

# ローカル・ラグ制御を伴うマルチメディア伝送サーバの構築

李 榮宰<sup>1</sup> 片桐 滋<sup>1</sup> 大崎 美穂<sup>1</sup>

**概要:** コンピュータに支援される協調的作業の実現に対する要請が高まる中で、高い品質をもつマルチメディア情報提示技術の開発が関心を集めている。そのような情報提示を実現するため、特に、大型液晶ディスプレイやビデオカメラ、ラウドスピーカ、マイクロホンなどの多数のマルチメディア機器を部屋型の空間に配置する“t-Room”のために、音声データと映像データを同期的に伝送することができるマルチメディアデータ伝送サーバを開発する。本サーバ中の音声伝送モジュールは、知覚レベルの遅延の知覚を低減することを目指すローカル・ラグ制御機能を持ち、映像伝送モジュールは、複数のディスプレイ間の映像データの同期機能を持つ。その開発においては、映像伝送モジュールが、その厳密に定義されている計算手順の存在に基づいて、マルチメディア同期における基準として用いられている。また、音声伝送サーバにおけるリングバッファに関わるデータ入出力を可変長の複数データブロック単位とすることで、その音声伝送の動作安定性を高めている。体系的な評価実験の結果、開発したマルチメディアデータサーバは、640 × 480 ピクセルから 1920 × 1280 ピクセルにわたる広範囲の映像データサイズと 512 バイトなどの種々の音声データブロックサイズにおいて、安定的に動作することが確認される。

**キーワード:** 同期的協調作業, ローカル・ラグ法, マルチメディアデータ伝送システム

YEONGJAE LEE<sup>1</sup> SHIGERU KATAGIRI<sup>1</sup> MIHO OHSAKI<sup>1</sup>

**Abstract:** Along with increasing demand for computer-supported cooperative work, the development of high-quality multi-media information displays has attracted vigorous research interests. To realize such information display, "t-Room" has been proposed that locates many multi-media devices, such as large size liquid crystal displays, video cameras, loudspeakers, and microphones in a room type layout. In this paper, we newly develop a multi-media data transmission server for t-Room by integrating the existing voice and video servers. The server consists of a voice data transmission module that has local-lag control function and a video data transmission module that allows the synchronized reproduction of multiple video data. In our development, a video data transmission module works as a standard in multi-media synchronization because the reproduction of video data is a rigid but stable computational mechanism. Additionally, we elaborate our new server's voice data transmission module, which is functionally equal to the existing voice server and improve its stability by introducing a mechanism that transfers multiple data blocks to or from the inside ring buffers. Finally, from systematic evaluation experiments, we show that our developed multi-media server stably runs with a wide range of video data sizes from 640 x 480 pixels to 1920 x 1280 pixels and several voice data block sizes, such as 512 bytes.

**Keywords:** Remote cooperative work, Local lag control, Multimedia data transmission

## 1. はじめに

近年のコンピュータの目覚ましい進歩等に基づき、マルチメディアデータをリアルタイム的に扱うことを可能とするメディアデータ伝送・処理システムが普及しつつある。そうした中で、遠隔地間をつないで協調作業を行う場を提

供し、さらにそこでの利用者にとっての同室感あるいは高臨場感を高めることで協調作業を支援する遠隔コラボレーションシステムへの関心が高まっている。遠隔地間でリアルタイム性を損なわず同期的作業を行う際、そこにはデジタルネットワークを経由するために生じる不可避な通信遅延が存在し、それはしばしば、同期的作業の大きな妨げとなる。このような問題の解決を目指し、筆者の一部は、自然な協調作業にとって欠かすことができない音声メデ

<sup>1</sup> 同志社大学大学院 理工学研究科  
Graduate School of Science and Engineering, Doshisha University

アと映像メディアに着目し、それらの知覚レベルにおける遅延あるいは同期のずれの減少を目指すラグ制御機能を持つ音声伝送サーバと映像伝送サーバの開発を行ってきた [5], [6], [7], [8], [9] .

音声伝送サーバに関する研究は、まず、映像データを共有する同期的協調作業をローカル・ラグの導入によって支援する考え方を参考に、音知覚における同期のずれの軽減を目指すローカル・ラグ制御機能を持つ音声伝送サーバの開発から始められた [5] . しかし、そのサーバは遅延量制御の精度が十分でなく、また 2 地点間接続のみに対応しているなど、機能は必ずしも満足できるものではなかった . そこで、その制約を改善することを目指し、それまでと全く異なる (サーバ内部におけるデータ処理のためのバッファリング等の機構) アーキテクチャを持ち、かつ多地点接続に対応できる新しい音声伝送サーバが開発された [6] . しかし、評価実験を通してその新しいサーバの基本的性能や有用性は示されたものの、遅延等の測定時間が短く、かつサーバ内部の計算処理に伴う遅延量もまだ小さくなく、さらなる改善が求められていた [6], [7] . こうした状況を受け、音声伝送サーバはさらに改善され、今では、音データの入出力に伴う高い精度のブロッキング制御が可能な OS 上で、処理の優先制御機能も持つサーバとなっている [8] .

