Vol.2017-GN-101 No.27
2017/3/11

大型ディスプレイ画面上に音像を生成する機能を持つ音声伝送サーバの開発

松尾雄真† 片桐滋† 大崎美穂†

概要：遠隔地をつなぐコラボレーションにおける参加者が，あたかも同室にいるような同室感を共有できることは重要である．そうした同室感の達成を目指すアプローチとして，遮音カバーで囲まれたスピーカユニットを用いて大型ディスプレイ上の適切な位置に音像の生成を試みる技術と，メディア信号の伝送において不可避の遅延を知覚レベルで軽減することを目指すローカル・ラグ法に着目している．本稿は，この音像生成技術を，ローカル・ラグ制御機能を持つ音声伝送サーバに組み込み，遠隔コラボレーションにおける同室感の向上を目指す新しい音声伝送サーバを開発した結果を報告するものである．評価実験を通して，6 チャンネルに及ぶ音声出力が精確に同期され，かつサーバ内の処理遅延もほぼ予想通りの 20ms 程度であったことを明らかにする．

キーワード：音響反射板，音声伝送サーバ

Development of Sound Transmission Server Producing Sound Images on Life-Size Display Screens

Yuma Matsuo† Shigeru Katagiri† Miho Ohsaki†

Abstract: For people who try to collaborate beyond the distance, it is valuable that they can share the feeling of “being in the same room”, which is expected to increase their collaboration quality. Two approaches have been studied to achieve such feeling: a method of generating sound images in a life-size display using sound-insulation-covered speaker units, and a method for alleviating lag in perception level by adding lag to the media data of a local site. Focusing on these two studies, we develop a new sound transmission server that integrates the above two methods. In this paper, we report the design and implementation of the server, and demonstrate that our server accurately synchronizes sound data for the 6-channel outputs of the speaker units with an expected additional delay of 20ms.

Keywords: Acoustic Plate, Sound Transmission Server

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，視覚メディアや聴覚メディアに関する対称性，即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感覚“同室感”を遠隔地間にいるシステム利用者同士が十分に共有できるまでには至っていない．

こうした問題を解決するアプローチの一つとして，大型ディスプレイの左右端に遮音カバーによって囲まれたスピーカユニットを用いる音像生成手法が提案され，その性能の検証が行われている[2]．しかし，その手法を実装した装置にはまだ遠隔地間で音声データを伝送する機能が実装されておらず，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までにはいっていない．

一方，遠隔コラボレーション支援システムの一つ t-Room の制御用に，データ伝送に不可避の遅延を知覚レベルにおいて軽減することを目指したローカル･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，メディアデータ伝送地点間の伝送遅延を計測し，その遅延分をあえて受け入れ，伝送を伴わないメディアデータにも同量の遅延を付加することによって，メディア知覚における時間的なズレの克服を目指すものである．

t-Room のような大型ディスプレイを用いる遠隔コラボレーションシステムにおける同室感を高める一つの方法として，上記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．統合によって，遠隔地にある（人物などの）視聴覚オブジェクトの映像と音像を大型ディスプレイ上の正しい位置（方向）に同期的に再生できることが期待される．本研究は，上記の目的をもって，遮音カバーつきスピーカユニットを制御する機能とローカル・ラグ機能とを併せ持つ音声伝送サーバを開発するものである．

2. 関連研究

2.1 遮音カバーつきスピーカユニットを用いる音像生成法

2.1.1 装置の概要

この音像生成手法（以下，遮音カバーつき音像生成法）とは，標準的な 2 チャンネルステレオ方式によって生成される音像が受聴位置によって偏ってしまう問題の軽減を目指して提案されたものである[5]．図 1 に，そこで用いられるスピーカユニットなどの装置の概要を図解する．大型ディスプレイの左右端に，L 字型の遮音カバーによって囲まれた 3 機のフラットスピーカから成るスピーカユニットが設置されることで特徴づけられる．左右のスピーカは，ディスプレイを挟むように対峙し，遮音カバーによって直接音を抑制し，同時にディスプレイ面から反射される音によ

† 同志社大学
Doshisha University.

