

C-07

Mac OS のためのローカル・ラグ制御機能をもつ音声伝送サーバの実装と性能評価

Evaluation and Implementation of Sound Transmission Server that have Local Lag Method for Mac OS

松尾 雄真† 片桐 滋† 大崎 美穂†
Yuma Matsuo Shigeru Katagiri Miho Ohsaki

1. はじめに

コンピュータネットワークの性能向上と普及に伴い、遠隔コラボレーション支援システムの研究開発が盛んに行われている¹⁾。しかし、未だに、視覚メディアや聴覚メディアに関する対称性、すなわち、同室にいる者どうしが同じ方向に映像や音像を知覚する状況、“同室感”を、遠隔地にいるシステム利用者どうしが十分に共有するには至っていない。また、単一メディアにおいても、遠隔地の利用者の知覚には時間的ずれが伴い、それが同期的共同作業の妨げとなっている。

こうした問題に対する 1 つのアプローチとして、大型ディスプレイの左右端に音響反射板で囲まれたスピーカユニットを取り付けた音像生成法が提案され、その性能の検証等が行われている²⁾。しかし、当手法を実装したシステムには、ネットワークを経由して遠隔地間で音声データの伝送を行う機能がなく、遠隔コラボレーション環境下における当手法の評価等を行うには至っていない。

一方、上述の知覚における時間的ずれの軽減を目指し、ローカル・ラグ機能³⁾を持つ音声伝送サーバの開発が進められてきた⁴⁾。このローカル・ラグ機能は、音データの発信元におけるその音データ自体の受聴に遅延(ラグ)を追加し、遠隔地にいる協調作業員間の受聴における時間的ずれの克服を目指すものである。

遠隔コラボレーションにおける同室感を高める一つの方法として、上述の音像生成法とローカル・ラグ機能を持つ音声伝送法とを統合する意義は大きいものと考えられる。統合手法は、遠隔地にある(協調作業員などの)視覚聴覚オブジェクトを、映像と音像の再生場所を一致させ、かつ知覚における同期のずれを小さくした上で、大型ディスプレイ上に実寸大で再生することを目指す。

本研究は、上記の目的をもって、音響反射板つきスピーカユニットを制御する機能を、ローカル・ラグ機能を持つ音声伝送サーバに組み込み、新たに Mac OS 上で動作する、音像生成機能とローカル・ラグ制御機能とを併せ持つ音声伝送サーバとして実現するものである。伝送すべき音源(音像に対応)の数は 1 つとは限らない。従って、音像生成制御システムは、同時に複数の音声ストリームを出力・制御する必要がある。本研究では、この機能を実現するため、音像生成制御システム開発の先行研究⁵⁾に倣い、Mac OS に付属の音声 API、CoreAudio を利用する。以下、本稿では、開発する音声伝送サーバの概要と実装仕様、そして実装したサーバの動作評価結果を報告する。

2. 関連研究

2.1 音響反射板つきスピーカユニットを持つ音像生成法

2.1.1 装置概要

音響反射板つきスピーカユニットを持つ音像生成法とは、大型ディスプレイ中に再生される音像の正しい方向定位の実現を目指して、提案されたものである⁶⁾。図 2.1 にその装置概要を図解する。大型ディスプレイと、その左右端に設置された、L 字型の音響反射版によって囲まれた 3 機のスピーカからなるスピーカユニット(左右それぞれ 1 ユニットずつ)で構成されている。音響反射板とディスプレイとによるスピーカ出力の反射によって、ディスプレイ上の指定位置に音像が生成されることを目指すものである。受聴位置がディスプレイ中央の前にあるとき、通常のスtereo再生と同様に出力音は受聴位置に到達する。受聴位置がディスプレイの左端あるいは右端に近いときは、出力音は音響反射板によって減衰・回折され、Stereo再生に伴う音像のずれが軽減されることが期待できる。