一方、映像伝送サーバは、上述の新しいアーキテクチャを持つ音声伝送サーバを基盤として開発が行われた [9] . ただし、音声伝送サーバと異なり、自地点において発生される映像データにラグを付加して表示する必然性は低く、このサーバでは、複数の遠隔地点から送られてくる映像データを同時に表示する際の映像間の同期をとることにラグ制御機能はふりむけられた [9] .

上述したように、音声伝送サーバは映像伝送サーバにも展開され、協調的作業を行う際の主たるメディアであるものと思われる視聴覚メディアのデータを、ラグ制御機能を伴って伝送・処理をする手段が整いつつある . しかし、それぞれの伝送サーバは、別々のプロセスとして開発されたものであり、同時に使用する際には、両サーバの処理開始時刻合わせや、各メディア間の同期を取ることは困難である .

こうした状況を受け、本稿では、これまでの音声伝送サーバと映像伝送サーバを一つに統合した、新しいマルチメディア伝送サーバの構築を行い、その性能評価を行う . なお、この統合に際しては、まず、新サーバ内にメディア間の同期機構が必要かどうかを検討し、仮に必要である場合には同期機構を実装、そしてさらに、その性能評価実験を行っていく . 視聴覚メディア間の同期機構の実装に関しては、音声メディアと映像メディアのどちらを基軸にするかが検討されるべきである . これは、評価実験において安定的なデータ出力動作を行えるメディアデータを統合の基軸とするようにする .



図 1 遠隔合奏における音声データフロー .

Fig. 1 Voice data flow in remote ensemble .

図 2 遠隔合奏にローカル・ラグ制御を用いた音声データフロー .

Fig. 2 Voice data flow using local lag control .

## 2. ローカル・ラグ制御を伴う音声伝送サーバ

### 2.1 ローカル・ラグ法

図 1 のように、離れた地点 A と B における遠隔合奏を想定する . 地点 A で演奏者が演奏した音データがネットワークに送出される際、地点 A の演奏者自身が聴く音をフィードバック音と呼び、ネットワークを介し、地点 B の演奏者が聴く音をフィードスルー音と呼ぶ . この時、地点 A から送出される音データはネットワークの通信遅延を伴って地点 B に到着する . 地点 B の演奏者はこの遅延量の乗った音データに合わせて演奏する . つまり遅延量のみ演奏自体は遅れてしまう . また地点 B から伝送される音データは、ネットワークの通信遅延を伴って地点 A に到着する . ここで、さらなるネットワーク通信遅延分に合わせるように地点 A の演奏者は演奏を遅らせる . こうした事象が繰り返され、両地点の演奏者が聞く相手の演奏音のずれは次第に増大し、やがて演奏は破たんしてしまう [5] .

そこでローカル・ラグ法を適応し、各地点の演奏者がともに同タイミングで同期的に音を聴くことができるように、ネットワークの通信遅延と同量の遅延 (ラグ量と呼ぶ) をフィードバック音に加える (このラグ量が付加されたフィードバック音をラグ・フィードバック音と呼ぶ) . 図 2 は遠隔合奏にローカル・ラグ法を用いた場合の音声データフローである . 図 2 のように演奏者は自身の演奏音をラグ・フィードバック音としてラグ量分遅れて聞くことになるが、合奏者同士の聴取のレベルにおいては同期の確保が可能となる . 前小節の遠隔合奏の破たんの問題に対し、ローカル・ラグ法の適用は、演奏地点のそれぞれにおいて、演奏者自身の演奏音と遠隔地から届けられる演奏音との間のネットワークの通信遅延に伴うずれを取り除くことで演奏の破たんの防止を目指す [5] .

### 2.2 サーバの概要：構成

ローカル・ラグ制御を伴う音声伝送サーバは、図 3 のように音響サーバと同期制御サーバから構成される . 音響サーバは、マイクロホンから音データを取得する機能、ネットワークを介した音データの送受信機能、音響サーバ間のネットワークの遅延を計測する機能、ラグ量を付加し音データを出力する機能を持つ . 一方、同期制御サーバ



図 3 音声伝送サーバの構成例。

Fig. 3 A configuration example of voice data transmission server .

