Loss Recovery and Adaptive Playout Control for Packet Voice Communications over IP

Yi Liang

Department of Electrical Engineering
Stanford University

April 19, 2000
Outline

- Background and Motivation
- Network Delay and Jitter Characteristics
- Voice Concealment Using Time Scale Modification and Wave Substitution
- Adaptive Playout Delay Adjustment
- Future Work and More Loss Reconstruction/Concealment Techniques
Voice Over IP

- Increased network bandwidth
- Universal presence of IP, easier and faster Internet access
- Availability of supporting hardware
Voice Over IP - Challenges

- The Internet protocol delivers service on a best-effort basis, with no guarantee of quality of service.
- The transmission mechanism is unreliable and connectionless.
- Transport layer – TCP guarantees reliable transmission by retransmission protocols; but not acceptable for live audio due to end-to-end latency.
Improving QoS – Efforts and Implementations

<table>
<thead>
<tr>
<th>Problems</th>
<th>Efforts</th>
<th>Implementations</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-to-End Delay</td>
<td>Minimize Delay</td>
<td>Optimize Algorithms</td>
</tr>
<tr>
<td>Delay Jitter</td>
<td>Absorb Jitter</td>
<td>Buffer Management</td>
</tr>
<tr>
<td>Packet Loss</td>
<td>Reconstruct/Conceal Loss</td>
<td>Many Loss Recovery Techniques</td>
</tr>
</tbody>
</table>
Delay Jitter

Local Host: cass.stanford.edu
Packet Size=200 bytes   Packet Interval=20 ms

Remote Host: bperp.stanford.edu
Loss Rate= 0.00%

Remote Host: mit.mit.edu
Loss Rate= 0.02%

Remote Host: tom.inria.fr
Loss Rate= 5.99%

Remote Host: public.cs.hr.cn
Loss Rate=11.38%
Jitter Absorption and Buffer Management

Sender

Packet sequence number:
k
K+1
K+2
K+3
K+4
K+5
time

Receiver

Playout

\[ d_a : \text{buffer absorption time or allowed max delay} \]

Packet dropped

k
K+1
K+2
K+3
K+4
K+5
Delay and Loss Rate

\[ d: \text{delay, a non negative random variable} \]

The p.d.f. of \( d: f(d) \)

\[ d_a: \text{Absorption Time} \]

Loss Rate: \( k = \int_{d_a}^{\infty} f(d) \, dd \)

The delay distribution varies as time - The need to modify playout speed
Time Scale Voice Modification

a) Error Concealment

- Conceal Lost Packet

b) Playback Speed Modification

- Modify Playback Speed
Waveform Similarity Overlap Add (WSOLA) Algorithm

input signal $v[n]$

output signal $q[n]$

[Verhelst, Roelands]
Search for Similar Waveform by Pattern Matching

The overlapped output is weighted by Hanning window

\[ q[n] = \sum_{k} w[n - y_k] \cdot v[n - y_k + x_k] \]

The cross-correlation between two segments to be overlapped

\[ c_i = \sum_{j=0}^{\Delta y - 1} v[i + j] \cdot v[x_{k-1} + \Delta y + j] \]

Postion of the search region

\[ \Delta x \propto \frac{1}{l_{out} / l_{in}} \]
To conceal one lost packet:

Introduced delay = 2 packet time

[Stenger, Younes, Reng, Girod]
Delay-optimized Time Scale Modification

a) Conventional Method
- Delay = 2 packet time
- Packet length extension = 50%

b) Improved Method
- Delay = 1 packet time
- Packet length extension = 1/3

What if double-packet loss?
Waveform Substitution Using Pattern Matching

The best match is found by maximum cross-correlation

\[
c[n] = \frac{\sum_{m=1}^{M} x[m] \cdot y[n+m]}{\sum_{m=1}^{M} (y[n+m])^2}, \quad n = 1, 2, \ldots, N
\]