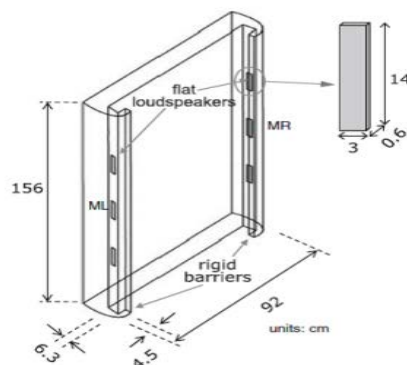


図 2.1 音響反射板で囲まれたスピーカユニットから成る音像生成法の装置概要(文献⁶⁾から引用)。

†同志社大学, Doshisha University

2.1.2 制御システム

当手法は、ディスプレイ上の音像位置の左右あるいは上下の移動を、両端スピーカの出力音の音圧レベルを制御することによって実現する。これまで開発されてきた、この制御を行うための制御システムの概要を図 2.2 に図解する。

システムは、スピーカ出力を直接操作するサーバと、マイクなどの入力を制御するクライアントによって構成されている。クライアントは、音声データを伝送すると同時に、再生すべき音像位置座標データをサーバに送信する。サーバは、送られてきた座標データを基にスピーカ出力を制御し、大型ディスプレイ上に音像を生成する。

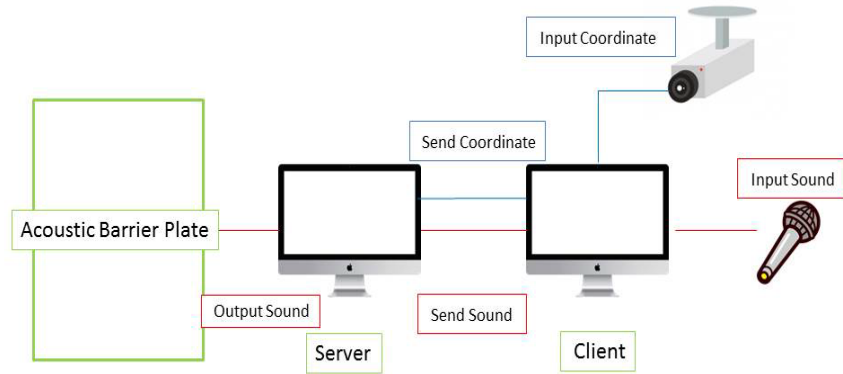


図 2.2 音響反射板を利用する音像生成法における制御システムの概要。

2.2 ローカル・ラグ制御機能を持つ音声伝送システム

2.2.1 音声伝送における同期のずれ

音声伝送における同期のずれの問題を説明するため、同期的協調作業の一例である、遠隔地間の利用者同士が合奏する、遠隔合奏をとりあげる。

まず、2種類の音声データストリームを定義する。演奏者自身のサイトで再生される“フィードバック音”と、演奏者による演奏音がネットワークを経由して遠隔地において再生される“フィードスルー音”である。

ある地点にいる演奏者は、自身が発するフィードバック音と遠隔地にいる合奏者が発するフィードスルー音を聞きながら演奏することになる。このとき、フィードスルー音は、伝送途上にあるネットワークやコンピュータ等の処理

に起因する、不可避的な遅延をもつ。演奏者は、相手に合わせようとして、自身の演奏を遅らせる。その遅れは、伝送によってさらに拡大される。結果的に、合奏、すなわち同期的協調作業の維持は困難になる。

2.2.2 ローカル・ラグ法

ローカル・ラグ法とは、原理的に不可避の伝送遅延の完全な除去に代え、知覚レベルにおける遅延の低減を目指すアプローチである。遠隔地間の伝送遅延を計測し、その伝送遅延に相当する遅延をフィードバック音に付加し、知覚における同期の達成を目指す。図 2.3 にその仕組みを図解する。

