Vol.2017-MUS-114 No.15
2017/2/27

WebRTC を用いた DAW 用遠隔指導支援システムの開発

野原祐一[†1]　辻靖彦[†2]

放送大学 大学院文化科学研究科　〒261-8516 千葉市美浜区若葉 2-11

E-Mail:　†1 y_nohara@nifty.com,　†2 tsuji@ouj.ac.jp

概要：近年，音楽制作の殆どがコンピュータ上で行われ，DAW (Digital Audio Workstation) の習得が必須となっている．DAW を遠隔で指導する場合，音響が，ステレオ通信ができないことや，周波数特性が不十分なことなどの理由から，現状のテレビ会議などでは満足にできない可能性が考えられる．そこで本研究では，WebRTC を用いて音楽制作，特に音楽表現の指導を考慮した遠隔指導支援システムを開発した．

音質の評価として，周波数特性，空間，レイテンシ，リップシンクにおける数値やグラフの測定に加えて，リアルタイム化した音響を用いて 2 名の講師のインタビューによる評価を行った．その結果，リアルタイム化した音響は，講師が問題把握や指導ができる品質である可能性が示された．反面，音響のレイテンシは，160ms 以上と遠隔指導の利用は可能であるが，合奏などにおけるリアルタイムの利用は難しいとの評価を得た．

キーワード：音楽制作，遠隔教育，表現指導，WebRTC, DAW

1. はじめに

1.1 研究背景と目的

近年，音楽制作の殆どが，コンピュータ上で行われ，DAW（Digital Audio Workstation）の習得が必須となっている．DAW の遠隔指導の先行研究において，非同期型の事例は存在するものの，同期型では殆ど行われていない．

大学における同期型の一般的な遠隔指導の事例として，「放送大学現代 GP プロジェクト」[1]がある．Web カンファレンスシステムを使用した同期型授業を実施し学生の主観評価により，特にゼミ・演習・購読などの形式の授業における有効可能性を示している．また，カメラ・マイクを通した授業であっても，ある程度の人数であれば遠隔側の学生を個別に十分把握でき，ホワイトボードや配布資料など映像・音声以外の情報を併用することで，多様な内容を扱うことができるとしている．

楽器の同期型の遠隔指導では，ピアノ，ドラム[2]，ヴァイオリン[3]および金管楽器[4]における研究実践が行われている．実践の結果，触覚，息遣い，カメラアングルの問題点が指摘されており，対面とは指導方法を変える必要性が指摘されている．また，ピアノとドラムの遠隔指導を実践した斎藤 [2] によれば，ドラム指導における映像と音響のズレ（リップシンク）が遠隔指導を困難にする可能性や，ジャズピアノにおいては指導回数が進むにつれて指導者の声が重要になる点を指摘している．

一方，入江ほか [5] は,音楽セッションを目的とした遠隔合奏支援システムを開発しており，遠隔における CD 相当品質での音響の共有を実現している．

以上の背景を踏まえて本研究では，遠隔地からでも対面に近い形で DAW の指導が可能なシステムの実現を目的とする．本報告では始めに，既存のツールにおける DAW の遠隔指導の実現可能性を検討するために音質の予備調査を行った．この結果を踏まえ，ブラウザ間のリアルタイムコミュニケーションを可能とする WebRTC に着目し，WebRTC の API を用いて DAW 用の遠隔指導支援システムを開発し，評価を行った．

1.2 WebRTC

WebRTC[6]は，W3C（*https://www.w3.org*）が提唱するリアルタイムコミュニケーション用の API の定義である．2016 年 11 月末時点で，まだ，Working Draft の状況である．具体的な機能は，プラグイン無しでボイスやビデオのチャット（Media Stream）や，テキストやバイナリのデータ送受信（RTC Data Channel）ができる．また，Web ブラウザ間のピアツーピア通信であり，効率的，高速，かつ暗号化により安全な通信ができる．ただし，ピアツーピア通信といえども，プロトコル整合（シグナリング）や NAT 越えなどにサーバを必要とする．また，アプリケーションや Web ページなどを格納する Web サーバも必要である．なお，開発言語は JavaScript である．