ってディスプレイ面中の指定位置に音像を生成することを目指す．

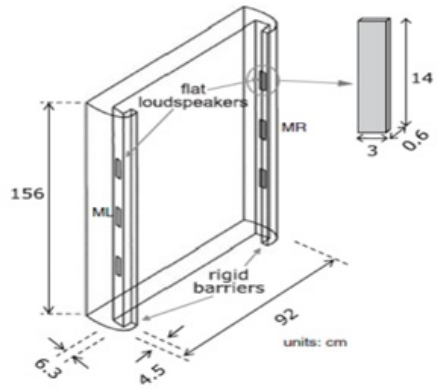

図 1　遮音カバーで囲まれたスピーカユニットから成る音像生成法の装置概要(文献[6]から引用)．

2.1.2 データの制御手順

この方式における音像位置は，左右のスピーカ出力の音圧レベルをタンゼント則[7]に基づいて変更することによって制御する．

先行研究で実装されてきた制御手順の概要を図 2 に図解する．手順は，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情報のデータをサーバに送信する．この時，クライアントは座標データの伝送路や音声伝送データの伝送路を設定した XML ファイルを自動生成し送信する．サーバは，送られてきた座標データをもとに 6 機のスピーカの音圧レベルを計算し出力することによって，大型ディスプレイ上に音像を生成する．この 2 つの機能を組み合わせることによって，ディスプレイ上に音像を定位することができ，遠隔間におけるメディアの対称性を保つ．

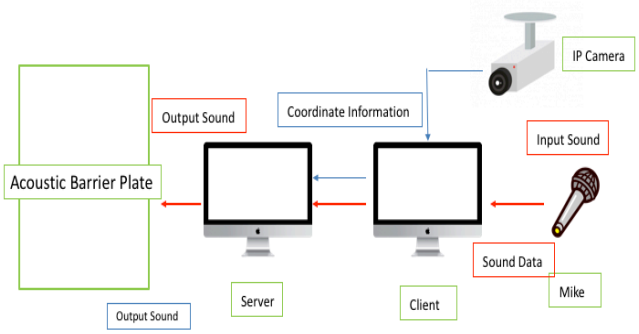

図 2　制御手順の実装装置．

2.2 音声伝送における同期のずれ

同期的協調作業の典型として，離れた 2 地点，即ち地点 A と地点 B とにおいて楽器を用いた遠隔合奏を行う事例を考える．地点 A においては，自身の演奏音である“フィードバック音”とネットワークを経由して伝送されてくる地点 B の演奏音である“フィードバック音”とが出力されることになる（地点 B においても，関係は対称となるが，同様に 2 種の音が出力される）．このとき，フィードスルー音には，伝送に伴う不可避の遅延が含まれる．従って，地点 A の演奏者がこの遅延を伴う地点 B のフィードスルー音に合わせて（同期的に）演奏しようとすると，自身の演奏が遅れる．この遅延が大きいとき，両地点の演奏者が相手に合わせようとするがゆえに,結果的に互いの演奏が遅れ，合奏が破綻に至る．

2.3 ローカル・ラグ制御機能

原理的に，伝送に伴う遅延を避けることは困難である．ローカル・ラグ法[3]は，この遅延の存在を受け入れ，人間の知覚レベルにおける遅延を小さくすることで前述の破綻のような状況を避けることを目指す．図 3 にその仕組みを図解する．手法はまず，2 地点間の伝送遅延を測定する．そして，フィードバック音にその伝送遅延と同量の遅延を付加することにより，フィードバック音とフィードスルー音の出力を同期させ，その地点における演奏者（コラボレーション作業者）の知覚的な遅延を緩和する．

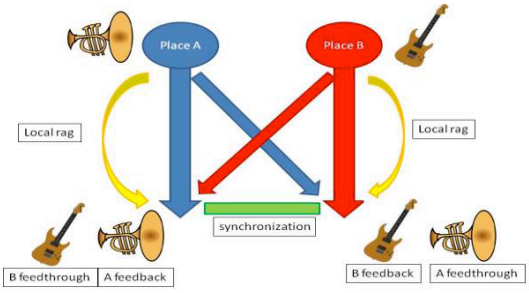

図 3　ローカル・ラグ法の概念．

2.4 ローカル・ラグ制御機能をもつ音声伝送システム

上述のローカル・ラグ制御機能を持つ音声伝送システムが開発されてきた[8][9]．システムは音声伝送サーバと同期制御サーバの 2 種類のサーバで構成されている(図 4)．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つ．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，音声伝送サーバ間の同期制御を支援する．

