

# リアルタイム遠隔セッションにおける 映像情報の必要性と遅延要件に関する検討

福井 達也<sup>1,a)</sup> 藤原 稔久<sup>1</sup> 椎名 亮太<sup>1</sup> 小野 央也<sup>1</sup> 谷口 友宏<sup>1</sup>

**概要:** 新型コロナウイルスの拡大により、ネットワークを介して音声や映像をやり取りし、演奏者が同じ場所に集まることなくリアルタイムにセッションを行う「遠隔合奏」の実現が求められている。本稿では、遠隔合奏における映像情報の必要性と遅延要件を明らかにすることを目的として、その評価方法を具体化するとともに、複数の演奏家による遠隔合奏実験を行った。実験によって得られた主観評価結果から、弦楽アンサンブルでは、遠隔合奏に低遅延映像情報があることでタイミング同期がしやすくなることと、その映像遅延要件が最小で 20~35ms であることを明らかにした。

**キーワード:** リアルタイム遠隔セッション、遠隔合奏、映像遅延、低遅延、主観評価

## Study on the Necessity and Delay Requirements of Video Information for Real-time Remote Sessions

TATSUYA FUKUI<sup>1,a)</sup> TOSHIHITO FUJIWARA<sup>1</sup> RYOTA SHIINA<sup>1</sup> HIROYA ONO<sup>1</sup> TOMOHIRO TANIGUCHI<sup>1</sup>

**Abstract:** Due to the spread of the new coronavirus, there is a growing need for real-time remote sessions over networks that eliminates the need for performers to gather in the same location. The purpose of this report is to clarify the necessity and delay requirements of video information in remote ensembles. In addition to specifying an evaluation method for this purpose, we experiment with remote ensembles of music performers. In string ensembles, the subjective evaluation results indicate that low-latency video information facilitates timing synchronization and the minimum video latency requirement is 20-35 ms.

**Keywords:** real-time remote sessions, remote ensembles, video delay, low-latency, subjective evaluation

### 1. はじめに

新型コロナウイルスの感染拡大により対面で人と人とが集まる活動が制限されたことから、音楽家による合奏や演奏活動、コンサート等の開催や実施が非常に難しいものとなった [1]。一部の音楽家によって個々に録音した曲をミキシングして曲を作る「オンライン合奏」や Zoom のような Web 会議サービスを利用してレッスンをを行う「遠隔レッスン」といった新しい音楽活動形態が模索されている。このような体験は音楽を楽しむ一つの形態として非常に有意

義であるものの、オーケストラや吹奏楽を嗜む学生やアマチュア音楽家にとっては、同時に複数人が演奏して一つの曲を作り上げる合奏こそが最も重要な体験であると思われる。その観点から、ネットワークを介して音声や映像をやり取りし、演奏者が同じ場所に集まることなくリアルタイムにセッションを行う「遠隔合奏」の実現が求められている。

遠隔合奏は、学術的には Network Music Performance(NMP) という分野として、古くより研究されてきた [2]。遠隔合奏において最初の障壁となるのは、ネットワークを介して音声情報をやり取りする際に発生する音声遅延である。音声遅延が大きい場合、遅れて聞こえてくる共演者の音に相互に同期しようとするためテンポが遅く

<sup>1</sup> NTT アクセスサービスシステム研究所  
NTT Access Network Service Systems Laboratories,  
Musashino, Tokyo 180-8585, Japan  
<sup>a)</sup> tatsuya.fukui.km@hco.ntt.co.jp

なることが知られており、演奏者が現実の合奏と同じように演奏できる音声遅延は10~15ms程度とされている[3]。この値は各演奏者が居る地点間の物理的な距離やネットワークの状態によっては実現できない場合もあることから、NMPでは、音声遅延が合奏の質や演奏者の体感に及ぼす影響[4]や、音声遅延が大きいときに演奏者が取りうる戦略[5]等、多くの研究がなされてきた。