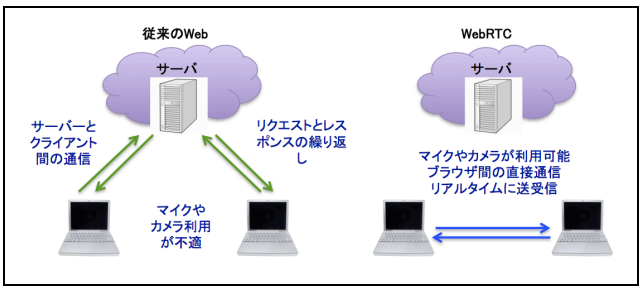

図 1　従来の Web と WebRTC の違い

1.3 学習コース

本研究が対象とする遠隔指導の学習内容を整理するため

に，公開されている MIDI 検定試験対策などの一般的なものを選択し，その指導や学習内容を検討した．ポイントをまとめたものを表 1 に示す．

表 1　学習コースの分析

学習コース	形態	DAW	音楽表現	音響	音響同期（推測）
MIDI 検定 3 級[7]	講義	不要	対象外	不要	―
MIDI 検定 2 級 1 次[7]	講義	不要	対象外	不要	―
MIDI 検定 2 級 2 次[7]	演習	必要	対象外	不要	―
MIDI 検定 1 級[7]	演習	必要	対象	必要	望ましい
ベーシック[8]	演習	必要	対象	必要	望ましい
アドバンス[8]	演習	必要	対象	必要	望ましい

音楽表現の指導や学習を行うには，音響が必要であり，その音響はリアルタイムであることが望ましいと推測できる（赤色枠）．ゆえに，MIDI 検定 1 級，ベーシック，アドバンスの 3 つのコースの遠隔指導を主として検討する（黄色の部分）．逆に講義や操作指導などの音楽表現を必要としない指導では，一般的なテレビ会議システムでも十分に対応できると考えられる．

なお，MIDI 検定関連を指導するには，（一社）音楽電子事業協会の認定指導者の資格が必要である．ただし，1 級は認定指導者の制度がない．また，ベーシック，アドバンスの両コースの指導には，ローランド（株）の講師資格が必要である．

1.4 予備実験

テレビ会議システムが DAW の指導に活用できるか否かを判断するために，一般的によく利用されている「Skype」と「TeamViewer」を使って音質を調査した．具体的には 2 台の PC 間で CD 品質相当の音響を，それらを使い送受信，録音して比較した．

表 2　予備実験の結果

比較項目	音源	Skype	TeamViewer
周波数特性	～21kHz	～11kHz	～7kHz
空間	ステレオ	モノラル	モノラル

実験結果を表 2 に示す．音源は 21kHz 付近まで音の成分があるにもかかわらず，Skype では 11kHz 付近，TeamViewer では 7kHz 付近までの音の成分のみとなり，周波数特性が悪くしている．また，ステレオからモノラルに変化させ，空間がわからなくなっていることも伺える．つまり，音響を必要とする DAW の遠隔指導には，これらのアプリケーションの適用では困難であることがわかった．

2. 研究方法

2.1 研究手順

音響を双方向にリアルタイム化した DAW の遠隔指導を実現するには，新たにシステム開発が必要であることが予備実験により明らかとなった．本章では，本研究で提案する遠隔指導支援システムの開発方針を示す．

なお，開発手法には プロトタイピングモデルを採用する．最後に周波数特性，空間，レイテンシ，リップシンクの測定評価および第三者の講師のインタビューによる評価を実施する．

2.2 開発方針

(1) WebRTC の採用

本システムの開発において WebRTC を採用することとした．表 3 に，Chrome と FireFox の 2016 年 11 月時点の WebRTC と関連する API の対応状況を示す．なお，表中の数字は対応バージョンを示す．