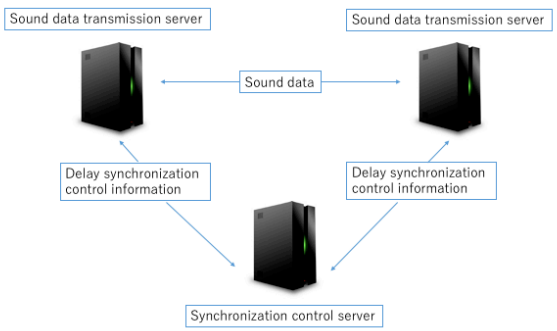

図 4　遠隔合奏支援システムの概要．

3. 提案音声伝送サーバ

遮音カバー付き音像生成法のデータ制御システムには，遠隔地から音データを受信する機能がなかった．また一方で，ローカル・ラグ制御機能を持つ音伝送システムには音像を制御する機能がなかった．それぞれの長所を生かし，短所を補う，自然な発展は，両者を統合することである．こうして本研究においては，両者の機能を統合する新しい音声伝送サーバを開発する．なお，新サーバは Mac OS 上で開発する．先行研究で， Windows の音声 API である ASIO を用いた複数音像の生成を適切に制御できない一方で，Mac OS の音声 API である CoreAudio を用いることでその問題を解決できることが明らかにされていたためである[6]．

図 5 に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い，言うまでもなく，音像生成には遮音カバーつき音像生成装置を用いる．映像サーバは，IP カメラから送られる映像から音像位置座標の抽出処理を行い，その座標情報を音声伝送サーバに送信する．

本システムは，役割が異なる 2 種類の音声伝送サーバ(図 5 のサーバ A およびサーバ B) を持つ．音声伝送サーバ B は入力された音声データを，同期制御サーバを兼ねる音声伝送サーバ A に送る．音声伝送サーバ A は映像サーバから送られてきた座標データを基に各スピーカに出力すべき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．なお，新システムの入力に関しては，システムの複雑化を避けるためにモノラル入力とする．

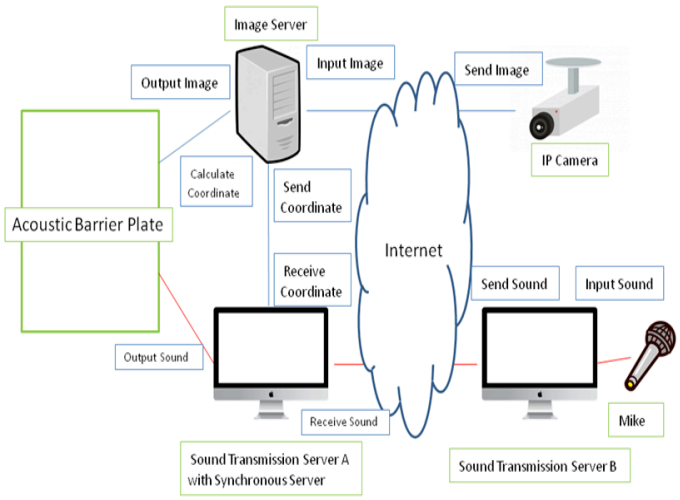

図 5 提案システムの概要図．

4. 評価実験

4.1 目的

新たに実装した音声伝送サーバの性能を，サーバ内のバッファリングや演算などに伴う処理時間（遅延量）と，遮音カバーつき音像生成法において特に重要となる 6 機のスピーカの出力間同期の性能，目標音像位置に応じた各スピーカ出力の音圧レベルの正確さと着目して，評価した．

4.2 音声伝送サーバの内部処理時間の計測

4.2.1 実験の概要

音声伝送サーバの処理時間を評価するために 1 台だけ稼働した状態で実験を行う．音声伝送サーバの処理時間を計測することを目的としているため，ローカル・ラグ制御と座標計算は行なっていない．

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 6 に示す．入力音声を電子メトロノームの出力とし，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によって，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を知ることができる．また，実装したシステムの諸元は表 1 の通りである．

表 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	Mac OS 10.11.6
プロセッサ	3.5GHz Intel(R) Core i5
メモリ	8GB
PC 用入力用オーディオインターフェース	Roland Quad-Capture
多チャンネル入力用オーディオインターフェース	MOTU 828k