一方で、遠隔合奏における映像情報の扱いに関する検討は十分になされていない。合奏を構成する主要素は音声で、音声品質が遠隔合奏の体感を決定づける支配要因であることから、音声遅延の短縮や音質の向上が優先課題であったことがその理由と想定される。しかしながら、5Gをはじめとする広帯域かつ低遅延なネットワークサービスの普及やDante[6]のように低遅延なネットワークオーディオ規格の発展に伴い、原信号が数Mbps程度の音声信号であれば、一般的な環境においても非圧縮かつ低遅延に伝送することの現実性が高まっている[7]。そのため、今後、さらに遠隔合奏を発展させていくためには、映像情報の扱いに関して検討を深めていくことが必要と考えられる。

遠隔合奏の映像情報の扱いを検討していくにあたり、まず明らかにすべきことは映像情報の必要性である。そもそも、遠隔合奏に限らず現実の合奏であっても視覚がどの程度用いられているかについて不明瞭な点が多い。現実の合奏における演奏者間のコミュニケーションについてまとめた[8]では、合奏を円滑に進めるコミュニケーションと合奏における発音や表現を同期させるという2つの用途で視覚情報が利用されるとしているが、楽曲や楽器、状況によっては視覚情報はそれほど重要ではないという示唆がされている。そのため、遠隔合奏において映像情報が果たす役割やその効果を明らかにしていくことが求められる。

次に明らかにすべきことは遠隔合奏における映像情報の遅延要件である。原信号がGbpsオーダとなる映像情報を伝送する場合、遅延の発生要因はネットワークの遅延だけではない。確保できる通信帯域に応じて適用する圧縮方式によって発生するコーデック遅延や、通信環境の揺らぎを吸収するためのバッファで発生するジッタ遅延、カメラやモニタ等の映像デバイスの内部で発生する処理遅延等、多くの遅延発生要因が存在する[9]。遠隔合奏で利用する低遅延な映像伝送システムを構成する際のネットワークや機器の制約条件を明確化するために、映像の遅延要件を確認することは重要である。

以上の点を鑑み、本稿では、遠隔合奏における映像情報の必要性と遅延要件を明らかにすることを目的として、その評価方法を具体化するとともに、複数の演奏家による遠隔合奏実験を行う。得られた主観評価結果をもとに、遠隔合奏における映像情報の必要性と遅延要件に関して考察する。

## 2. 関連研究

### 2.1 遠隔合奏における映像情報に関する研究

NMPの研究動向については[2]が詳しいが、映像情報を直接の対象として分析した事例は少ない。そのため、主目的が映像でなくとも、映像に関して何らかの知見を得ている研究を紹介する。

2020年11月に行われたリモートコンサートイベントBiennale Tecnologiaでは、映像を用いての遠隔合奏が行われたが、一般的なネットワークとWeb会議アプリケーションを用いて映像伝送システムを構成したことにより、映像遅延は数100ms程度と非常に大きくなり、音声との同期が取れず、演奏者にとって映像は有害でさえあったと評されている[11]。3拠点を映像と音声で接続して合奏を行った結果を分析した[12]では、演奏者が映像を見るのが非常に少なかったという結果が示されている。その理由として、音声に対して映像遅延が大きかったこと、そもそも現実の合奏において演奏者同士が見合うことが無いこと、見ることも音を聴くのに集中していることが推測されている。指揮者と2人の歌手によるオペラの遠隔合奏を行い、主観評価結果と生体信号によって演奏者の体感品質を分析した報告[13]では、評価者が映像の重要性を高く認識しているという結果が示されており、遠隔合奏における映像の必要性が示唆されている。

上記のように、遠隔合奏における映像の必要性に関してはネガティブな結果が多いものの、利用されているシステムは映像遅延が大きいものばかりであり、映像遅延が十分に小さい状況でその必要性が評価されていない。

映像遅延の変化に関する影響を評価した実験としては、前述の[13]にて音声遅延を15ms~135ms、映像遅延を60~180msとそれぞれ制御できる環境を用いて演奏者の体感品質が分析されている。しかしながら、本報告は実験当時の通信環境を想定しているために、音声遅延を十分に小さく設定したまま映像遅延を変化させる条件が設定されておらず、演奏者間の遅延条件も統一されていないために、映像遅延要件を正しく導けるものとはなっていない。また、映像と音声を送受信する通信路に遅延を付加できるエミュレータを挟み込むことで、遠隔合奏における通信遅延影響を評価した報告[14]では、通信遅延が大きくなるにつれて合奏体感が劣化することが示されており、遠隔合奏が実現可能な通信遅延条件が明らかとなっている。しかしながら、本報告では映像と音声の遅延が同時に増加する実験環境を用いているため、合奏に直接影響のある音声遅延に起因した体感劣化が支配要因となっており、映像の遅延要件を切り分けて分析できていない。

