Networked Music Performance (NMP)

Xiaoyuan Gu, Matthias Dick, Ulf Noyer and Lars Wolf
Institute of Operating Systems & Computer Networks
Technical University Braunschweig
Agenda

- Why Networked Music Performance
- Related work
- What is NMP
- Design considerations
- Evaluation
- Conclusions & future work
Motivation

- IT has penetrated into nearly every aspect of the work and life of human beings
- The market of networked entertainment is growing
- The usage of Internet as music databases has been well established and exploited
- Emerging interests in exploring the nature of Internet for new paradigms of networked music
- Emerging applications: networked collaborative composition, networked conducting, and distributed musical performance.

Our focus!
Limitations of Tradition

- Requires physical presence of the musicians
- Not an easy task to find a common timeslot
- Time and costs on traveling
- Find a player of the desired level
- Different versions of the sheet music

→ A basic need to improve the way of music performance for sakes of flexibility, economy, efficiency, productivity and creativity.
Definition

- A concept of rehearsals/concerts via networks with acceptable audio quality
- **Bandwidth-demanding**: Mono PCM 0.7Mb/s, and up to 27.6 Mbps for high definition multi-channel natural audio
- **Highly-delay sensitive**: 120ms E2E delay upper-bound for real-time interactive apps, 20ms desired for music for professionals.
- **Strict requirement on audio stream synchronization**: clocks of PCs, latencies from sound device, NIC, and rhythm adjustment etc.
Only a few studies on this topic present in literature

Earlier work was focused on mono audio transport over ATM Network

Some experiments on using MIDI to convey synthetic audio

The Master Class approaches

Xu and Copperstock’s work

The SoundWire Project

The Conductor-driven Scheme

**Conclusion:** limitations of current work and the demand for further work
Design: The Architecture

- 4 major components: client, server, compression & communication
- Application boundaries

Two application scenarios:
- real-time rehearsal & rehearsal on-demand
- Targeted at home Internet users constrained by the last-mile bottleneck link

Two application scenarios: real-time rehearsal & rehearsal on-demand

Two application scenarios: real-time rehearsal & rehearsal on-demand
Application Scenarios

Local Musician 1

Access Router

Remote Musician 2

Access Router

Remote Musician 3

Access Router

Remote Musician 4

Access Router

Server + Virtual Remote Musicians
The Client

- The user interface to NMP
- Service configuration: instrument selection and tuning, music piece determination, partner selection, rhythm control, starting-point signaling etc.
- Sound Card Manipulation: duplex is a must, allow controlling the related latency due to buffering
- Clock Synchronization: all client clocks are in sync with that of the server using NTP
The Server

NMP Server

Session Management
- User Repository Maintenance
- Performance Coordination
- Clock Synchronization

Audio Stream Manipulation
- Audio Synchronization
- Audio Mixing
- Audio Transcoding

Audio Repository Management
- Storage
- Recording
- Packet Composition

A point of centralized control

Session Management: user repository maintenance, performance coordination, clock synchronization

Audio Stream Manipulation: audio stream synchronization, audio mixing, transcoding, packet composition

Audio Repository Management: storage of the performance examples for either emulation of the remote musicians or playback of the live performance
The Communication

- A packet-switched paradigm
- Hybrid delivery mechanism: *Unicast clients’ mono audio to server and multicast the multi-channel audio from server back to clients*
- Audio Data Transport: *RTP over UDP, standard compatible*
- Session Control: *Proprietary session management protocol over TCP. Support for session management, performance coordination. Message object serialization.*
The Compression

- A trade-off between bandwidth efficiency and latency
- MP3 and MPEG-4 AAC not for real-time due to buffering
- ADPCM is the selected codec for real-time rehearsal scenario
- The usefulness of NMP is decided by E2E delay
- Buffer size has a significant impact on delay

Compression as Library:
*Flexibility in choosing the optimal codec, configure the desired algorithm, quality, and buffer size.*
The Implementation

- **Test-bed:**
  *PCs in a controlled LAN*

- **Operating Systems**
  *Linux: OSS + ALSA*

- **Programming Language**
  *C++*

- **Graphical User Interface**
  *QT*

  *Envision of cross-platform portability*

- **Hardware**
  *PCs, Fast Ethernet switches, multi-channel sound cards and speaker systems, oscilloscope and sweep generator.*
Evaluation: Test Procedure

- Targeted codecs: MP3 and ADPCM
- Evaluation categories: object + subjective
- Measurement metrics: PSNR and MOS
- Peak Signal to Noise Ratio: levels 50dB, 70dB, 90dB
- Mean Opinion Score: 1-5
- Test sequence: publicly available audio samples
Test Configurations

(a) Test configuration A - distortion due to compression

(b) Test configuration B - end-to-end processing distortion

Networked Music Performance © Xiaoyuan Gu, Matthias Dick, Ulf Noyer and Lars Wolf

Nov. 29, 2004, NIME’04 at Globecom’04, Dallas, USA
Test Configurations (cont.)

(a) Test configuration C - single client latency

(b) Test configuration D - end-to-end latency
## Audio Quality

<table>
<thead>
<tr>
<th>Instrument</th>
<th>ADPCM PSNR</th>
<th>LAME MP3 PSNR</th>
<th>ADPCM MOS</th>
<th>LAME MP3 MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>bass</td>
<td>0.00</td>
<td>0.50</td>
<td>1.00</td>
<td>1.50</td>
</tr>
<tr>
<td>400</td>
<td>1.00</td>
<td>2.00</td>
<td>2.50</td>
<td>3.00</td>
</tr>
<tr>
<td>1k</td>
<td>2.00</td>
<td>3.00</td>
<td>3.50</td>
<td>4.00</td>
</tr>
<tr>
<td>harp</td>
<td>3.00</td>
<td>4.00</td>
<td>4.50</td>
<td>5.00</td>
</tr>
<tr>
<td>horn</td>
<td>4.00</td>
<td>5.00</td>
<td>5.50</td>
<td>6.00</td>
</tr>
<tr>
<td>quar</td>
<td>5.00</td>
<td>6.00</td>
<td>6.50</td>
<td>7.00</td>
</tr>
<tr>
<td>trpt</td>
<td>6.00</td>
<td>7.00</td>
<td>7.50</td>
<td>8.00</td>
</tr>
<tr>
<td>vioo</td>
<td>7.00</td>
<td>8.00</td>
<td>8.50</td>
<td>9.00</td>
</tr>
<tr>
<td>castanets</td>
<td>8.00</td>
<td>9.00</td>
<td>9.50</td>
<td>10.00</td>
</tr>
<tr>
<td>youcandothat</td>
<td>9.00</td>
<td>10.00</td>
<td>10.50</td>
<td>11.00</td>
</tr>
</tbody>
</table>

PSNR (dB) vs. MOS
**Latency Distribution**

**Breakdown of E2E Delay (ms)**

- Server processing delay, 10.5, 57%
- Client audio capturing and playback delay, 2.67, 15%
- Client audio coding/decoding, packetization and other overhead, 5.17, 28%
Conclusions

- The proposed application suffices the real-time constrains and the required audio quality in the LAN.
- Different audio compression schemes and multi-channel audio were supported.
- There exists loose-couplings between MOS and PSNR

Future Work

- To extend the application toward larger scale networks
- To add the support for MPEG-4 AAC
- To consider realistic network conditions
- End-system adaptation schemes and QoS support
- To adopt better object measurement metric like PEAQ