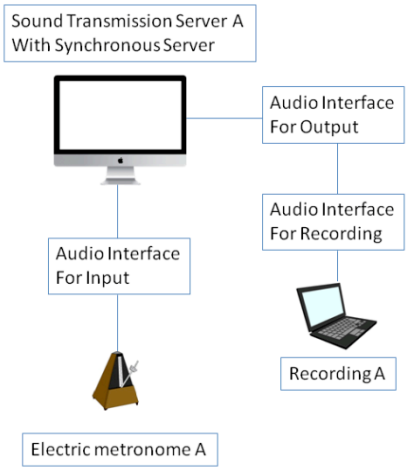

図 6 音声伝送サーバ 1 台稼動時の実験環境図．

4.2.2 直接音とフィードバック音の差

図 7 は，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に関する遅延計測の結果を示している．なお本稿における遅延の計測は全て，電子メトロノーム音の波形の立ち上がり時刻の目視による比較観測によって行った．観測は 5 分毎に行った．図中縦軸は，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観測時刻における差である．差は，直接音時刻からフィードバック音時刻を引くことで求めた．折れ線グラフの中の黒の縦バーは，3 回の計測に伴う標準偏差を示している．結果より，直接音とフィードバック音の差は， 20ms から 30ms の間にほぼ収まっていることがわかる．サーバに用いたコンピュータの（OS によって提供されたアプリを用いた）単純な音データの入出力処理時間は約 20ms であった．従って，これに準じる処理時間を示したサーバの実装結果

は，基盤としたコンピュータの性能をほぼ十分に利用したものであったように考えられる．

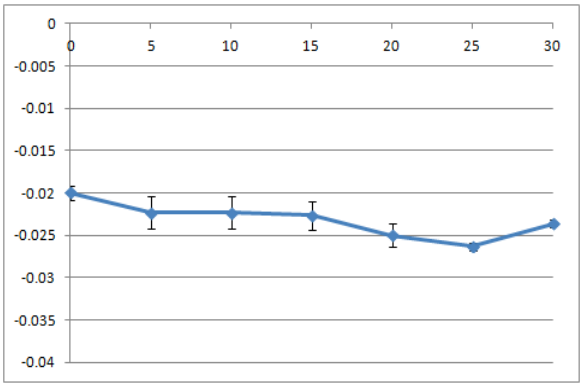

図 7　1 台の音声伝送サーバのみを稼動した際の直接音とフィードバック音の差（縦軸：単位は秒）．横軸は観測時刻（単位は分）．

4.3 データ伝送を伴う場合の音声伝送サーバの性能評価

4.3.1 実験の概要

2 台の音声伝送サーバを稼働して音声伝送サーバの処理時間の計測を行った．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 8 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 を用いてオーディオインターフェースを介してライン入力を行った．音声伝送サーバから出力される音を録音するために，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，録音用のコンピュータを用いて録音を行った．比較は，電子メトロノームの出力である直接音と，音声伝送サーバが（サーバ間伝送を行わずに）直接出力するフィードバック音，サーバ間の伝送を経たフィードスルー音との間で行った．実験は，不測の影響を排除するため LAN を用いて行った．また，音声伝送サーバの諸元と計測機器は表 1 と同様である．

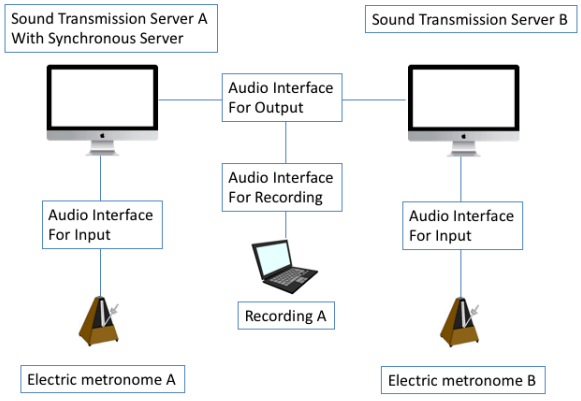

図 8　音声伝送サーバ 2 台稼動時の実験図．

4.3.2 直接音とフィードバック音の差

図 9 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．縦軸と横軸は図 7 等と同様である．サーバ起動時に遅延の差が最大になるが，25ms から 35ms の間に収まっていることがわかる．

