Enhancement of Dynamic Local Lag Control for Networked Musical Performance

Makuhari Messe, Tokyo, Japan    IEEE GCCE 2014

Mya Sithu, Yutaka Ishibashi, and Norishige Fukushima
Graduate School of Engineering
Nagoya Institute of Technology
Nagoya, Japan

Oct. 10, 2014
Outline

- Background
- Previous work
- Purpose
- Networked haptic drum system
- Assessment system and methods
- Assessment results
- Conclusions and future work
Networked musical performance has become one of the most demanding real-time applications.

Synchronization quality of sound and interactivity may seriously be deteriorated.
To achieve high synchronization quality of sound, Irie et al. used the local lag control for a networked ensemble.*1

- The local lag control*1,2 buffers the local information for a constant time (called the local lag).
  The interactivity is degraded.

- They set the local lag to the same value as the network delay from the local terminal to the other terminal.
  The interactivity may seriously be degraded when the network delay is large.

They assumed that the network delay from the local terminal to the other terminal is equal to that in the opposite direction (called the *symmetric delay case*).

- Usually, in a network, the network delay from the local terminal to the other terminal is different from that in the opposite direction (called the *asymmetric delay case*).

High synchronization quality of sound may not be achieved.
We subjectively investigated the effect of the local lag control on the synchronization quality of sound, interactivity, and comprehensive quality in the joint musical performance. *3

- There exists the optimum value of local lag for joint musical performance.
- The optimum value of local lag is dependent on the network delay from the other terminal to the local terminal.

We proposed the dynamic local lag control, which dynamically changes the local lag according to the network delay from the other terminal to the local terminal.\(^4\)

The dynamic local lag control sets the local lag to the optimum local lag value according to the network delay.

\[ \Delta = 0.637 \cdot D + 6.578 \]

**\( D \):** The MU delay, which is defined as the time interval from the moment a media unit (MU) is generated at the other terminal until the instant the MU is output at the local terminal.

The dynamic local lag control can achieve higher synchronization quality of sound and keep interactivity higher than the local lag control.

Purpose

Problem
The dynamic local lag control can be used for only two users.*4

It is important to be able to employ the control for three or more users so as to perform music like orchestra music.

This work
- We enhance the dynamic local lag control so that three or more users can play musical instruments together through a network.
- We make a comparison among the dynamic local lag control, the adaptive $\Delta$-causality control*5, and no control by QoE (Quality of Experience) assessment subjectively and objectively.
- We use a networked haptic drum system*4 for networked musical performance in the assessment.

Networked haptic drum system

3D virtual space

User 1’s drumstick
High-hat cymbals
Snare drum

Terminal 1

Headset
PC 1
Haptic interface device
PC 2
Display
Switching hub

User 1

Terminal 2

User 2

Network

Terminal 3

User 3

User 2’s drumstick
User 3’s drumstick
Floor tom
Bass drum

User 1’s drumstick
High-hat cymbals
Snare drum

User 2’s drumstick
User 3’s drumstick
Floor tom
Bass drum

Networked haptic drum system
Enhanced dynamic local lag control

- $\Delta_{ij}$: The MU delay from terminal $i$ to terminal $j$
  - ($i, j = 1, 2, \text{or } 3$)
- $\Delta_{ii}$: The local lag at terminal $i$

Terminal 1

- $\Delta_{11} = \max(\Delta_{21}, \Delta_{31})$
- To synchronize the output time between received MUs and local information, the additional buffering time is added to $\Delta_{j1}$ by $\Delta_{j1} - \Delta_{11}$ ($j = 2$ or $3$).
- Note that the additional buffering time of the larger one between $\Delta_{21}$ and $\Delta_{31}$ is zero.
The network emulator generates an additional constant delay $D_{ij} (i, j = 1, 2$ or $3)$ for each packet transmitted between the terminals.

For simplicity, we use only PC 1 at each terminal, and each subject employs only the right drumstick.

For two terminals, we used both right and left drumsticks or only right drumsticks, and we obtained the similar results.

$D_{ij} = D_{ji}$ ($D_{ij}$ and $D_{ji}$ are a pair of constant delays between terminals i and j)

$\Delta_{ij} = D_{ij} + \alpha$ ($\alpha =$ buffering time + processing time)
Assessment methods (1/5)

- We compare QoE among the following three types of control:
  - The enhanced dynamic local lag control
  - The adaptive $\Delta$-causality control
  - No control

**The adaptive $\Delta$-causality control**

- The local lag at each terminal is set to the maximum MU delay among the terminals.
- The synchronization quality is perfect, but the interactivity is the poorest when the network delay is large.