は、音響サーバの接続要求を受け付け、音響サーバ同士の接続を制御する。音響サーバの詳細については 2.5 で述べる。最小構成として音響サーバは演奏を行う各地点に 1 台ずつ設置する必要があるが、同期制御サーバは接続地点のいずれかに 1 台設置する必要がある。いずれのサーバも LINUX ベースとする OS が搭載された計算機上で動作するように実装している。

### 2.3 サーバの動作

伝送サーバを 2 地点間で接続する場合を想定する。例えばネットワークの通信遅延が 20 ミリ秒である場合、地点 A からネットワークを介して送出された音は地点 B で 20 ミリ秒遅れて出力される。ここにローカル・ラグ法を適用すると、地点 A のフィードバック音に 20 ミリ秒のラグ量が付加され、地点 A と地点 B の演奏者が同時に同じ音を聴取できるようになるために遠隔地同士の合奏が可能となる。遠隔合奏支援のための音声伝送サーバの主な処理の流れは以下の通りである。

- (1) 音響サーバが音データの送受信を行う他の音響サーバへの接続要求を同期制御サーバに出す。
- (2) 同期制御サーバが音響サーバからの接続を受け付け、各音響サーバの情報を通知する。
- (3) 同期制御サーバが音響サーバ同士を接続させる。
- (4) 同期制御サーバが音響サーバ間のラグ量を計測させる命令を出す。
- (5) 各音響サーバがラグ量を計測し、同期制御サーバに通知する。
- (6) 同期制御サーバが最大のラグ量を計算し、各音響サーバに付加するラグ量を通知する。
- (7) 音響サーバ同士が音データの送受信を開始する。
- (8) 音データを受信した各音響サーバはラグ量を付加して音データを出力する。

### 2.4 サーバの実装

本稿の 2.3 で述べた通り、音響サーバは、音データの入出力や送受信、ラグ量の制御などの機能を担っている。それらの機能はサーバ内の複数のスレッドによって実現されている。音響サーバを構成している各スレッドの図を図 4 に示し、その役割を述べる。

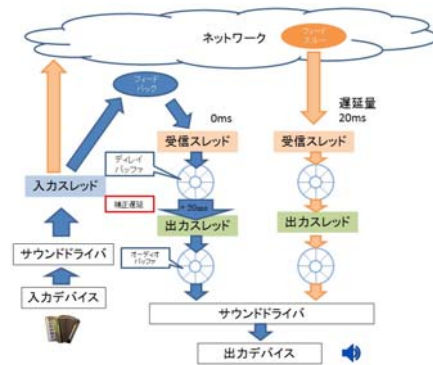


図 4 音響サーバ内のデータフロー。

Fig. 4 Voice data flow in voice server.

- 入力スレッド...入力デバイス (サウンドカード等) からドライバを経由して音データを取得し、接続している全ての音響サーバに音データを送信するスレッド。
- 受信スレッド...送信元が送出した音データを受信し、音響サーバ自身の時間情報を基に受信時刻を記録してリングバッファに格納するスレッド。音データのチャンネル毎に 1 つ存在する。
- 出力スレッド...現在時刻と受信スレッドで記録した受信時刻を比較し、その比較を基に受信スレッドのリングバッファから音データのみを取り出し、オーディオ出力用バッファに転送するスレッド。音データを取り出すタイミングを調整することで、ローカル・ラグ法のラグ量付加機能を実現する。音データはオーディオ出力用バッファから Portaudio のコールバック関数を使用し、出力デバイスに転送される。出力スレッドは受信スレッドと 1 対 1 の関係にある。

音響サーバでは、音の入出力ライブラリに Portaudio を使用している。Portaudio はクロスプラットフォームに対応した OSS の入出力ライブラリであり、C 言語又は C++ 言語で簡単にプログラムを書くことができる。Portaudio のオーディオ再生処理には同期処理と非同期処理の 2 種類がある。同期処理は出力デバイスでオーディオの再生が完了するまで処理をブロックするので、オーディオ再生に安定性はあるものの処理のアイドル状態が発生し遅れてしまう可能性が高くなる。非同期処理はオーディオの再生完了を待たずして処理が返されるので、リアルタイム性が求められる場合において用いられることが多い。入出力デバイスの準備が整った場合に Portaudio のコールバック関数が適切なタイミングで入出力を行う。本音響サーバが開発された当初、入出力処理は非同期処理で行われていたが、評価実験の結果、入力時に非同期処理を用いると、音のデータサイズを小さくすると音が途切れてしまうことが確認されたため、現在は入力に同期処理が、出力に非同期処理が用いられている。

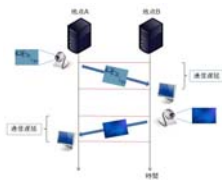


図 5 遅延を伴う 2 地点間の映像伝送.

Fig. 5 Lag-affected video data transmission between two points.



図 6 映像伝送システムの構成例.

Fig. 6 A configuration example of video data transmission server.

### 3. ラグ制御を伴う映像伝送サーバ

#### 3.1 映像再生における同期の崩れへの対処

図 5 のように離れた地点 A と地点 B において同期的コラボレーション作業を行うことを想定する。また、地点 A と地点 B の間で映像データが送受信され、それらの映像が重ねて表示されるものとする。この時、地点 A から送出される映像データはネットワークの通信遅延に伴って地点 B に到着する。地点 B の利用者はネットワークの通信遅延分遅れた地点 A の映像にタイミングを合わせて行動する。同様に地点 B からの映像データにはネットワークの通信遅延を伴って地点 A に到着する。このように時間の経過と共に伝送される映像における遅延は増大し、両地点における利用者間の行動の同期性は失われてしまう。同期の崩れの解決策はネットワークの通信遅延を取り除くことである。しかし、ネットワークの通信遅延は不可避なものであるため、その遅延を取り除くことは困難である。そこで、この映像伝送サーバの開発においては、せめて、異なる地点から送られてくる映像間の同期を保つこと、また自地点（ローカル）映像と他地点映像を重ねて（あるいは並べて）表示する際の映像間の同期を保つことが目指されている。