以上のように、これまでの研究では、遠隔合奏を無理なく行える音声遅延環境を前提として映像遅延を変化させる

状態の評価が十分に行われていない。

## 2.2 合奏における視覚利用に関する研究

共演者と共に演奏を作り上げる合奏という形態では演奏者同士が高いレベルで協調するために様々な言語・非言語的な手がかりが使われており、なかでもジェスチャーやアイコンタクトのような視覚的な手がかりは非常に多く用いられているとされている [8]。合奏において明確に視覚情報を用いているのは指揮者であり、[15] では、音楽家が指揮者の指揮軌道が減速する瞬間から拍を認識していることが示されている。他方、小編成で指揮者を利用しないアンサンブルにおいても視覚情報を用いていることが示されている。フルートとクラリネットの身体動作について分析した [16] では、ソロで演奏しているときよりもデュオで演奏しているときの方が身体動作がよりインタラクティブで相互の協調が見られると報告されており、小編成のアンサンブルにおいて視覚情報を活用していることが推察される。しかし、向かい合わせにした2台のピアノの間に衝立を設置して視覚情報の有無を変えながらデュオを行い、身体動作や視覚が音楽に与える影響を分析した [17] では、ピアニストの身体の動きが同期しているほど2人の発音タイミングや表現が同期する傾向にあるものの、視覚情報についてはむしろ無い方が発音タイミングが同期することが報告されている。一方で、このような結果はテンポが一定の曲を使用しているためとも考えられている [8]。視覚を演奏同期に明確に用いていることを示す研究としては、弦楽四重奏における弓の動きをセンシングして合奏を分析した [18] がある。同研究では、フレーズの先頭における第一ヴァイオリンの弓の動きが後に続くテンポを提示していることが明らかにされている。ただし、同様の分析においてフレーズの途中では特徴的な動きはなく専ら聴覚情報によって音の同期を図っているとも報告されており、視覚情報の利用箇所は限定されていることも示されている。

以上のように、合奏において視覚は重要な情報源として利用されているものの、その利用頻度や演奏に与える影響は様々であり、合奏形態や曲目によっても異なることがわかる。

## 3. 遠隔合奏実験

遠隔合奏における映像の必要性和、その映像遅延要件を明らかにするために、音楽家による遠隔合奏実験を行う。本章では、本実験の環境や内容と評価方法を示す。実験に際しては、関連研究の動向を鑑み、音声遅延と映像遅延が十分に小さいシステムを構築し、映像遅延にフォーカスした評価を行っていくこととする。

### 3.1 評価者

運弓という全身動作のある弦楽では視覚をより高頻度に

用いているという関連研究の知見を踏まえ、本実験では検討対象をクラシックの弦楽アンサンブルに絞ることとしている。本実験には、バイオリン、ビオラ、チェロからなる弦楽三重奏9組の合計27名が評価者として参加した。評価者は、いずれも音楽大学に在学中または卒業して10年以内の専門家であり、10年以上の楽器演奏歴を有している。メンバー間のアンサンブル経験は様々であり、本評価の前に共演したことのあるメンバーもいれば、初対面となるメンバーも存在した。

### 3.2 曲

関連研究の知見を踏まえ、本実験では演奏する曲目を統一している。本実験で利用する曲はW.A. モーツァルト作曲 アイネ・クライネ・ナハトムジーク第一楽章とし、実験では冒頭から27小節目までの約90秒間を演奏することとした。これは、冒頭のユニゾンの音を同期させるために視覚が利用されることと、評価者にとって演奏経験が豊富な曲であるため高い精度で演奏できることが期待できるためである。本楽曲の編成は弦楽合奏であり、本来は第二バイオリンパートが存在するが、編曲は行わず、第一バイオリン、ビオラ、チェロのパート譜をそのまま演奏することとした。

