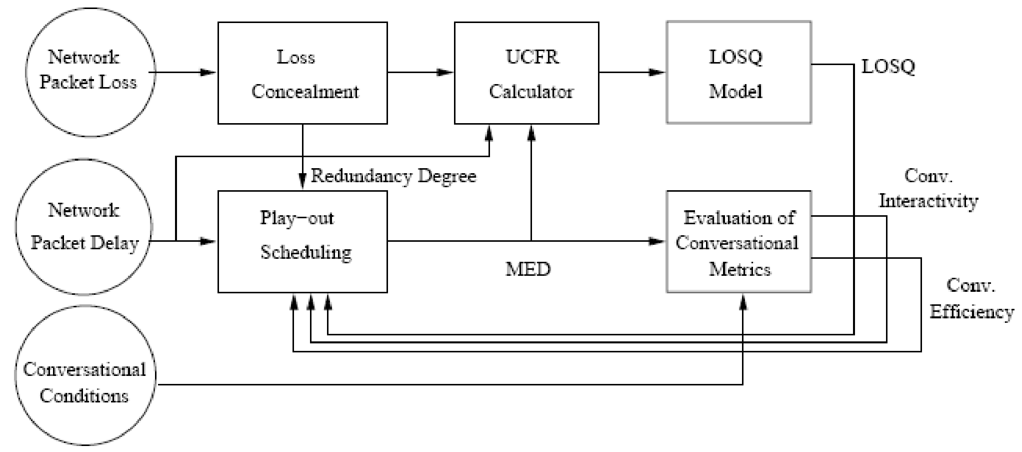
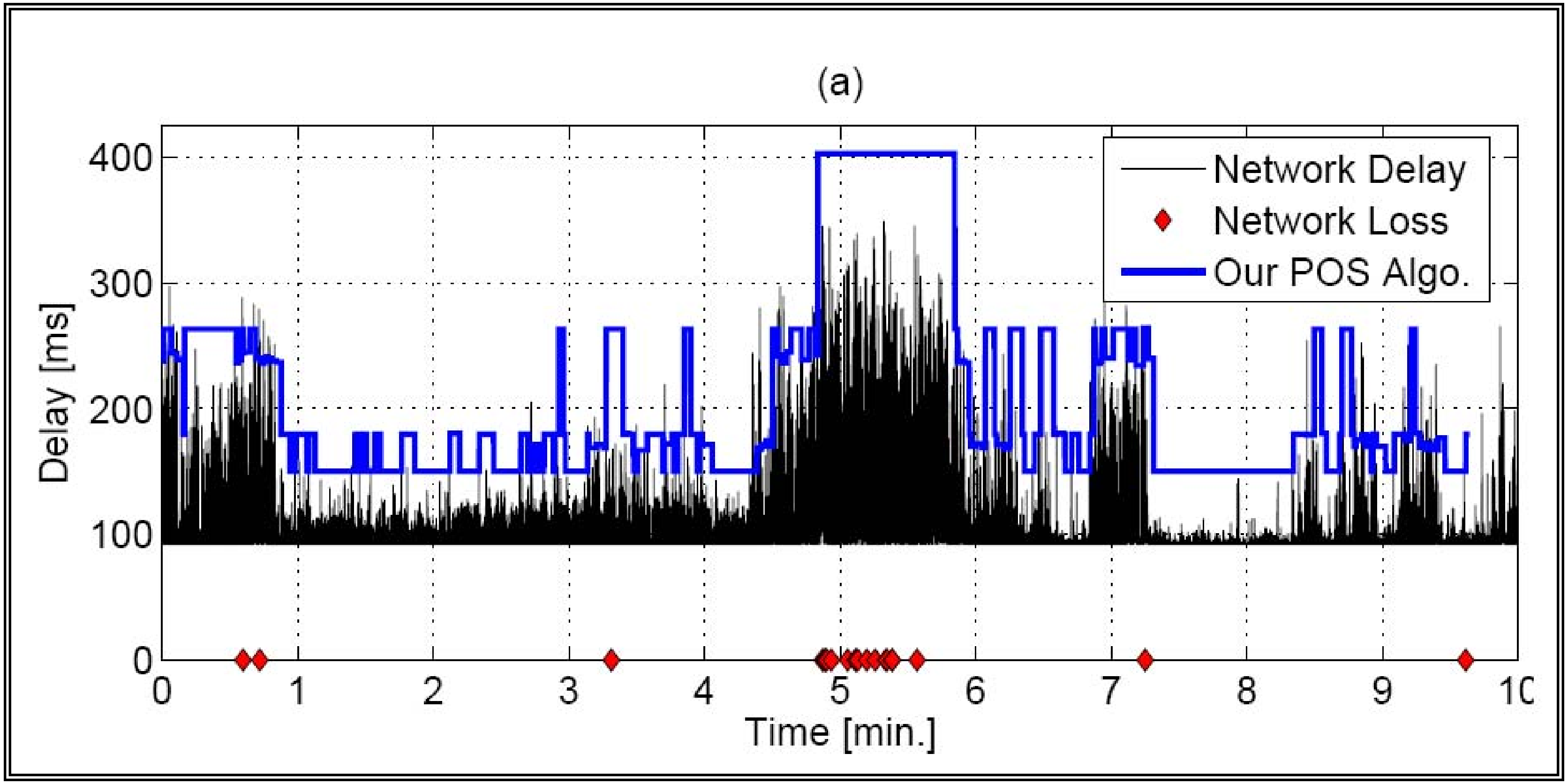


Figure 14: Proposed network control by POS/LC.