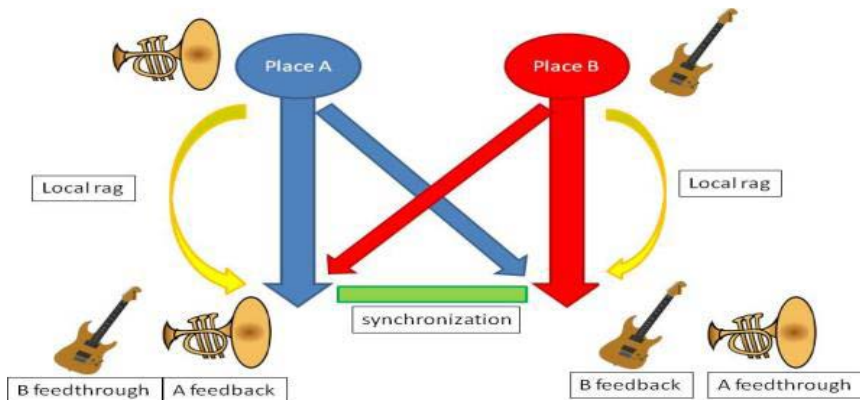


図 2.3 ローカル・ラグ法の概念。

2.2.3 実装システム

上述のローカル・ラグ制御機能を持つ、遠隔地間の協調的同期作業を支援するための音声伝送システムが開発されてきた。本稿で新たに開発するシステムは、その基本を当音声伝送サーバに拠っている。

実装された音声伝送システムは、Linux 上で開発され、音声伝送サーバと同期制御サーバとの2種類のサーバで構

成されている（図 2.4 参照）。音声伝送サーバは、ローカル・ラグ制御機能と他の音声伝送サーバとの伝送遅延を計測する機能を持っている。一方、同期制御サーバは、音声伝送サーバ間の伝送遅延の最大量を全ての音声伝送サーバに通知し、音声伝送サーバの同期制御を支援するサーバである。

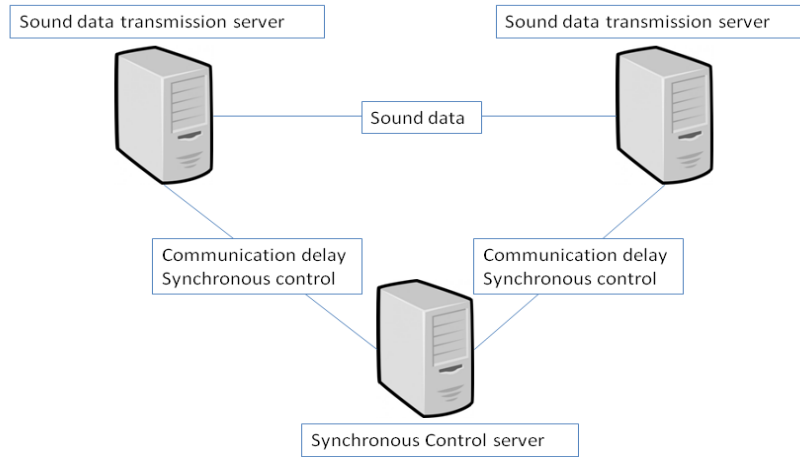


図 2.4 ローカル・ラグ制御機能を持つ音声伝送システムの構成。

3. 提案システム

図 3.1 に、本研究で開発を進める、新システムの概要を図解する。システムは、遠隔地間のデータの伝送をネットワーク経由で行う。またシステムは、音像と映像との双方の同期的生成（再生）を行う。

音像再生部に関しては、先行する音像生成システムと同様に、3機のスピーカを音響反射板で囲むスピーカユニットを大型ディスプレイの左右端に設置する。

映像処理サーバは、IPカメラによって送られてきた映像

から音源位置の抽出処理を行い、その位置（座標）情報を音声伝送サーバに送信する。

本システムは、やや役割が異なる2種の音声伝送サーバ（図中のAおよびB）を持つ。なお、いずれのサーバもMac OS 上に実装する。音声伝送サーバBは、入力された音声データを同期制御サーバ兼音声伝送サーバAに送る。音声伝送サーバAは、送られてきた座標を基に各スピーカに出力すべき音の音圧レベルを計算し、受信した音声を、それぞれのスピーカに多チャンネル出力する。