#### 3.2 サーバの概要：構成

既開発の映像伝送システムは、図 6 のように映像サーバと同期制御サーバから構成されている。映像サーバは、カメラから映像データを取得する機能、ネットワークを介した映像データの送受信機能、映像サーバ間のネットワークの遅延を計測する機能、ラグ量を付加し映像データを出力する機能を持つ。一方、同期制御サーバは、映像サーバの接続要求を受け付け、映像サーバ同士の接続を制御する。映像サーバの詳細については 3.5 で述べる。最小構成としては遠隔合奏支援システムと同様に、映像サーバが各地点に 1 台ずつ設置する必要があるが、同期制御サーバは接続地点のいずれかに 1 台設置する必要がある。いずれのサーバも LINUX ベースとする OS が搭載された計算機上で動作するよう実装されている。

#### 3.3 サーバの動作

映像伝送サーバを 2 地点間で接続する場合を想定する。

例えばネットワークの通信遅延が 20 ミリ秒である場合、地点 A からネットワークを介して送出された映像は地点 B で 20 ミリ秒遅れて出力される。ここにローカル・ラグ法を適用すると、地点 A の映像データに 20 ミリ秒のラグ量が付加され、地点 A と地点 B の利用者が同タイミングでそれぞれの映像を視覚できるようになるために遠隔地同士の同期的協調作業が可能となる。映像伝送システムの主な処理の流れは以下の通りである。

- (1) 映像サーバが映像データの送受信を行う他の映像サーバへの接続要求を同期制御サーバに出す。
- (2) 同期制御サーバが映像サーバからの接続を受け付け、各映像サーバの情報を通知する。
- (3) 同期制御サーバが映像サーバ同士を接続させる。
- (4) 同期制御サーバが映像サーバ間のラグ量を計測させる命令を出す。
- (5) 各映像サーバがラグ量を計測し、同期制御サーバに通知する。
- (6) 同期制御サーバが最大のラグ量を計算し、各映像サーバに付加するラグ量を通知する。
- (7) 映像サーバ同士が映像データの送受信を開始する。
- (8) 映像データを受信した各映像サーバはラグ量を付加して映像データを出力する。

#### 3.4 サーバの実装

本稿の 3.2 で述べた通り、映像サーバは、映像データの入出力や送受信、ラグ量の制御などの機能を担っている。それらの機能はサーバ内の複数のスレッドによって実現されている。映像サーバを構成している各スレッドの図を図 7 に示し、その役割を述べる。

- 入力スレッド...入力デバイス（カメラ等）からドライバを経由して映像データを取得するスレッド。
- 送信スレッド...接続している全ての映像サーバに映像データを送信するスレッド。
- 受信スレッド...送信元が送出した音データを受信し、映像サーバ自身の時間情報を基に受信時刻を記録してリングバッファに格納するスレッド。
- 出力スレッド...現在時刻と受信スレッドで記録した受信時刻を比較し、その比較を基に受信スレッドのリングバッファから映像データのみを取り出し、映像を出

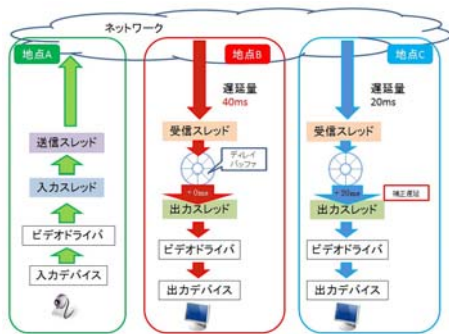


図 7 映像サーバ内映像データフロー.  
Fig. 7 Video data flow in image server.

力するスレッド．映像データを取り出すタイミングを調整することで，ラグ制御機能を実現する．

映像伝送システムでは，コンピュータビジョンライブラリとして OpenCV を採用している．OpenCV はマルチプラットフォームに対応した OSS のライブラリであり，C 言語や C++ 言語などの言語がサポートされており，容易に映像処理が可能である．

### 3.5 ローカル・ラグ制御を伴う音声伝送サーバとの差異

音あるいは音声は，発信者自身も聴取するメディアであるためにローカル・ラグによる知覚レベルの改善を試みる対象であるが，映像は，発信者自身が見るものではなく，本質的にローカル・ラグ制御の対象ではない．しかし，複数の遠隔地間で映像の同期性は保つ必要があるため遠隔地間でタイミング合わせるべくラグの制御が必要となる．t-Room[2]などで標準的に考えているような全身像や部屋全体の映像を自分がフィードバック的に見るシーンは考えにくいですが，例えば手の動作などの身体の一部を遠隔地の人の手などの動きと重ねるといふことが必要となる場合がある．よって，このような場合は音声伝送サーバと同様にローカル・ラグ制御の対象になる可能性がある．