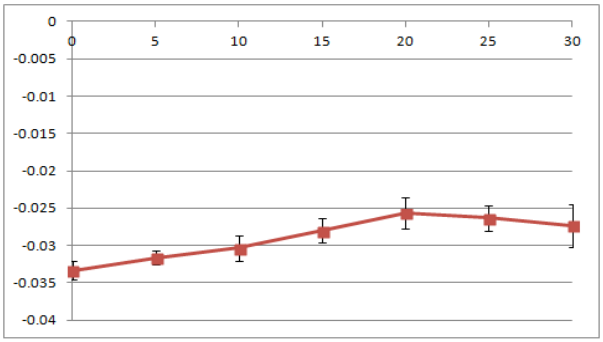

図 9　音声伝送サーバ A のフィードバック音と直接音 A の差（縦軸：単位は秒）．横軸は観測時刻（単位は分）．

4.3.3 直接音とフィードスルー音の差

図 10 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．縦軸と横軸に関しては図 7 等と同様である．グラフより通信遅延の差は 25ms から 30ms の間に収まっていることがわかる．

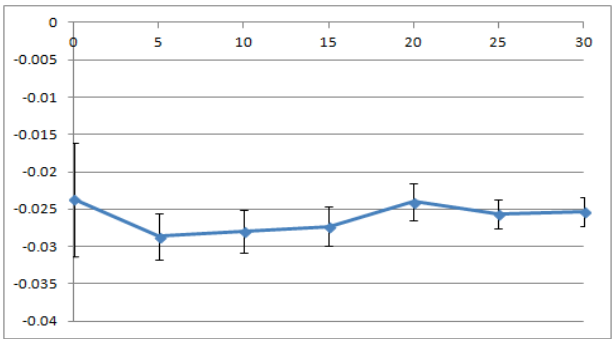

図 10　音声伝送サーバ B のフィードスルー音と直接音 B の差（縦軸：単位は秒）．横軸は観測時刻（単位は分）．

4.4 音像位置制御に関する音声伝送サーバの性能評価

4.4.1 実験の概要

図 11 に示すような，2 台の音声伝送サーバを接続し映像処理サーバから送られてくる座標情報を受け取り，受け取った座標に従い音圧レベルを計算し，音を出力する利用環境を想定する．録音環境等は前節の評価実験と同様である．

生成すべき音像位置を明示的に制御するために，映像処理サーバを代行する，決められた座標を 30ms ごとに送信するテストプログラムを用意した．ここで 30ms としたのは，映像処理サーバが処理する予定の映像のフレームレートが 30fps であることに依る．座標情報を送るコンピュータと音声伝送サーバ A および B を HUB で接続し，音声伝送サーバがともに座標情報を受信し，それぞれにおいて出力すべき音データの振幅を計算するように設定した．なお，

自地点の入力音に対して，他地点の音源座標情報に合わせた音振幅計算を施すことを避けるため，フィードバック音の再生は取りやめた．また，録音環境の都合により，音声伝送サーバ B のフィードスルー音のみを録音する．

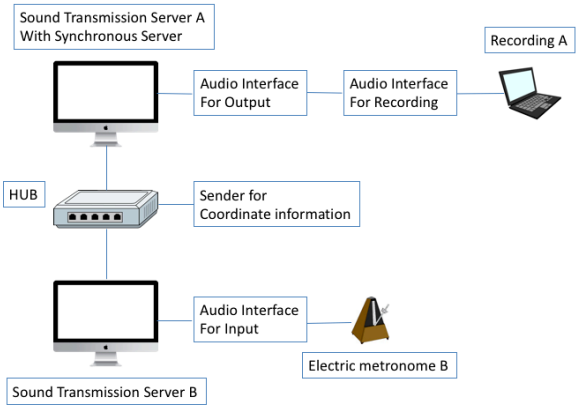

図 11 音像位置制御に関する評価実験の構成．

4.4.2 結果

図 12 は，座標情報から音圧レベル計算を伴った音声伝送サーバ B のフィードスルー音と直接音との時間差である．約 20ms から 25ms の間で動作していることが確認できる．サーバの音声の入出力が 22ms 付近なため安定して動作していることがいえる．