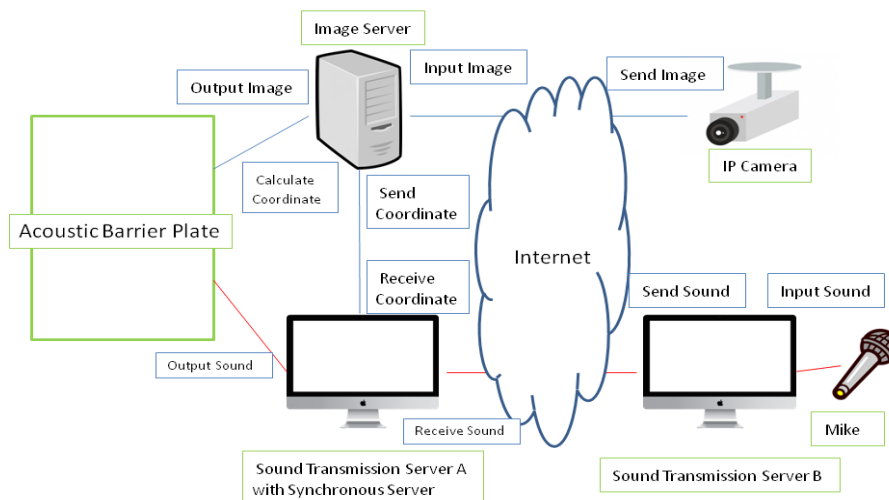


図 3.1 提案システムの構成。

4. 評価実験

4.1 目的

先行して開発されてきた、ローカル・ラグ制御機能を持つ音声伝送システムは、Linux 上に開発されてきた。しかし、Linux が通常用いている音声 API は、複数の音声出力ストリームを同時に扱うことが難しく、例えば複数人の音声や複数の楽器音などの音像の同時再生が期待される音像再生システムの開発には適さない。そこで本研究では、Mac OS の CoreAudio を用いて音像生成法のための制御システムを実装した例を基に、ローカル・ラグ制御機能を持つ音声伝送システムを Mac OS 用に移し変え、合わせてそこに音響反射板つきスピーカユニットを用いる音像生成システムも組み込む実装を行う。

本システムの性能を左右する要点は、6 機のスピーカの出力間同期の実現にある。この点に着目し、実装したシステムの出力信号の観測を通して、出力間同期の質を評価する。

4.2 実装の概要

図 3.1 に示すシステムを、表 3.1 の性能を持つコンピュータ上に実装した。また、スピーカユニットとコンピュータをつなぐ（多チャンネルの）音声チャンネルには、やはり表 3.1 に示すオーディオインターフェースを利用した。

表 3.1 実装システムの諸元。

音声伝送サーバ A, B	iMac 27 Inch 2013 Late
OS	10.11.5
プロセッサ	3.5GHz Intel(R) Core i5
メモリ	8GB
PC 入力用オーディオインターフェース	Roland Quad-Capture
多チャンネル入出力用オーディオインターフェース	MOTU 828k

4.3 評価実験の概要

音声伝送サーバを 1 台だけ動作させた場合の実験と、2 台の音声伝送サーバを接続して動作させた場合との、2 種の評価実験を行った。1 台の音声伝送サーバのみを用いた場合の機器の構成を図 4.1 に示す。入力音声を電子メトロノーム A の出力とし、音声伝送サーバ A のフィードバック音と電子メトロノーム A の直接音の差を計測する。この差を計測することで、サーバにおける計算処理が要する遅延の大きさや安定性等を知ることができる。なお、音声伝送サーバ A は同期制御サーバも兼ねている。