また，映像データは入出力の際，映像ディスプレイにおける動画表現の基本単位となる静止画に対応するフレーム単位で扱われる．したがって映像伝送システムではリングバッファのブロックサイズをフレームサイズと同一とし，システム内での映像データの扱いはすべてフレーム単位で行う．一方，音データはある程度の自由度のある固定長のブロック単位で扱うことができ，映像データのフレームのように一定量のデータをまとめて処理する必要がないため，マルチメディアとして音と映像を扱う場合，メディア間の同期を確保するためには，映像伝送システムと音声伝送システムが互いのデータ処理の違いを考慮した同期制御を行う必要がある．

## 4. メディア間同期を考慮したラグ制御を伴う音声・映像伝送サーバ

### 4.1 マルチメディア化における問題点

音声伝送サーバと映像伝送サーバは，別々に，また継時的に開発されてきた経緯もあり，それらを同時に稼働させる場合であっても，それぞれは別々のプロセスとして動作する．プロセスが別であれば，同期制御サーバのラグ量の制御系が複数必要になるので，実質の同期制御サーバが複数必要となる．また，各システムの実行のタイミングにずれが生じてしまい映像と音の同期が困難なものとなる可能性が高い．よって，こうした既存音声伝送サーバと映像伝送サーバを一つに集約・統合することが望まれる．

しかし，別々のプロセスであれば，それぞれを異なる計算機上で走らせることが可能であるが，一つに統合されたサーバは一台の計算機で走らせることとなり，その計算機における負荷の増大が予想され，結果的に同期的協調作業が破たんしてしまう恐れもある．したがって負荷を軽減させる必要があるのかどうかをまず調査しなければならない．

また，二つの伝送機能の一つに集約するだけでは，そうして作られた伝送サーバは映像データと音データの間同期をとる機構を持っていないため，音と映像の間にずれが生じてしまう可能性がある．人間の知覚の上限は映像メディア先行の場合，約 130 ミリ秒であり，音声メディア先行の場合，約 50 ミリ秒であるとされている．したがって，同期機構を構築するためには，この人間の知覚の上限を超えない範囲にずれを抑え込む必要がある．

さらに，同期機構を構築するには音と映像のどちらを基軸にし，同期をとるべきかという問題もある．基軸にしたい映像メディアが安定的に出力されなければ，システムが破たんしてしまうため，この評価が重要であると考えられる．

### 4.2 サーバの概要：構成

統合に伴う上記の論点を考慮し，本稿で提案する新しいマルチメディア伝送サーバは，音声と映像のメディア間同期を考慮したラグ制御をもつ．この新しい伝送サーバにおけるマルチメディアサーバは，これまでの音声伝送サーバと映像伝送サーバのすべての機能を踏襲している．また，同期制御サーバについても，本稿 2.3 と 3.3 で述べた通りの機能を保持している．最小構成としては，音声伝送サーバと映像伝送サーバと同様にマルチメディアサーバが各地点に 1 台ずつ設置する必要があるが，同期制御サーバは接続地点のいずれかに 1 台設置する必要がある．いずれのサーバも LINUX ベースとする OS が搭載された計算機上で動作するよう実装している．

### 4.3 サーバの動作

新しい伝送サーバを 2 地点間で接続する場合を想定す

る．例えばネットワークの通信遅延が 20 ミリ秒である場合，地点 A からネットワークを介して送出された音・映像は地点 B で 20 ミリ秒遅れて出力される．ここにローカル・ラグ法を適用すると，地点 A の音・映像データに 20 ミリ秒のラグ量が付加され，映像の場合は地点 A と地点 B の利用者が同タイミングでそれぞれの映像を視覚できるようになり，音の場合は地点 A と地点 B の演奏者が同時に同じ音を聴取できるようになるために遠隔地同士の同期的コラボレーション作業が可能となる．本伝送サーバの主な処理の流れは以下の通りである．

- (1) マルチメディアサーバが音・映像データの送受信を行う他のマルチメディアサーバへの接続要求を同期制御サーバに出す．
- (2) 同期制御サーバがマルチメディアサーバからの接続を受け付け，各マルチメディアサーバの情報を通知する．
- (3) 同期制御サーバがマルチメディアサーバ同士を接続させる．
- (4) 同期制御サーバがマルチメディアサーバ間のラグ量を計測させる命令を出す．
- (5) 各マルチメディアサーバがラグ量を計測し，同期制御サーバに通知する．
- (6) 同期制御サーバが最大のラグ量を計算し，各マルチメディアサーバに付加するラグ量を通知する．
- (7) マルチメディアサーバ同士が音・映像データの送受信を開始する．
- (8) 音・映像データを受信した各マルチメディアサーバはラグ量を付加して音・映像データを出力する．