表 3　Web ブラウザの対応状況

API	Chrome	FireFox
WebRTC	23	22
Media Capture and Streams	21	17
Web Audio API	10	25
Web MIDI API	43	Extension
Media Stream Recording	49	29
Audio Output Devices API	49	-
Screen Capture	Extension	Extension

なお，本システムでは Chrome のみを対応ブラウザとして開発する．

セキュリティの問題より Media Capture and Streams は，Chrome 47 から，HTTPS ONLY（localhost を除く）となる．また，Web MIDI API においても SYSEX の MIDI メッセージの通信を行う場合も，HTTPS ONLY である．そのため，Web サーバは HTTPS 対応が必須である．

(2) SkyWay の採用

SkyWay[9]とは，プロトコル整合や NAT 越えなどの機能を提供するクラウドで，現在，NTT コミュニケーション（株）が WebRTC アプリケーションの開発者向けに無料で提供している．このクラウドを利用することにより，サーバを用意することなく，開発者はアプリケーションのプログラミングに集中できることになる．

SkyWay では，PeerJS[10]のフレームワークを使ってシグナリングを行っているが，本家よりも品質改善や機能強化が図られていることや，ドキュメントが日本語対応されていることも大きなメリットである．

セキュリティの問題より，Chrome 34 より画面共有（Screen Capture）の機能提供が停止したため，SkyWay ScreenShare Library[11]（Extension）で対応する．

(3) 音響の CD 品質，ステレオでの通信の実現

標準の PeerJS（SkyWay も同様）では，音響のステレオ通

信ができない．これは，自動生成される SDP（Session Description Protocol）に，ステレオでの送受信指定がされないことによる．そのため，自動生成される SDP に，ステレオでの送受信指定を追加するカスタマイズを施す．

また，CD 品質相当にするには，MIT Codec の Opus で，48kHz サンプリング，128kbps の情報速度，ステレオ送受信指定で実現できる．

(4) マルチトラック通信の実現

1 つのセッションで映像や音声で複数のストリームを扱う必要がある．ただし，現バージョンでは，センション接続中のストリームの追加や削除までは対応していない．

(5) 指導や学習の記録

テキスト，音響，必要に応じて映像付きで，指導や学習を記録する機能を設け，分析や評価に生かすデータを残すことができるようにする．本来，すべてのストリームを記録の対象とすべきであるが，マルチトラックレコーディングが簡単にできないことや，映像の編集や変換などに多くの CPU 能力を必要とする．ゆえに，相互の音響と音声のみの方向とする．また，再生や編集は，本システムでは対応せず，外部のアプリケーションで行うことにする．

2.3 開発環境

開発における主な環境を表 4 に示す．

表 4　主な開発環境

項目	仕様
iMac	CPU: Core-i3 3.06GB，メモリ: 12GB，OS: MacOSX 10.XX（最新版）
Windows PC	CPU: Core-i3 266GB，メモリ: 6GB，OS: Windows7 Professional 64bit
ネットワーク	1000Base-TX
その他	開発ツールは，Chrome のデベロッパーツールを利用． Web サーバは，MacOSX に標準添付の Apache と OpenSSL で構築．

3. システム開発

3.1 システム構成，機能要件

図 2 にシステム構成を示す．

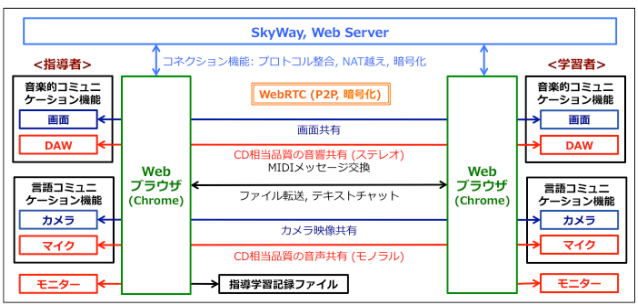

図 2　システム構成

青色の部分は，サーバとの通信である．SkyWay を介し，Peer と接続するための機能，Web サーバからアプリケーションを受け取る機能などである．また，サーバ間との通信は暗号化する．