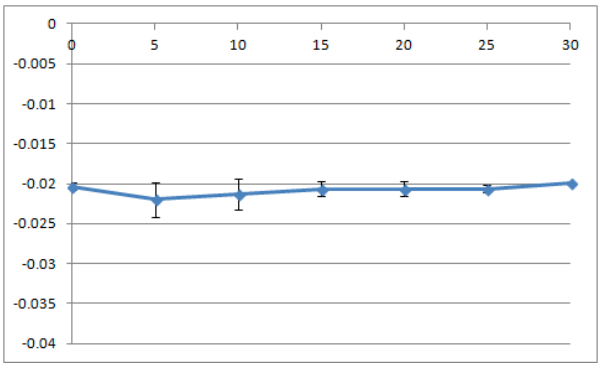

図 12 音圧計算を伴った音声伝送サーバ B のフィードスルー音と直接音 B との差（縦軸：単位は秒）．横軸は観測時刻（単位は分）．

4.5 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，スピーカ間の同期が取れていることである[6]．全 6 機のスピーカ間の同期性能を確認するため，音声伝送サーバ起動時と起動から 30 分経過した時点における同期の状況を観測した．図 13 と図 14 は 1 台の音声伝送サーバを稼働した時の，各スピーカの出力を図解している．図中の縦軸は振幅で，横軸は時間である．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った．波形の立ち上がり部に注目すると，全スピーカの出力音で正確に同期が取れていることがわかる．この同期が取れている状況は，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確認することができる．また，2 台の音声伝送サーバを接続した場合においても，同様の正確な同期の実現を確認することができた．

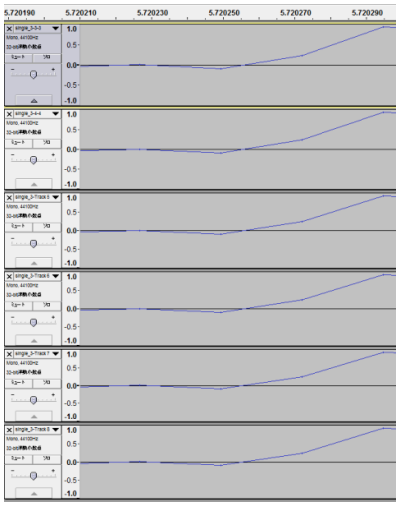

図 13 1 台の音声伝送サーバを用いた際の起動直後の 6 スピーカ出力音．縦軸は音振幅，横軸は時間．

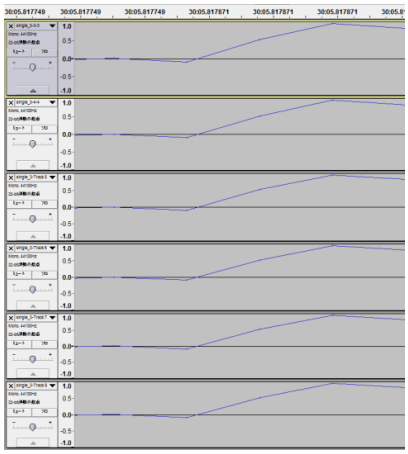

図 14 1 台の音声伝送サーバを用いた際の起動から 30 分後の 6 スピーカ出力音．縦軸は音振幅，横軸は時間．

4.6 出力信号の音圧レベル計算に関する観測

4.6.1 概要

指定された座標に音像を生成するために，音圧レベルを制御する音の振幅重み係数を受信音に掛ける．重み係数は 1.0 のとき最大出力とし，0 の時は出力しない．図 15 は，遮音カバーつき音像生成法における座標構成と 6 機のスピーカの番号を対応づけた図である．左上を 0 とし，横軸の最大座標は 1080，縦軸のそれは 1920 とした．スピーカに関しては，左上から順に 1 から 6 まで番号を付与した．各チャンネルの音信号の振幅が制御の意図通りに設定されているかどうかの確認をするために，重み係数と目標音像位置との関係が明確な音像位置(540,480)を例として利用した

重み係数を表 2 に示す.

表 2 目標音像位置(540, 480)に対する音圧係数.