#### 4.4 マルチメディア化に伴う実装の要点

音声伝送システムと映像伝送システムを個別に，別のプロセスで起動・実行した時の CPU の使用率は約 90 %程であり，サーバに高い負荷を与えてしまう．そこで，単一のプロセスで実行されるように実装したところ，CPU 使用率が約 70 %と，約 20 %減少した．次に，同期制御サーバが複数必要となる問題は，マルチメディアサーバが使用する同期制御サーバのラグ量の制御系を一つに集約することで実現した．また，実行の基点を合わせるべく，音声伝送機能と映像伝送機能を実現させる各スレッドの生成から起動までを一つのクラス内(メソッド)に集約した．最後に同期機構であるが，本研究における同期機構は音声伝送機能の安定性を保ち，かつ直接音とフィードバック音のずれが 33 ミリ秒付近であるような機構とする．なぜなら，今回使用するカメラデバイスのフレームレートが最大 30FPS であるために映像 1 フレーム間の間隔が 33 ミリ秒であるためである．基点は映像伝送システム内の出力スレッド内の安定した入出力を確認できたため，映像メディアとした．従って，音声メディアの出力タイミングを映像メディアのタイミングに合わせるように実装した．

## 5. 性能評価実験

### 5.1 実験概要

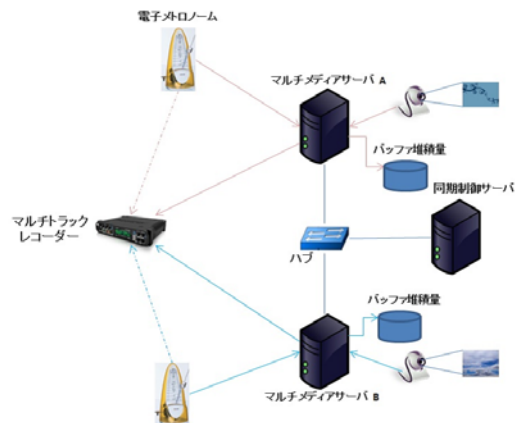


図 8 実験環境.

Fig. 8 Experiment environment.

本実験は音声伝送システムと映像伝送システムを統合した際の同期機構の必要性を検証するための実験である．実験に使用したサーバの使用は表 1 のとおりである．今回の

表 1 マルチメディアサーバの仕様．

機種	Dell Optiplex 9010
カーネル	2.6.32-431.20.3.el6.x86_64
OS	CentOS6.5
CPU	Intel®Core™i-7-3770 Processor 3.40GHz
コア数	4
メモリ	8GB DDR3 SDRAM
カメラデバイス	Logicoool HD Pro Webcam C920(30FPS)
ビデオカード	Intel HD graphics 2500
サウンドドライバ	ALSA
サンプリングレート	44.1kHz
量子化ビット	16bit
チャンネル数	1(モノラル)
サウンドカード	SoundBlaster X-Fi Digital Audio

実験環境は図 8 のように設定した．正確な精度を調査するためマルチメディアサーバと同期制御サーバは同一 LAN 内で接続した．音声メディアについては，120BPM の電子メトロノーム音をマルチトラックレコーダとマルチメディアサーバ(図 8 中のマルチメディアサーバ A, B)に入力，マルチトラックレコーダに直接入力される電子メトロノーム音(直接音)はラグを含まないので，レコーダに録音される直接音とマルチメディアサーバ A, B からの出力であるフィードバック音との時間的なずれを計測することで，マルチメディアサーバ内の計算処理に起因するラグ量を求めることができる．ずれの計測は，レコーダに記録された音声データを波形編集ソフトを用いて行った．なお，映像メ

ディアについては、カメラデバイスの入力能が最大 30FPS であることから 1 フレームあたりの入力時間は約 33 ミリ秒と推測可能であるために、映像メディアの出力スレッド内にてリングバッファの堆積量を調査する。すなわち、音声メディアをこの 33 ミリ秒付近に計算処理を安定して抑え込むことができれば、各メディアは同期的な動作を示すことになる。

本実験において、マルチメディアサーバ内の計算処理に所要する時間は音声メディアの場合、入力するデータサイズに左右され、映像メディアの場合、入力映像の描画サイズに左右される。音声メディアのデータサイズは先行研究で有用であると示した 128, 256, 512[8] と設定し、映像メディアの描画サイズは今回用いたカメラデバイスの最大値である 1920 × 1280 ピクセルと最小値である 640 × 480 ピクセル、さらに先行研究で最適であると示した 1280 × 720 ピクセル [9] とする。音声メディアの解析は 5 分毎に、前述した波形編集ソフトを用いる方法で計測した。なお、本実験の 1 回あたりの試行時間は 30 分とし、試行は各音声データサイズ毎にそれぞれ 3 回ずつ独立に行った。

## 5.2 結果と考察

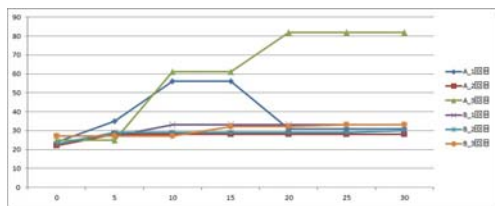


図 9 フィードバック音に伴う誤差 (描画サイズ 1920 × 1280 ピクセル) (以下, 図 14 まで, 描画サイズを除いては同様のため, 省略形で記載)。

Fig. 9 Delay added to feedback sound for "1920 × 1280" video data.