**No control**

- The control outputs information on receiving it.
- The interactivity is the highest, but the synchronization quality is most seriously damaged when the network delay is large.
Assessment methods (2/5)

Case 1 (Three pairs of constant delays are different from each other.)

- $D_{12} = D_{21} = 50$ ms and $D_{23} = D_{32} = 0$ ms
- $D_{13}$ and $D_{31}$ are simultaneously changed from 0 ms to 150 ms at intervals of 50 ms.

Case 2 (A pair of constant delays is different from the other pairs.)

- $D_{12} = D_{21} = D_{23} = D_{32} = 0$ ms
- $D_{13}$ and $D_{31}$ are changed from 0 ms to 150 ms at intervals of 50 ms at the same time.

Case 3 (Three pairs of constant delays are the same.)

- All the constant delays are simultaneously changed from 0 ms to 150 ms at intervals of 50 ms.

The constant delays, cases 1, 2, and 3, the three types of control are selected in random order for each trio of subjects.
Assessment methods (3/5)

Each subject hits the snare drum repeatedly by the right hand.

- High-hat cymbals
- Snare drum
- Floor Tom

He/she needs to move his/her right drum stick between the snare drum and the floor tom.

- Each trio of subjects hit the snare drum with rhythm 1 at 60 beats per minute (bpm).
- A different rhythm and tempo (i.e., rhythm 2 and 100 bpm) were not handled here because results of the other combinations of rhythm and tempo were almost the same as those of rhythm 1 at 60 bpm*1.

Subjective QoE Assessment

Each subject was asked to base his/her judgment on the following QoE:

- Synchronization quality of sound
- Interactivity
- Comprehensive quality (weighted sum of above two qualities)

Five-grade quality scale

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Excellent</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
</tr>
<tr>
<td>2</td>
<td>Poor</td>
</tr>
<tr>
<td>1</td>
<td>Bad</td>
</tr>
</tbody>
</table>

Number of subjects: 15
Age: Between 20 and 28
Gender: Male and female

We obtain Mean Opinion Score (MOS).

Each stimulus: 30 seconds
Total assessment time: One and half hours
Objective QoE Assessment

- Objective assessment was also carried out at the same time as the subjective assessment.
- We adopted the root mean square error of sound at terminals as an objective assessment measure.

**Root mean square error of sound**: Square root of the mean square error which denotes the average of squared differences between the output times of three sounds (i.e., between the output times of sound at terminals 1 and 2, between those at terminals 1 and 3, and between those at terminals 2 and 3).

- The local lag ($\Delta$) was also employed as an objective assessment measure since it is closely related to the interactivity.
Subjective assessment results (1/4)

Case 1 at terminal 1

Synchronization quality of sound

MOS of synchronization quality of sound

- Enhanced dynamic local lag control
- Adaptive Δ-causality control
- No control

95% confidence interval

$D_{13}$ and $D_{31}$ (ms)
Subjective assessment results (2/4)

Case 1 at terminal 1

Interactivity

MOS of interactivity

Enhanced dynamic local lag control

Adaptive \( \Delta \)-causality control

No control

\( D_{13} \) and \( D_{31} \) (ms)
Subjective assessment results (3/4)

Case 1 at terminal 1

Comprehensive quality

MOS of comprehensive quality

Enhanced dynamic local lag control

Adaptive Δ-causality control

No control

I 95% confidence interval

$D_{13}$ and $D_{31}$ (ms)
Subjective assessment results (4/4)

Case 1 at terminal 2

Comprehensive quality

Enhanced dynamic local lag control

Adaptive $\Delta$-causality control

No control

MOS of comprehensive quality vs. $D_{13}$ and $D_{31}$ (ms)
Conclusions

- We enhanced the dynamic local lag control so that three or more users can play musical instruments together over a network.
- We investigated the effect of the control by subjective and objective QoE assessments in a networked haptic drum system for three terminals.
- We made a comparison among the enhanced dynamic local lag control, adaptive $\Delta$-causality control, and no control.
- The enhanced dynamic local lag control is the most effective.
- We can estimate the MOS values from the root mean square of sound and/or local lag with high accuracy.
Future Work

- We will investigate the influence of packet loss by QoE assessment.
- It is also important to examine the effect of the enhanced dynamic local lag control in real environments where multiple users play the same or different kinds of real musical instruments.