REAL-TIME HIGH-QUALITY AUDIO STREAMING

Juan-Pablo Cáceres
Network Sound and Data
Center for Computer Research in Music and Acoustics (CCRMA)
Stanford University
Perform Music On-Line in Real-Time with the Highest Audio Quality Possible
Goals of High-Quality Audio over Networks

**Maximize Audio Quality** for available networks conditions

**Minimize Latency**

But more important, **Minimize Jitter**

Adjustable **Number of Channels**

**Audio routing** flexibility

**Multiple peers**
Keep Delay **Constant**

Maximize Audio Quality
Why Latency Matters

Saint Lawrence String Quartet (Quintet)

25 ms One-way Delay

Banff Centre, Alberta, Canada
Quartet

Stanford, Anechoic Room
Viola
Why Latency Matters

Saint Lawrence String Quartet (Quintet)

25 ms One-way Delay

‣ What Happens Naturally with Delay?
‣ Now, they are conscious...
‣ ...and they try to be stable
‣ The whole Quintet
Some Historical/Technical Foundations

Basic Principle: *Uncompressed Audio*

**Year 2000**

**McGill**  Xu & Cooperstock

**Stanford**  Chafe, Wilson, Leistikow, Chisholm, Scavone
UDP and TCP revisited

**UDP**
- Unreliable
- Connections-less
- Datagram Oriented

**TCP**
- Reliable
- Connected (virtual circuit)
- Byte-Stream Oriented

Delay Constant?
Why TCP is problematic

Event At Sender Site | Network Messages | Event At Receiver Site
--- | --- | ---
Send Packet 1 | Receive Packet 1 | Receive ACK 1
Send Packet 2 | Receive Packet 2 | Receive ACK 2
Send Packet 3 | Receive Packet 3 | Receive ACK 3
Receive ACK 1 | Send ACK 1 | 
Receive ACK 2 | Send ACK 2 | 
Receive ACK 3 | Send ACK 3 | 

TCP supports byte streams flowing in both directions. However, a connection may sometimes show data flowing in only one direction. Remember that, in general, a single TCP connection is established, and then at a later time for the same pair of ports to be involved in a second connection. Note that because TCP connections come and go, it is possible for a connection be closed, and then at a later time for the same pair of ports to be involved in a second connection.

TCP is a byte-oriented protocol, which means that the sender writes bytes into a TCP connection and the receiver reads bytes out of the TCP connection. Although “byte stream” describes the service TCP offers to application processes, TCP does not, itself, create a stream of bytes. Instead, TCP on the source host buffers a sufficient number of bytes from the sending process to fill a reasonably sized packet, and then sends the packet to its peer on the destination host. TCP on the destination host then empties the packet into a receive buffer, and the receiving process reads from the receive buffer at its leisure. This situation is illustrated in Figure 5.3, which, for simplicity, shows how TCP manages a byte stream.
TCP makes the underlying delay elastic and ever-increasing

Use UDP
A JackTrip Session
What is Jack?

Demo: Jack
Real-Time Audio: Under the Hood

Audio

UDP3  UDP2  UDP1

UDP1  UDP2  UDP3

Audio

Juan-Pablo Cáceres | CCRMA | Stanford University | jcaceres@ccrma.stanford.edu
Real-Time Audio: Under the Hood

Audio Process Callback
JackAudio Threads

Non-blocking
Process plug-ins
Non-blocking
Ring Buffer
Non-blocking
Ring Buffer

Audio Inputs (N-Channels)
Audio Outputs (N-Channels)

UDP Sender Thread
UDP Receiver Thread

To Peer
From Peer
Network
Network
Real-Time Audio: Under the Hood

UDP1  UDP2  UDP3
Audio In

Audio Out

UDP Sender Thread
Audio Process Callback
Non-blocking
Blocking
Non-blocking
UDP Receiver Thread
Audio-Outputs (N-Channels)
Audio-Inputs (N-Channels)

Ring Buffer initial "symmetric" pointer positions

write position
read position

empty packets
full packets

Round Robin

Reset pointers and clear older packets
Under-run Condition
Loop through the last read packet
Silent Mode
Wavetable Mode
No new slots to read from
Ring Buffer initial
full packets
empty packets

Juan-Pablo Cáceres | CCRMA | Stanford University | jcaceres@ccrma.stanford.edu
Real-Time Audio: Under the Hood

Sender end

UDP1


UDP2


UDP3


time

Receiving end

UDP1


Lost Packet

UDP3


time

Redundancy

Juan-Pablo Cáceres | CCRMA | Stanford University | jcaceres@ccrma.stanford.edu
Parameters that matter

Queue Length (latency/jitter tradeoff)

Redundancy (bandwidth)

Audio Bit Resolution (bandwidth)

8 / 16 / 24 / 32
Parameters that matter

Packet size (latency)
Sampling Rate (bandwidth)
Packet Size / Latency

The smaller the packet size, the lower the latency (sampling rate constant)
Sampling Rate / Latency

For the same packet size

64 samples → 48kHz: 64/48000 = 1.3 ms
64 samples → 96kHz: 64/96000 = 0.7 ms

The higher the Sampling Rate, the lower the delay (packet size constant)
A Simple (default) JackTrip Session

at CCRMA (Server)

```
jacktrip -s
```

server mode

at UK (Client)

```
jacktrip -c [ccrma-IP-number]
```

client mode
More Control over a JackTrip Session

at CCRMA (Server)

```
jacktrip -s -n 8 -q 4 -r 2
```

eraser  
mode  

Queue  

num chans  

redundancy

at UK (Client)

```
jacktrip -c [ccrma-IP-number] -n 8 -q 4 -r 2
```

client  
mode
A JackTrip Session (demo)
QoS Network “Audible Distances”

Plucking the Net

Longer Distance = Lower Pitch

Short Distance = Higher Pitch

Short Jittery/Lossy Wireless

JackTrip Server

client

client

client
Net vs. Net Collective

Chris Chafe | Chopper
JamLink

http://www.musicianlink.com/
More Information

http://ccrma.stanford.edu/groups/soundwire/

JackTrip at Google Code:
http://code.google.com/p/jacktrip/

JackTrip Mailing List:
http://groups.google.com/group/jacktrip-users