### 3.3 演奏に関する指示

本実験では、予め演奏に関して以下の3点を指示した。

- 3人が同じ場所に集まっている状況を想定する
- 本番の演奏を行っている状況を想定する
- 声によるコミュニケーションを行わない

1点目は、遠隔合奏特有の工夫を防ぐための指示である。遠隔合奏における演奏者は、遅延が大きい場合、相手と演奏タイミングが同期せずとも意図的に無視して演奏する、といった特殊な戦略を採ることがある [5] が、今回は遅延要件を抽出する上で障害となりうるため、除外した。2点目と3点目は、視覚を利用した演奏を促すための指示である。合奏における演奏方略は練習と本番で異なり、コミュニケーション手段が限定される本番ではアイコンタクトなどの非言語的な手がかりの使用が増えるとされている [8]。映像の遅延要件より抽出しやすくするために、本指示を行った。ただし、映像の無い環境で演奏を行う実験のみ、冒頭で声による開始合図を許可している。

### 3.4 実験環境

図1に実験環境を示す。各防音室の映像は、カメラによって撮影された後にHDMIエンコーダにおいてIP信号に変換される。その後、NWエミュレータを経由した後にHDMIデコーダにおいてHDMI信号に戻された後、スプリッターによって分岐して他の防音室のモニタに表示される。防音室間でやり取りする映像の遅延は、経路中にある

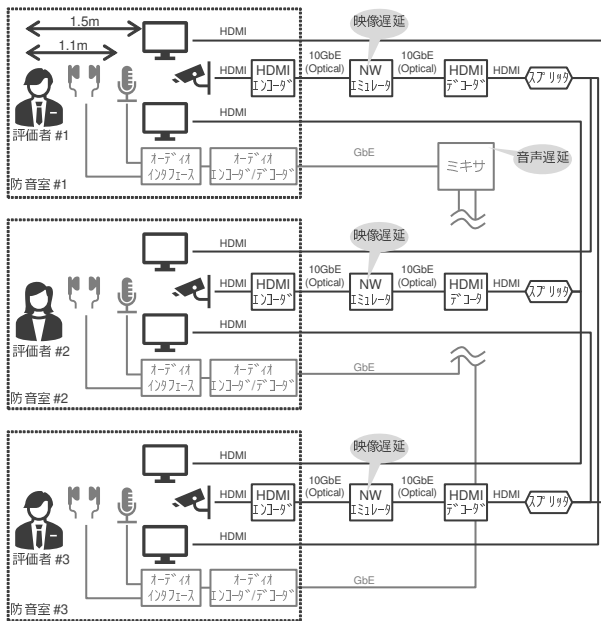


図 1 実験環境

Fig. 1 Experimental Environment

表 1 映像・音声パラメータ

Table 1 Parameter of video and audio

パラメータ	値
映像解像度	1920 × 1080
映像フレームレート	120 fps
映像コーデック	非圧縮
音声サンプリングレート	48 kHz
音声ビット深度	24 bit
音声コーデック	非圧縮

NW エミュレータで遅延を印可することにより任意の値に制御することが可能である。

各防音室の音声は、マイクによって集音された後にオーディオインターフェース装置を経由してオーディオエンコーダ/デコーダに入力されて IP 信号に変換される。その後、ミキサにおいて音声信号に戻された後、各部屋向けにミキシングして IP 信号に再変換し、オーディオエンコーダ/デコーダに送信される。再度に、オーディオエンコーダ/デコーダにて音声信号に戻された後、オーディオインターフェース装置を経由してイヤホンから出力される。なお、今回の評価では、イヤホンで出力されるのは他の部屋の音声のみであり、自分の演奏音のループバックは行っていない。防音室間でやり取りする音声の遅延は、ミキサで付加遅延を制御することにより任意の値に制御することが可能である。また、表 1 に本実験で用いた映像信号と音声信号のパラメータを示す。