2 つの緑色の Web ブラウザ間が WebRTC の通信である．紺色が映像，赤色が音響や音声，黒色がデータと分類している．この通信も暗号化する．また，音楽に関連する通信を「音楽的コミュニケーション」，会話に関連する通信を「言語コミュニケーション」とした分類も行う．

これらの機能および仕様を表 5 に整理する．

表 5　主な機能と仕様

音楽的コミュニケーション機能	
画面共有	V8 or V9: 960×540px
音響共有	Opus: 48kHz, 128kbps, Stereo
データ共有	MIDI メッセージ交換
言語コミュニケーション機能	
カメラ共有	V8 or V9: 640×480px
音声共有	Opus: 48kHz, 128kbps, Mono
データ共有	テキストチャット，ファイル転送
コミュニケーションの記録機能	
指導や学習の記録	テキストのファイル出力 音響や音声の録音とファイル出力

3.2 画面イメージ

図 3 に画面イメージを示す．

図 3　画面イメージ

4. 評価

4.1 システム評価

音楽的コミュニケーション，言語コミュニケーションの各々の性能や品質を数値で測定する評価である．測定項目は，周波数特性，空間，レイテンシ，リップシンクとした．

4.1.1 評価基準

レイテンシでは，西堀ほか [12]によると，演奏における遅延では，検知眼 30ms，許容眼 50ms である．この値を，音楽的コミュニケーションに適用する．また，ITU-T 勧告 G.114 でのエンドツーエンド遅延の範囲[13]を，表 6 に示す．この値を，言語コミュニケーションに適用する．

表 6　ITU-T 勧告 G.144 のレイテンシ値

項　目	レイテンシ
通話に際し利用者が許容できる値	150ms 以内
事業者が提供する役務として許容できる値	400ms 以内
一般的なネットワークの品質として許容できない値	400ms 以上

リップシンクでは，赤井田ほかのアナウンスと打楽器(クラベス) の 2 種類のリップシンクずれの先行研究[14]があり，それを，表 7 に示す．

表 7　アナウンスと打楽器（クラベス）のリップシンク値

リップシンク		アナウンス	打楽器 (クラベス)
検知眼	音進み	46ms	23ms
	音遅れ	122ms	56ms
許容眼	音進み	78ms	56ms
	音遅れ	182ms	130ms

アナウンスは，言語コミュニケーションに，打楽器（クラベス）は，音楽的コミュニケーションに適用する．

4.1.2 評価結果

(1) **webrtc-internals**

webrtc-internals とは，セッションのイベント，シグナリングの経過，送受信中の統計データを確認できる WebRTC の機能である．また，送受信している映像の Frame Size, Frame Rate も確認することも含まれる．

そこで，各種データを採取する前に，帯域ごとに送信するスクリーンキャプチャーした映像の Frame Size, Frame Rate を採取し，表 8 にまとめた．また，できるだけ CPU への負荷をかけないように配慮した．まず，帯域制限をかけることにより影響を受けるのは Frame Rate であることがわかる．つまり，キャプチャーの Frame Rate が落ちる前の帯域制限値が，快適に利用できる最小値といえる．この場合，1Mbps 以上が利用できることになる．なお，音響は帯域制限に影響を受けていない．

表 8　帯域制限と Frame Size と Frame Rate の変化

帯域制限（送信側）	Screen Capture Stream		
	Width	Height	Frame Rate
Programing	960px	540px	30fps
Capture	960px	540px	30fps
1Gbps	960px	540px	30fps
8Mbps	960px	540px	30fps
4Mbps	960px	540px	30fps
2Mbps	960px	540px	30fps
1Mbps	960px	540px	30fps
512kbps	960px	540px	10～20fps
256kbps	960px	540px	0～10fps