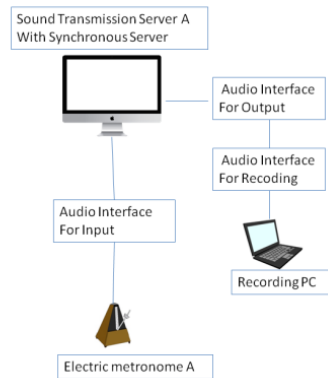


図 4.1 1 台の音声伝送サーバを動作させた計測実験における機器構成。

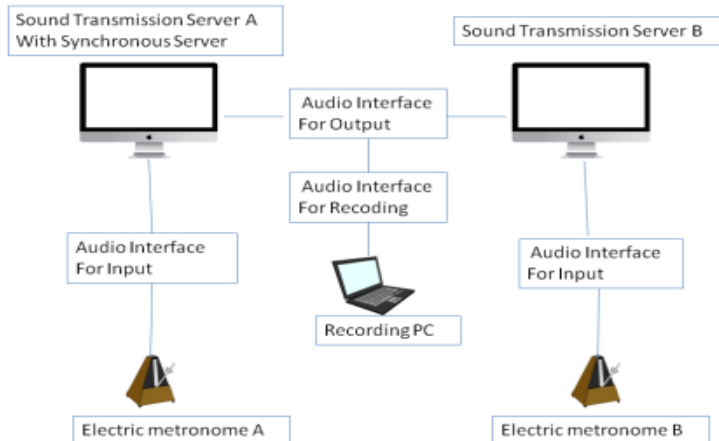


図 4.2 2 台の音声伝送サーバを接続して行った計測実験における機器構成。

2台の音声伝送サーバAとBを用いる場合の機器の構成を図4.2に示す。音声伝送サーバAは音声伝送サーバ兼同期制御サーバである。入力する音声は電子メトロノームAとBから入力用のオーディオインターフェースを用いて、ライン入力を行う。以下電子メトロノームAとBから出力される音を、それぞれ直接音AとBと呼ぶ。音声伝送サーバから出力される音声を録音するために、オーディオインターフェースも2台用意し、出力用と録音用に振り分ける。録音用のPCで録音を行う。録音する音声に関しては、音声伝送サーバAに入力した音を音声伝送サーバA自身が再生するAのフィードバック音、音声伝送サーバBに入力した音を通信し音声伝送サーバAで再生するBのフィードスルー音を録音し、フィードバック音とフィードスルー音の遅延量を評価する。なお、サーバの接続は、サーバ以外の遅延に対する影響要因を排除するため、LANを用いて行った。

また、音声出力の計測は、表3.2に示すような機器あるいはソフトウェアを用いて行った。

表3.2 計測に用いた機器とソフトウェア。

録音用ソフトウェア	SONAR X1 PRODUCER
OS	Windows Vista Business
プロセッサ	Intel(R) Core(TM)2 Duo 1.2GHz
メモリ	2.0GB

なお、計測結果の信頼性を高めるため、それぞれの音出力の組み合わせ毎に、30分間の録音を行い、その録音された音信号全体に関して遅延の計測を行った。また、そうした30分のデータを用いた計測実験を3度繰り返した。