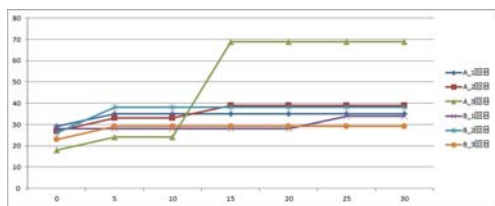


図 10 フィードバック音に伴う誤差 (描画サイズ 1280 × 720)。

Fig. 10 Delay added to feedback sound for "1280 × 720".

映像メディアの描画サイズが 1920 × 1280 ピクセル, 1280 × 720 ピクセル, 640 × 480 ピクセルの 3 パターンと音声メディアのデータサイズが最適なデータサイズ 256[8] の場合の直接音とフィードバック音の解析結果のみを図 9, 10, 11 に示す。図の縦軸は直接音とフィードバック音のずれを示しており, 単位はミリ秒である。また横軸は 5 分ご

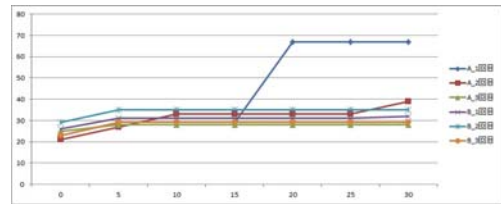


図 11 フィードバック音に伴う誤差 (描画サイズ 640 × 480)。

Fig. 11 Delay added to feedback sound for "640 × 480".

との計測時点である。図の系列はマルチメディアサーバの A, B の何回目の思考であるかを表している。なお, 映像メディアの結果であるが, いずれの場合においても堆積量が無かったため, 図は示さない。データサイズ 128, 256 においては基本的に 20 ミリ秒 ~ 40 ミリ秒に抑えられているものの, 突出したずれが発生している。これは, 各スレッドの処理のタイミングが合わず, 出力スレッド内で出力するデータが確保できず, 安定的な音声データの供給が行えなかった, もしくは Portaudio の出力用のコールバック関数の割り込みが適切に行われず, 呼び出されていないために発生したと考える。このような突出したずれは, 同期的共同作業を行う際にはクリティカルなものになってしまうため, いかにかこのようなずれが起こらないようにするかがキーとなるのは明らかである。安定的な音声データの供給についてはマルチメディアサーバ内の入出力の処理系を改良する必要があり, Portaudio の出力用のコールバック関数の割り込みが適切に行われず, 呼び出されていない場合, マルチメディアサーバのスペックを向上や Portaudio のバージョンアップなどで対処できる可能性がある。データサイズ 512 においては, 100 ミリ秒を超えるずれが発生しており, 当然のごとく, 同期的作業を行うには適していないと考える。データサイズが 512 の場合はフレーム超が 512 / 44100 の約 11.6 ミリ秒であり, 音声 が 1 つ供給されないと, 約 11 ミリ秒のずれが生じてしまう。データサイズ 128, 256 と同様に割り込みなどの処理が行えないと大きなずれが生じてしまうため, 今回のような結果となったと考えられる。

## 6. 性能評価実験

### 6.1 実験概要

本実験は, 前節の音声メディアにおける突出したずれが起こらないようにする, つまり, いかなる条件下においても計算処理に伴う遅延量を 33 ミリ秒付近に安定的に抑え込む機構を導入し, その性能を評価することが目的である。マルチメディアサーバは音声メディアをデータサイズ単位で扱っていることから, データサイズのフレーム長は, 例えばデータサイズが 128 である場合, 128 / 44100 の約 2.9 ミリ秒となる。従って, 安定的に 33 ミリ秒分の音声データを供給するためにはデータサイズが 128 である場合, 約 11 個分のデータを受信スレッドの持つバッファにため込

み,出力スレッドが約11個分のデータをPortaudioに供給することで実現できる.このような機構を今回導入した.サーバの仕様と実験環境,測定・解析方法も5節と同様である.

実験条件についてであるが,第5章と同様に,音声メディアのデータサイズを128,256,512と設定し,映像メディアの描画サイズを1920×1280ピクセル,640×480ピクセル,1280×720ピクセルとした.音声メディアの解析は5分毎に,前述した波形編集ソフトを用いる方法で計測した.なお,本実験の1回あたりの試行時間は30分とし,試行は各音声データサイズ毎にそれぞれ3回ずつ独立に行った.

## 6.2 結果と考察

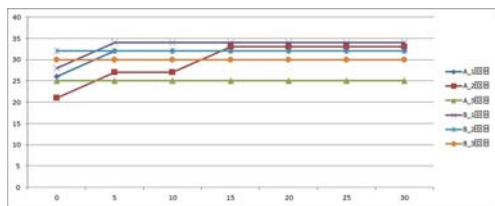


図 12 フィードバック音に伴う誤差 (描画サイズ 1920 × 1280).