(2) **音楽的コミュニケーション**

周波数特性では，図 4～5 に示すとおり，20kHz 以上の音の成分がある音源が，20kHz 程度までは送信できている．また，図 6 に示すとおり，空間では音の成分分布が，L から R に分散しており，ステレオで出力されていることがわかる．

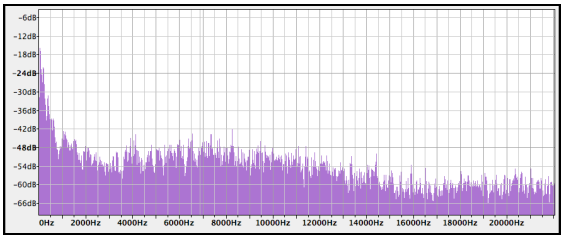

図 4　音源のスペクトラム

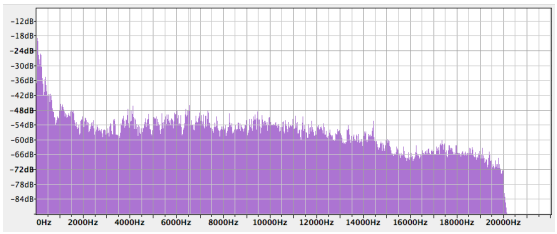

図 5　送受信後の音響のスペクトラム

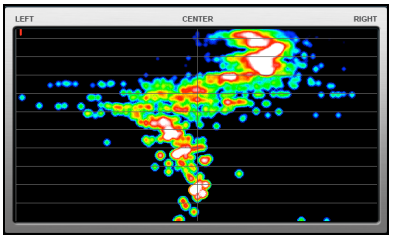

図 6　送受信後の音響の空間

表 9 にレイテンシとリップシンクの測定結果を示す．

音響のレイテンシは 160ms 以上と大きく，許容眼(50ms)を満たせない．映像との同期のため，音響を遅らせている可能性が考えられる．なお，実際の指導では学習者の制作した音響を聴きながら問題把握や指導を行う為，一般的には合奏に求められるレベル（検知眼:30ms，許容眼:50ms）を満たさなくても運用できる．

MIDI のレイテンシは，合奏に求められるレベルを満たしていた．8Mbps の結果が悪いのは，測定時に何だかの負荷が掛かった可能性が考えられる．負荷がなければ，22ms 程度の結果が出たと推定できる．

リップシンクでは， SONAR の譜面表示を利用した．縦線はナビゲータと考え，発音から音符が赤に変わるタイミングで認知すると仮定して測定した．その場合，概ね検知眼 （23ms），許容眼（56ms）を満たしている．なお，1Gbps は発音から音符が赤に変わるまでは許容眼を満たしていないが，縦線が音符に来てから発音までの許容眼（130ms）を

満たしてはいる結果となった.

表 9　レイテンシ，リップシンクの測定結果

帯域制限	レイテンシ（音響）	レイテンシ（MIDI）	リップシンク
1Gbps	160ms	21ms	79ms
8Mbps	185ms	27ms	27ms
4Mbps	203ms	22ms	15ms
2Mbps	204ms	24ms	▲6ms
1Mbps	220ms	31ms	56ms

音響も MIDI も 1Mbps 未満では，テンポやリズムに狂いが生じ，測定を断念した．また，リップシンクではフレーム落ちが多くなり測定できなかった.

(3) 言語コミュニケーション

図 7～8 に示すとおり，周波数特性では，8kHz, 16kHz あたりの音の成分に若干の変化はあるものの，概ね同様のスペクトラムであった．また，図 9 に示すとおり，空間では音の成分分布が Center に集中しており，ステレオ からモノラルに変わっていることがわかる.