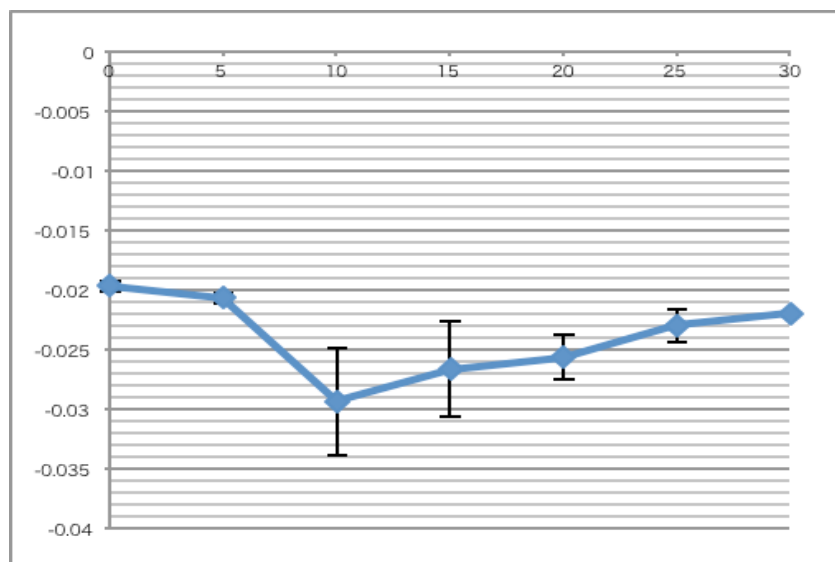


図4.1 音声伝送サーバ1台を動作させたときの直接音とフィードバック音の差の平均。

4.4 実験結果

4.4.1 音声伝送サーバ1台を動作させたときの直接音とフィードバック音の差の計測

図4.1は音声伝送サーバ1台のみを動作させたときのフィードバック音に関する遅延計測の結果を示している。横軸は、録音（録音時間は30分）と計測（5分おきに計測）を行った時間を示している。縦軸は、電子メトロノームAと音声伝送サーバから出力されたフィードバック音の差（秒）である。折れ線グラフ中の黒の縦バーは、3回の計

測に伴う標準偏差を示している。結果から、直接音とのフィードバック音との差は約20ms~30msの間に収まっていることがわかる。

4.4.2 2台の音声伝送サーバを接続したときの直接音とフィードバック音の差の計測

図4.2は直接音Aと音声伝送サーバAのフィードバック音の差を示している。縦軸と横軸は先述と同様である。グラフより、音声伝送サーバの起動時に最も差が大きく、時間の経過と共に差が小さくなっていることがわかる。