本実験環境における遅延を付加していない状態の、デバイス両端における片道の映像遅延と音声遅延の測定結果を表 2 に示す。同表より、本実験環境は、映像遅延と音声遅延はいずれも遠隔合奏を行うのに十分な低遅延性を有して

表 2 実験環境の遅延性能

Table 2 One-Way Video and Audio Delays in the Experimental Environment

	最小値	最大値	平均値	中央値
映像遅延	1.8ms	9.8ms	5.97ms	6.0ms
音声遅延	8.6ms	9.6ms	9.14ms	9.2ms

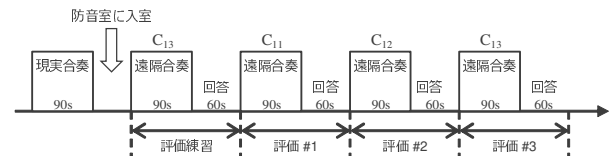


図 2 実験 1 の流れ

Fig. 2 Experiment 1

表 3 実験 1 の遅延設定

Table 3 Conditions of Experiment 1

#	映像遅延	音声遅延
C <sub>11</sub>	映像なし	9ms
C <sub>12</sub>	150ms	9ms
C <sub>13</sub>	6ms	9ms

いと判断できる。

### 3.5 実験の流れ

#### 3.5.1 準備

合奏の練習や環境への適応が不十分であると合奏中の演奏者のふるまいが正しく評価できない恐れがあることから、実験の前に準備を行っている。具体的には、3人1組の評価者は、物理的に集合した状態で音出しや今回演奏する曲の練習を行った。次に、評価者は防音室に入室し、音量や映像の画角の調整を行った後、評価環境を利用した遠隔合奏練習を実施した。

#### 3.5.2 実験 1：映像必要性評価

実験 1 の流れを図 2 に示す。実験 1 では、まず評価者は防音室の外で物理的に集合して曲 (1) を演奏し、現実の合奏の感覚を確認する。その後、評価者は防音室に入室し、練習を含む、各回毎に異なる環境で合計 4 回 (結果に採用しない練習 1 回を含む) の評価を行い、現実の合奏との違いを評価する。

実験 1 で利用した遅延設定を図 4 に示す。なお、環境 C<sub>11</sub> は、2 枚のモニタには何も描画されず音声だけが接続されている環境を示している。評価者には、同表に示す 3 種類の環境が存在すること、各評価で利用する環境は評価者にわからないようにランダムに設定することを予め伝えている。ただし、実験 1 は設定パターンが少なく各設定は評価者にとっても自明であることから、同図 2 のように実際は C<sub>13</sub> → C<sub>11</sub> → C<sub>12</sub> → C<sub>13</sub> という固定的な順番で実施している。

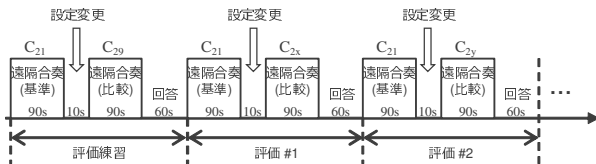


図 3 実験 2 の流れ

Fig. 3 Experiment 2

表 4 実験 2 の環境設定

Table 4 Conditions of Experiment 2

#	映像遅延	音声遅延
C <sub>20</sub>	6ms	9ms
C <sub>21</sub>	20ms	9ms
C <sub>22</sub>	35ms	9ms
C <sub>23</sub>	50ms	9ms
C <sub>24</sub>	80ms	9ms
C <sub>25</sub>	100ms	9ms
C <sub>26</sub>	20ms	20ms
C <sub>27</sub>	35ms	35ms
C <sub>28</sub>	6ms	35ms
C <sub>29</sub>	100ms	35ms