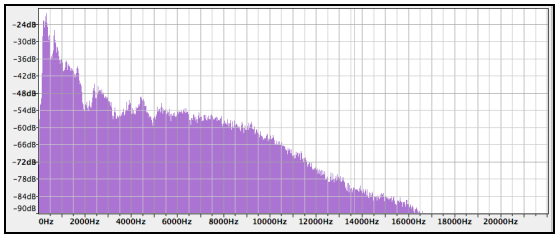

図 7　音声のスペクトラム

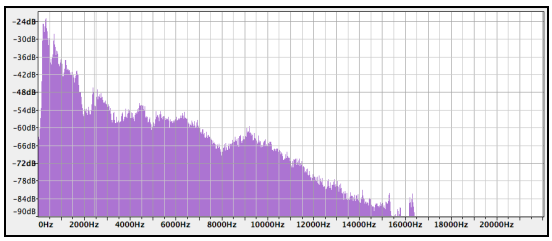

図 8　送受信後の音声のスペクトラム

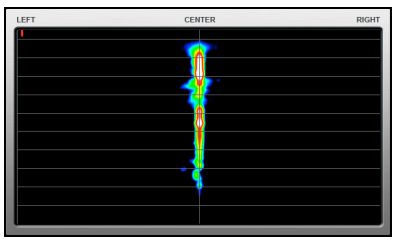

図 9　送受信後の音声の空間

表 10 にレイテンシとリップシンクの測定結果を示す.

レイテンシは，検知眼（150ms）を満たせないが，許容眼（400ms）は満たしている．また，音響に比べ全体的に値が大きい．なお，音声は問題なく聞き取れる.

リップシンクでは，本来は，音声で評価すべきであるが，今回はより厳しい楽器を対象とした．それは，楽器の遠隔指導の参考として試したいことによる．鍵盤を抑えてから発音までは，概ね検知眼（58ms），許容眼（130ms）を満たしている．なお，2Mbps では，アナウンスでの許容眼（182ms）は満たしているので会話では問題はないと考えられる.

表 10　レイテンシ，リップシンクの測定結果

帯域制限	レイテンシ	リップシンク
1Gbps	210ms	0ms
8Mbps	212ms	55ms
4Mbps	280ms	126ms
2Mbps	356ms	155ms

なお, 1Mbps 以下では，フレーム落ちが多く，測定を断念した.

(4) コミュニケーションの記録機能

周波数特性，空間のグラフより，音響共有や音声共有と同等の品質で録音ができたと考えられる.

4.2 第三者講師による評価

4.2.1 目的と方法

前節で CD 相当のステレオの音響であることは示した．実際の講師が,「遠隔からの学習者の音響より問題把握ができるか」を調べるために，第三者講師による評価を行った.

評価手順を示す．本システムを用いて学習者の再生をコミュニケーションの記録機能を使い映像ファイル化し，それを観てもらった後にヒアリングを実施した．学習者の音源には, MIDI で制作した簡単なジャズとボサノバを用いた．そして，第三者の講師には，音楽制作の指導経験を有する 2 名の講師に依頼し，メールや電話によりヒアリングを実施した.

4.2.2 結果と考察

以下に，得られた結果と考察をまとめる.

(1) 音響品質が学習者の問題把握や指導ができるか

両講師とも，音響については，問題把握するには問題のない品質であるとの回答であった．これより，対面同様の音響で遠隔指導ができる可能性が示された.そのことに加えて両講師は，今回の音源に入っていなかったヴァイオリンなどの弦楽器や，フルートやオーボエなどの木管楽器があるとより指導可能性を判断しやすいと指摘している.この点は今後の課題と考えられる.

なお，1 名の講師は，音楽表現を遠隔から指導するには，リアルタイムなコミュニケーションが必要であると述べた.