Outline

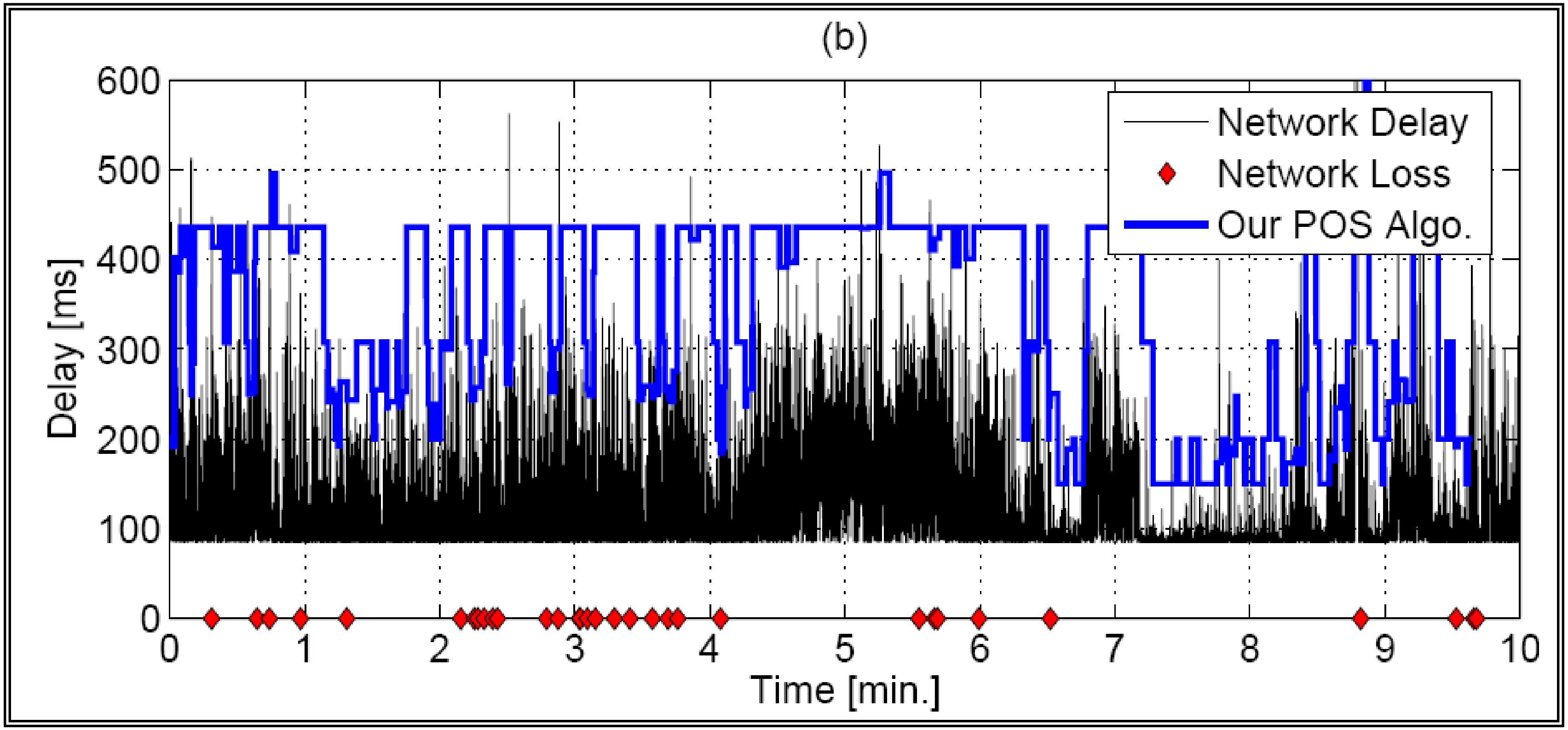
- Introduction
- Network environment & network control
- Trade-offs in CVCQ attributes
- Previous work & our contributions
- Subjective tests
- Design of adaptive POS & LC schemes
- Experimental results

US-Switzerland



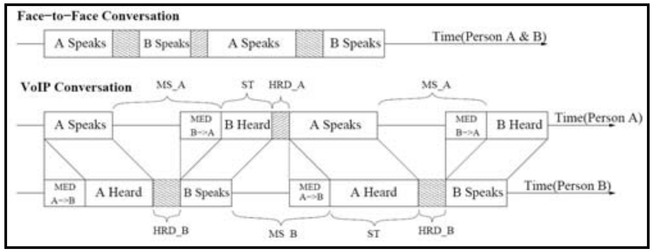
Our: P-optimum:

US-China



Our: P-optimum:

Conclusions



- Conversational quality
  - Listening only speech quality
  - Conversational interactivity and efficiency
  - Subjective tests and just-noticeable difference
  - Optimize via intermediate quality metrics

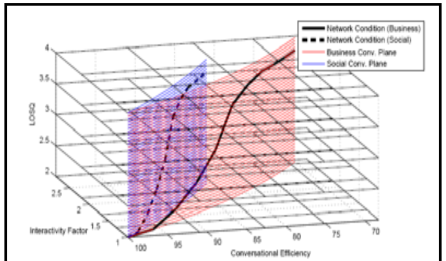


Figure 12: The planes represent the conditions due to the conversational type. The curve on each plane represents the conditions imposed by the network.

- Trade-offs achieved via network controls
  - Loss concealments via redundant piggybacking
  - Suitable mouth-to-ear delays via play-out scheduling