Adaptive Playout Time Control with Time-scale Packet Modification

Yi Liang

July 12, 2000
Outlines

- Three playout schemes
- The algorithm - adaptive playout time calculation according to network delay
- Implementation - time-scale modification of voice packet
- Performance comparison and voice samples
Playout Time Control

1. Fixed Playout Time:
   Constant playout time throughout whole trace. Simply do jitter absorption.

2. Adjust Playout Time between Talk Spurt:
   Adapt playout time to network delay; but playout time is constant in talk spurts.  
   \([\text{Ramjee, Moon}]\)

3. Adjust Playout Time in Talk Spurt:
   Adaptive playout time control + Time-scale modification of voice packet
Adaptive Delay Estimate

- Linear recursive filtering
  \[ d_{\text{est}}_i = \alpha \cdot d_{\text{est}}_{i-1} + (1 - \alpha) \cdot \text{delay}_i \]
  \[ \text{var}_i = \alpha \cdot \text{var}_{i-1} + (1 - \alpha) \cdot |d_{\text{est}}_i - \text{delay}_i| \]
  \(d_{\text{est}}\): network delay of current packet
  \(d_{\text{est}}\): current estimate of delay
  \(\text{var}\): estimate of delay variance
  \(\alpha\): coefficient between 0.5 - 1

- Histogram of past delays
  - Past 30 – 40 delays are stored in a circular array, both the array and the histogram get updated once a new packet is received.
  - Loss rate specified by user.
  - Playout time calculated according to histogram and specified loss rate.
Size of Circular Array

Performance vs. window size

Avg buffering delay (ms)

Late loss rate (%)

Window size for histogram (packet)
Implement - Waveform Similarity Overlap Add (WSOLA) Algorithm

WSOLA Review

input signal $v[n]$

output signal $q[n]$

[Verhelst, Roelands]
WSOLA Working on Long Blocks

[Stenger, Younes, Reng, Girod]
Improved WSOLA on Short Block of Voice

- Working on short block (one packet) of voice in order to avoid additional delay.

- Difference from long block operation
  - Fewer pitch periods in the block (2-8 in a 30 ms packet).
  - Harder to find correlation; need to look into prior packet for stretching.
  - Divide into fewer overlapping segments (two or three); can speed up computation as a result.

- Advantages
  - No delay introduced.
  - Flexible operation, no overlapping needed between packets in concatenation.
Improved WSOLA on Short Block of Voice

**Stretch**

1 2 3 4 5 6

Divide into Overlapping Segments

0 1 2

1 2 3

1 2 3

0 1 2 3 4 5 6

Realign Segments and Add

0 1 2 3 4 5 6

Output Packet

**Shrink**

1 2 3 4 5 6

Divide into Overlapping Segments

1 2 3

2 3 4

1 2 3

2 3 4 5 6

Output Packet

\[ \frac{1}{2} \quad \frac{2}{3} \quad \frac{3}{4} \quad 5 \quad 6 \]
Time-scale Modification on One Packet
Examples

Stable Pitch Period

Packet Extension
125% packet length

Packet Shrinking
87% packet length
Examples (2)

Transition Period

Packet Extension
134% packet length

Packet Shrinking
83% packet length
Performance Comparison

- Trace: Stanford -> Chicago, One-way
- Trace: Stanford -> Germany, One-way
- Trace: Stanford <-> China, Round-trip
- Trace: Stanford <-> Japan, Round-trip

Adaptive Playout Time Control with Time-scale Packet Modification

Yi Liang
Voice Examples (1)

Adaptive Playout Delay

STD of delays: 4.3 ms
STD of playout time: 3.1 ms

1.wav
2.wav
Voice Examples (2)

Adaptive Playout Delay

- STD of delays: 17.6 ms
- STD of playout time: 10.8 ms

1.wav
2.wav

STD of delays: 17.6 ms
STD of playout time: 10.8 ms
Conclusions and Future Work

- Constant playout time adaptation within talk spurt provides the best performance in terms of loss rate and buffering delay.
- Voice packet can be scaled to 50% - 200% of its original size without hurting sound quality.

What’s next ...
- Study the performance improvement for delay traces with different characteristics.
- Implementation in the T2 processor.