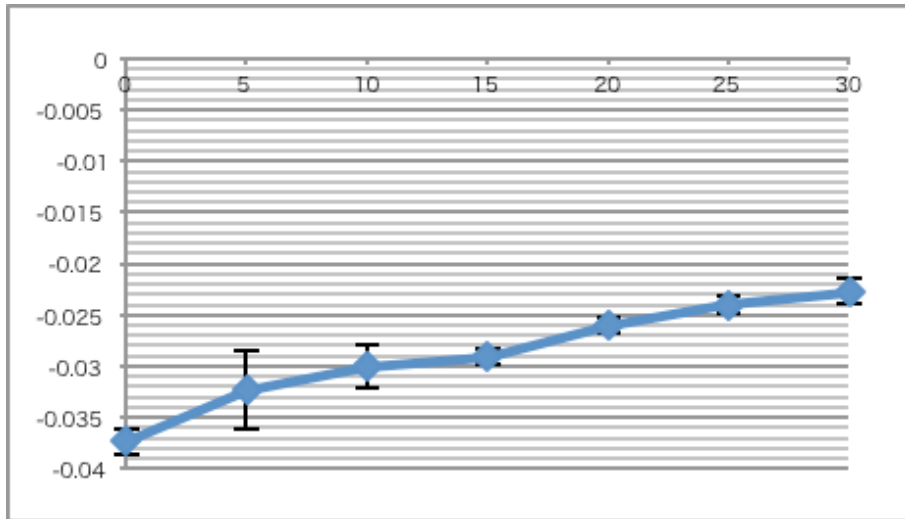


図 4.2 音声伝送サーバAのフィードバック音と直接音Aの差の平均。

4.4.3 2台の音声伝送サーバを接続したときの直接音とフィードスルー音の差の計測

図 4.3 は直接音Bと音声伝送サーバBのフィードスルー

音の差のグラフである。縦軸と横軸は先述と同様である。グラフより通信遅延の差は 25ms～30ms の間に収まっていることがわかる。

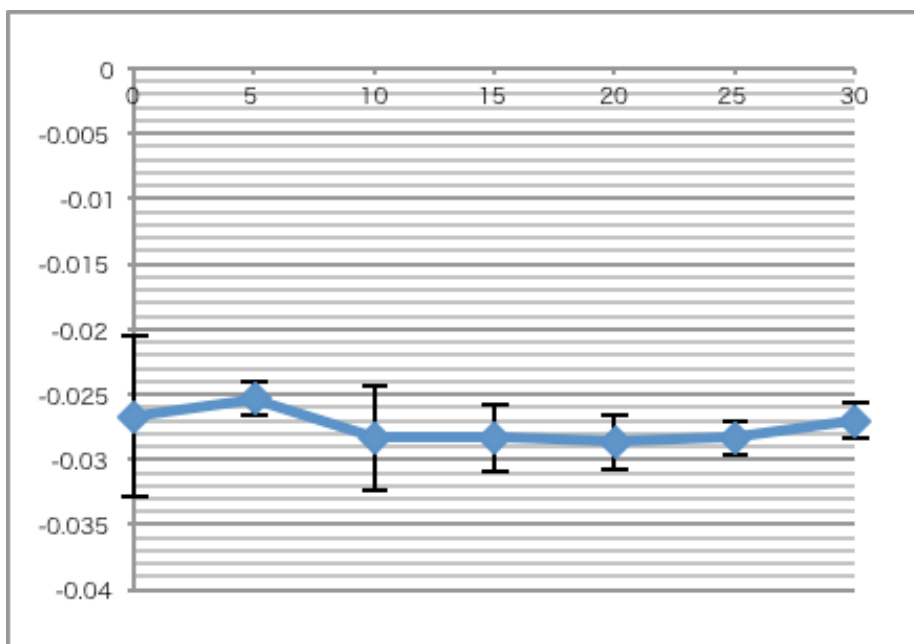


図 4.3 音声伝送サーバBのフィードスルーと直接音Bとの差の平均。

4.4.4 多チャンネル出力時の出力信号の観測

多チャンネル出力をする際に重要な点は、スピーカ同士の同期がとれていることである⁴⁾。全6機のスピーカの出力の同期がとれているか否か、この点を確認するため、音声伝送サーバ起動時と動作から30分経過した時点における出力の状況を観測した。図 4.4 は、音声伝送サーバを1