ただし，本人は，会話のリアルタイム化は求めているが，音響のリアルタイム化の必要性までは言及していない．この点を再度質問したが，明快な回答が得られなかった．

(2) 本システム中のテレビ会議機能の評価

両講師とも，Skype などの既存の Web 会議システムの方がよいとの回答であった．実運用を行うに辺り操作性，デザインの向上が今後の課題と考えられる．

(3) カメラの必要性

1 名の講師は，カメラがなく，音声のみでも指導は可能と回答した．これは，実際の指導では，音響や音声が重要であり，相手の顔を見なくても問題ないということになる．しかし，斎藤の先行研究[4]では，「回数が進むにつれ，指導者のピアノの音響や演奏映像の重要性が下がり，逆に指導者の表情や音声が重要になっている．特に，指導者の音声が一番重要となっている」と指摘している．この結果を比較すると「音声が重要である」点に違いはないが，「指導者の表情が重要になってくる」点については反する結果となった．

指導者と学習者の立場の違いから発生している可能性があり，学習者目線での評価を行う必要があると考えられる．

5. まとめ

本研究では， DAW の遠隔における指導を実現するために，WebRTC を用いた遠隔指導支援システムを開発した．周波数特性，空間，レイテンシ，リップシンクの測定結果によるシステム評価を行った結果，リアルタイムの合奏を行うには困難な面もあるものの，DAW の遠隔指導において本システムは安定して利用可能であることがわかった．さらに，2 名の第三者講師によるインタビュー評価を行った結果，リアルタイム化した音響は，講師が問題把握し指導できる品質であり，本システムを用いることで対面同様に問題把握し指導できると回答が得られた．

今後の課題として，以下が考えられる．

・本システムの実運用を踏まえた操作性の向上
・本システムを用いた DAW の遠隔指導実践による学習効果の検証

謝辞 第三者講師による評価にご協力頂いた皆様，また，合同ゼミなどでご意見を賜った皆様に，謹んで感謝の意を表する．

参考文献

[1] “放送大学現代 GP プロジェクト”. http://u-air.net/GP/index.html, (参照 2016-11-30).
[2] 加納暁子, “遠隔教育における器楽指導の実践と課題について”,教育実践総合センター紀要 7,2008.3,pp.211‐18
[3] 千葉圭説, “インターネットを利用した遠隔地との管楽器レッスン”,北翔大学生涯学習システム学部研究紀要 14,2014,pp.99‐103
[4] 斎藤忠彦. 遠隔演奏システムを活用した音楽教育のデザインと今後の方向性-試行的な実践を通して. 信州大学教育学部研究論集, No.1, 2009.7, pp117–26.
[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. 研究報告マルチメディア通信と分散処理(DP5), No.23, 2009.11, pp1–8.
[6] “WebRTC 1.0: Real-time Communication Between Browsers”. http://www.w3.org/TR/webrtc/, (参照 2016-11-30).
[7] 音楽電子事業協会, 日本シンセサイザープログラマー協会. ミュージッククリエイターハンドブック: MIDI 検定公式ガイド. ヤマハミュージックメディア, 2012.
[8] “コンピュータミュージック科”. http://www.roland.co.jp/school/course/enjoy/cm/, (参照 2016-11-30).
[9] “SkyWay”. http://nttcom.github.io/skyway/, (参照 2016-11-30).
[10] “PeerJS: Simple peer-to-peer with WebRTC”. https://github.com/nttcom/peerjs, (参照 2016-11-30).
[11] “SkyWay ScreenShare Library”. https://github.com/nttcom/SkyWay-ScreenShare, (参照 2016-11-30).
[12] 西堀佑, 多田幸生, 曽根卓朗. 遅延のある演奏系での遅延の認知に関する実験とその考察. 情報処理学会研究報告[音楽情報科学], No.127, 2003.12, pp37-42.
[13] ITU, One-way transmission time. ITU-T Recommendation G.114, 2003.5.
[14] 赤井田卓郎, 黒住幸一, 岡田清孝, 林俊一, 深谷崇史. リップシンク～映像と音声のタイミング～. NHK 技研だより, No.88, 1997.5, pp11-18.