スピーカ番号	音圧係数
1	0.50
2	0.50
3	0.50
4	0.50
5	0
6	0

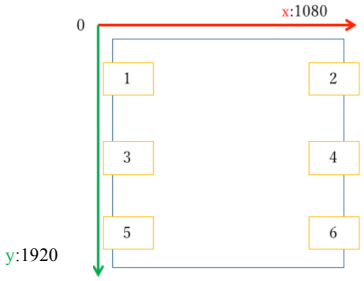

図 15 遮音カバーつき音像生成法における座標と 6 機のスピーカの配置.

4.6.2 出力信号の振幅の観測

目標音像位置(540,480)に対する重みを掛け合わせた出力信号を観測した．音圧の計算には 2 チャンネルスピーカの標準的なパンニング法であるタンゼント則[7]を用いた．スピーカを番号順に上から並べ，縦軸・横軸に関しては図 13 等と同様である．図 16 は重み係数がすべて 1 のときの音信号である．振幅の最大値が，想定通りの 0.5（左右対の和が 1 になるように正規化しているため）付近にあることが確認できる. 図 17 は音像位置(540,480)に対応する音出力である．表 2 のようにスピーカ 1 から 4 の振幅の最大値がほぼ 0.25 となっており，スピーカ 5 と 6 の出力音はなく，表 2 の重み係数が適切に反映されていることわかる.

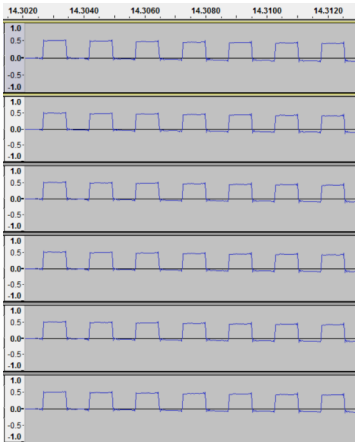

図 16 重み係数を全て 1 としたときの音出力．縦軸は音振幅，縦軸は時間.

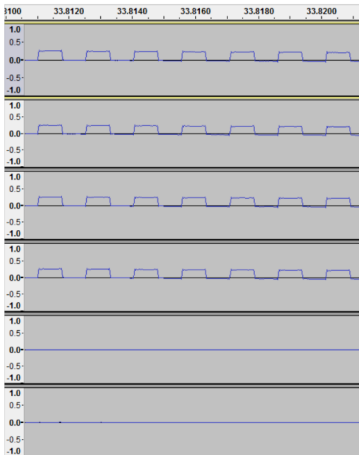

図 17 目標音像位置(540,480)に対応する重み係数を用いたとき音出力．縦軸は音振幅，縦軸は時間.

5. まとめと今後の課題

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ディスプレイ上に音像を生成する遮音カバーつき音像生成法を統合し，新たに音声伝送サーバを Mac OS 上に実装し，その性能評価を行った．電子メトロノーム音を入力とした観測実験の結果，1 台のサーバを稼働させた場合でも 2 台のサーバを接続稼働させた場合でも，処理遅延量は想定通りにほぼ 20ms から 30ms の間に収まっていた． また，6 機のスピーカの音出力間の同期はサーバ起動から 30 分経過した時点においても高い精度で確保され，各スピーカの出力音の振幅制御も正確に行われていることを確認できた

なお今回は，複数音源入力（複数音像生成）を原理的に可能とする実装環境，Mac OS 及び CoreAudio の環境を用いたものの，サーバの実装は単一音源の音入力をのみ仮定して行った．今後，複数音源の入力に拡大する必要がある.

参考文献

[1] K. Hirata, Y. Harada, T. Takada, S. Aoyagi, Y. Shirai, N. Yamashita, and J. Yamato; The t-Room: Toward the Future Phone, NTT Technical Review, vol.4, no.12, pp.26-33 (2006).
[2] 柴田; 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] D. Stuckel and C. Gutwin; The Effects of locallag on Tightly-Coupled Interaction in Distributed Groupware: Computer Supported Cooperative Work, pp.447-456 (2008).
[4] 大島; 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] G. Pablo Nava, K. Hirata, M. Miyoshi; A loudspeaker design for sound image localization on large flat screens, Acoust. Sci. and Tech., vol. 31, no. 4, pp.278-287 (2010).
[6] 中谷; 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).
[7] Y. Makita; On the Directional Localization of Sound in the Stereophonic Sound Field, EBU Review no.73-A, pp.102-108 (1962).