台のみ動作させたときの、各スピーカの出力の状況を表している。図中、横軸は時間軸であり、縦軸は音信号の振幅である。なお、縦方向に6機のスピーカの出力を並べている。

計測に用いた入力音は、電子メトロノームのパルス波である。信号波形の立ち上がり部に注目すると、全スピーカの出力間で、正確に同期がとれていることがわかる。この

同期がとれている状況は、サーバ起動直後でも、起動から30分経過した場合でも、同様に観察できた。

また、図 4.5 より、2 台の音声伝送サーバを接続した場合においても、同様の正確な同期を観察できた。

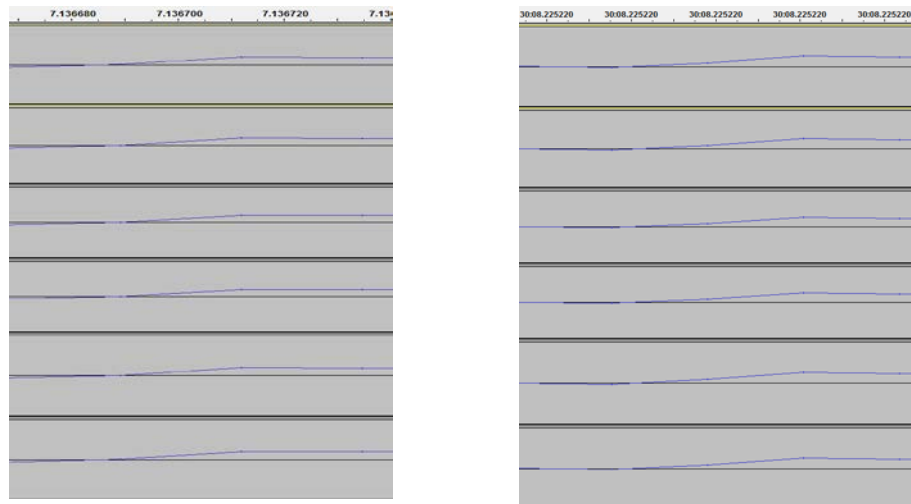


図 4.4 1 台の音声伝送サーバのみを動作させたときの 6 チャンネル音出力（左：起動時、右：30 分後）。

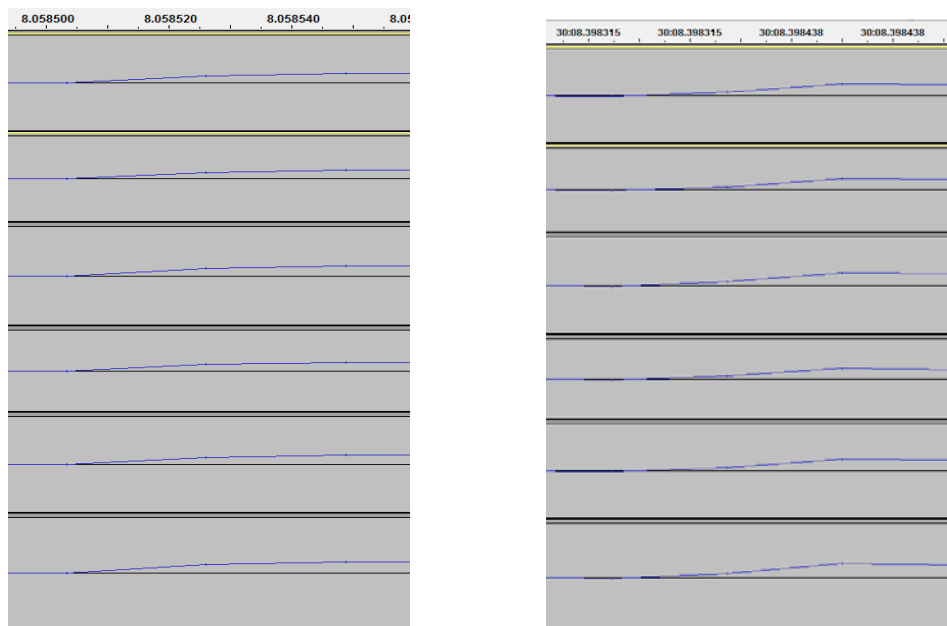


図 4.5 2 台の音声伝送サーバを動作させたときの 6 チャンネル音出力（左：起動時、右：30 分後）。

5. まとめ

ローカル・ラグ制御機能を持つ音声伝送サーバと 6 機のスピーカの出力を制御して大型スピーカ上に音像を生成する制御サーバを統合した、新しい音声伝送サーバを、Mac OS 上に実装し、その性能の評価を行った。

1 台の音声伝送サーバのみを動作させたときのフィードバック音の（基準となる直接音に対する）遅延と、2 台の音声伝送サーバを接続したときのフィードスルー音の遅延は、20ms~30ms の間に収まっていた。しかしその一方、2

台の音声伝送サーバを接続した際のフィードバック音の遅延は、音声伝送サーバ起動時にやや大きく、しかも時間経過と共に徐々に小さくなっていく様子が確認された。この現象の原因については、引き続き検討が必要である。

また、6 機のスピーカに対応する 6 チャンネルの音出力間の同期を観測した結果、サーバ起動時においても起動から 30 分間経過した時点においても、正確な同期がとられていることが確認できた。

計測結果の一部に、遅延の原因の解明を要する結果が含

まれてはいたが、全体としては、高々30ms程度の遅延に収まる安定した音声伝送が実現できていることを確認できた。今後は、本伝送サーバを用いた音像位置制御や高い同室感をもつ遠隔コラボレーション支援の実験を進めていきたいと考えている。

参考文献

- 1) 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).
- 2) 柴田真尚, “マルチメディア遠隔コラボレーション支援システムのための音場制御システムの構築”, 同志社大学修士論文, (2010).
- 3) Dane Stuckel and Carl Gutwin;The Effects of locallag on Tightly-Coupled Interaction in Distributed Groupware : Computer Supported Cooperative Work, pp.447-456 (2008).
- 4) 大島義博, “ローカル・ラグ制御をもつ音声伝送サーバの遠隔合奏による評価”, 同志社大学修士論文, (2015)
- 5) 中谷彰皓, “複数音像の同時生成を可能とする多チャンネル音出力制御システムの開発”, 同志社大学修士論文, (2015)
- 6) Gabriel Pablo Nava , Keiji Hirata , Masato Miyoshi , “A loudspeaker design for sound image localization on large flat screens ” , Acoustical Science and Technology Vol.31, No.4, pp.278-287, (2010).