Can only conceal a limited number of lost packet using prior information

Advantage: Delay=0

[Goodman, Wasem]
Conceal Consecutive Lost Packets Using Hybrid Method

- Delay is reduced to one packet time; voice quality is good even for double-packet loss
- Can conceal more consecutive lost packets; but voice quality degrades as the number of consecutive lost packets goes up
Adaptive Playout Delay Adjustment

[Ramjee, Moon]

Average Playout Delay = 5.47ms
Late Loss Rate = 10.82%

Average Playout Delay = 5.76ms
Late Loss Rate = 5.63%
Less Bursty Loss
Audio Demos

I. How Well WSOLA-Based Concealment Works?
- Voice Stream Suffered from Simulated Random Packet Loss
- Concealed Voice Using Delay-optimized Time-scale Modification

\[ \text{Introduced Delay} = 1 \text{ packet time} \]

<table>
<thead>
<tr>
<th>Original</th>
<th>Loss</th>
<th>Concealed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1: 20% Loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2: 30% Loss</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

II. Playback Speed Modification

<table>
<thead>
<tr>
<th>Original</th>
<th>Scaled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

III. Adaptive Playout Delay Adjustment

<table>
<thead>
<tr>
<th>Original</th>
<th>Loss</th>
<th>Playout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Conclusions

<table>
<thead>
<tr>
<th>Concealment Algorithm</th>
<th>Output Voice Quality</th>
<th>Introduced Delay (# of packets)</th>
<th>Capability of Concealing Burst Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waveform Substitution</td>
<td>Fair</td>
<td>0</td>
<td>Very Poor</td>
</tr>
<tr>
<td>Time-scale Modification</td>
<td>Good</td>
<td>2</td>
<td>At the Cost of Increased Delay</td>
</tr>
<tr>
<td>Delay-optimized Time-scale Modification</td>
<td>Good</td>
<td>1</td>
<td>Fair</td>
</tr>
</tbody>
</table>
Future Work

- Audio Signal Processing
- Architecture and Protocols
- Loss Avoidance at Network Layer
- Protect Real-time Voice from Data, Reserving the Bandwidth (Token Bus, RSVP, DiffServ)

VoIP

- Receiver-based Loss Recovery and Concealment
- Scalable (layered) Encoding Scheme
- Error Protection and Data Reconstruction
- Joint Source and Channel Coding
- Loss-resilient Real-time Data Transmission
Combating Loss

Receiver-based Recovery
- Insertion
  - Silence Substitution
  - Packet Repetition
- Interpolation
  - Waveform Substitution
  - Pitch Replication
- Regeneration
  - Time-scale Modification

Sender-based Recovery
- Active
  - Retransmission ARQ
- Passive
  - Interleaving
  - Forward Error Correction
  - Media Independant
  - Media Dependant
Forward Error Correction

### Various Encoding Schemes

<table>
<thead>
<tr>
<th>Standard</th>
<th>Codec</th>
<th>Bit Rate (kb/s)</th>
<th>MOS</th>
<th>Delay (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711</td>
<td>PCM</td>
<td>64</td>
<td>4.3</td>
<td>0.125</td>
</tr>
<tr>
<td>G.726</td>
<td>ADPCM</td>
<td>32</td>
<td>4.0</td>
<td>0.125</td>
</tr>
<tr>
<td>G.728</td>
<td>LD-CELP</td>
<td>16</td>
<td>4.0</td>
<td>0.625</td>
</tr>
<tr>
<td>GSM</td>
<td>RPE_LTP</td>
<td>13</td>
<td>3.7</td>
<td>20</td>
</tr>
<tr>
<td>G.729</td>
<td>CS-ACELP</td>
<td>8</td>
<td>4.0</td>
<td>15</td>
</tr>
<tr>
<td>G.723.1</td>
<td>ACELP</td>
<td>6.3</td>
<td>3.8</td>
<td>37.5</td>
</tr>
<tr>
<td>US Dod FS1015</td>
<td>LPC-10</td>
<td>2.4</td>
<td>2.3</td>
<td>22.5</td>
</tr>
</tbody>
</table>
References

For references and more voice samples:
http://www-ise.stanford.edu/~yiliang/voip.htm