Fig. 12 Delay added to feedback sound for "1920 × 1280".

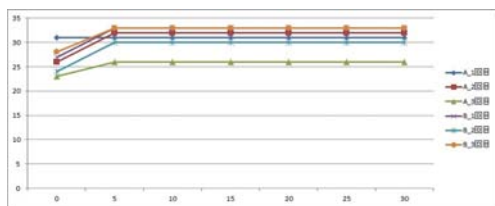


図 13 フィードバック音に伴う誤差 (描画サイズ 1280 × 720).

Fig. 13 Delay added to feedback sound for "1280 × 720".

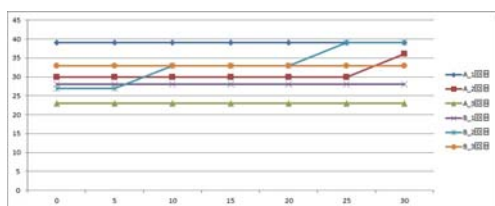


図 14 フィードバック音に伴う誤差 (描画サイズ 640 × 480).

Fig. 14 Delay added to feedback sound for "640 × 480".

映像メディアの描画サイズが1920×1280ピクセル,1280×720ピクセル,640×480ピクセルの3パターンと音声メディアのデータサイズが256の場合の直接音とフィードバック音の解析結果のみを図12,13,14に示す.図の縦軸は直接音とフィードバック音のずれを示しており,単位はミリ秒である.また横軸は5分ごとの計測時点である.図の系列はマルチメディアサーバのA,Bの何回目の

試行であるかを表している.なお,映像メディアの結果であるが,いずれの場合においても堆積量が無かったため,図は示さない.以下の結果と5節の結果を比較すると全ての条件下において突出したずれは発生しなかった.しかし,安定的に音声データを供給しているにもかかわらず,数ミリ秒のずれが発生した.これは,各スレッドの呼び出し割り込みが上手く行えなかった,もしくは5節の考察と同様にPortaudioの出力用のコールバック関数の割り込みが適切に行われず,呼び出されていないために発生したと考える.これらのことからマルチメディアサーバのスペックを向上させることで改善できる可能性が高いと考える.以上のことから今回の同期機構は有用であると考えられる.

## 7. おわりに

本稿では,既存の音声伝送サーバと映像伝送サーバの集約・統合,および音声・映像メディア間同期の実現を目指し,新しいマルチメディア伝送サーバを構築した.性能評価実験の結果,本稿で紹介したメディア間の同期機構はほぼ期待通りに動作し,開発した新サーバは,視聴覚メディアを同時に伝送するため有用な手段足り得ることが明らかになった.今後は,遠隔コラボレーションの支援ツールとして,様々な協動的作業の支援課題において評価を行っていく予定である.

## 参考文献

- [1] Sara A. Bly, Steven R. Harrison, and Susan Irwin: Media spaces: bringing people together in a video, audio, and computing environment, Communications of the ACM, vol. 36, no. 1, pp. 28-46 (1993).
- [2] Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemi Aoyagi, Yoshinari Shirai, Naomi Yamashita, and Junji Yamato; The t-Room: toward the future phone, NTT Technical Review, vol. 4, no. 12, pp. 26-33 (2006).
- [3] 井ノ上直己: 超臨場感コミュニケーション技術の研究開発概要, 情報通信研究機構季報, vol. 56, nos. 1/2, pp. 3-7 (2010).
- [4] Dane Stuckel and Carl Gutwin; The Effects of local lag on Tightly-Coupled Interaction in Distributed Groupware: Computer Supported Cooperative Work, pp.447-456 (2008).
- [5] 入江洋介, 青柳滋己, 高田敏弘, 平田圭二, 梶克彦, 片桐滋, 大崎美穂; t-Room のための遠隔合奏支援システムの構築, 情報処理学会, Vol. 2009-GN-73, No. 3 (2009).
- [6] 竹森幸輝, 前田佳奈, 岩原正典, 片桐滋, 大崎美穂; ローカル・ラグ制御機能とログ同期機能を持つ音響サーバの開発, オーディオビジュアル複合情報処理研究会 (2012).
- [7] 前田佳奈, 竹森幸輝, 岩原正典, 片桐滋, 大崎美穂; ローカル・ラグ機能を持つ音響サーバを用いた遠隔合奏の評価, オーディオビジュアル複合情報処理研究会 (2012).
- [8] 岩原正典; ローカル・ラグ制御機能と同期機能を持つ音響サーバの構築, 同志社大学大学院工学研究科情報工学専攻修士論文 (2013).
- [9] 村上昂, 飯田卓也, 片桐滋, 大崎美穂: 遠隔コラボレーション支援のための映像伝送システムの開発とその遅延評価, 情報処理学会, 第86回グループウェアとネットワークサービス研究会, No. 19 (2013).