### 3.5.3 実験 2：映像遅延要件評価

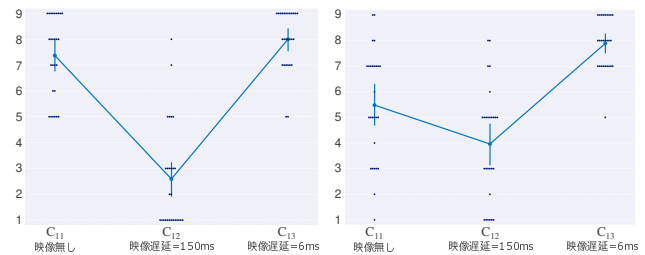
実験 2 の流れを図 2 に示す。実験 2 では、画質の主観評価で利用されている二重刺激妨害尺度法 (DSIS: Double Stimulus Impairment Scale Method)[19] を参考に、基準環境と比べて比較対象の環境でどの程度遅延を知覚したかについて評価することで、遅延要件の抽出を行う。具体的には、評価者は、基準となる最短遅延の環境 (映像遅延=6ms、音声遅延=9ms) で合奏を行った後に、遅延状態を変更して再度合奏し、2 回目の合奏において基準環境に対して知覚した遅延の程度を評価する。本評価は後述する 10 種類の環境を 1 回ずつ、合計 10 回 (結果に採用しない練習 1 回を含む) 実施する。なお、実験に際しては 5 回試行した後に休憩時間を確保している。

実験 2 で利用した環境を 4 に示す。実験に際しては、結果に採用しない最初の評価練習では C<sub>29</sub> を比較環境として必ず利用しているが、その後の 9 回の試行については C<sub>20</sub>~C<sub>28</sub> の 9 種類を、参加者の組ごとに異なるランダムな順番で 1 回ずつ利用している。留意すべき点として、基準環境 C<sub>20</sub> に対して、比較環境も C<sub>20</sub> であるケースが含まれていることが挙げられる。評価者には、大小さまざまな遅延設定の環境がランダムに提示されることと、基準環境とまったく変わらない環境が比較環境に含まれている点を予め伝えている。

## 4. 実験結果

### 4.1 実験 1

実験 1 の結果を図 4 に示す。同図において、プロットは各評価者の回答を示しており、実線は平均値、バーは 95% 信



(a) 映像と音声に遅延を感じまし (b) 現実の合奏と比べて、他者と演奏たか？ のタイミングは合わせやすかったですか？

図 4 実験 1 の主観評価結果

Fig. 4 Subjective Evaluation Result of Experiment 1

頼区間を示している。評価者は、文章で定義された 5 段階の評価に加えて、その間を選択可能な 9 段階のスケールで回答を行っており、最高値である 9 は、現実の合奏と同等と定義している。

図 4(a) は、現実の合奏と比較した遅延の知覚度合を評価した結果である。同図より、映像のない C<sub>11</sub> と比較して、映像遅延=150ms の C<sub>12</sub> の環境では遅延を強く感じていることがわかる。これは、低遅延な音声環境を準備したとしても、遅延の大きい映像があるとかえって遠隔合奏の体感を劣化させるということを示している。一方、映像遅延=6ms の C<sub>13</sub> の環境は、C<sub>11</sub> と比較して知覚する遅延が小さくなることがわかる。これは、低遅延な音声に加えて低遅延な映像を提供することで、遠隔合奏で知覚する遅延が低減されることを示している。さらに、C<sub>13</sub> における平均値は 8 であり、現実の合奏と比べてもほとんど遅延を感じていないことがわかる。これにより、映像遅延=6ms、音声遅延=9ms の環境は、遅延の観点では現実の合奏とほぼ同等の環境であるといえる。

図 4(b) は、低遅延な映像が加わることによる合奏体感の向上度合いを評価した結果である。同図より、映像のない C<sub>11</sub> の環境と比較して、映像遅延=150ms の C<sub>12</sub> の環境は演奏における同期をかえって困難にすることがわかる。一方で、映像遅延=6ms の C<sub>13</sub> の環境は、映像のない場合と比べて合奏における音の同期をしやすくすることが確認できる。

以上の結果より、遠隔合奏では、低遅延な音声に加えて低遅延な映像を提供することにより、音楽家の合奏体感を向上させることがわかった。

### 4.2 実験 2

実験 2 の結果を図 5 に示す。同図は、実験 1 と同様に 9 段階のスケールで得られた回答結果とその平均値を示している。最高値である 9 は、パラメータ変更前と同等と定義している。映像遅延=6ms の C<sub>20</sub> の結果は、評価者には設定が変更したように思わせながら、実際には C<sub>20</sub> のまま変

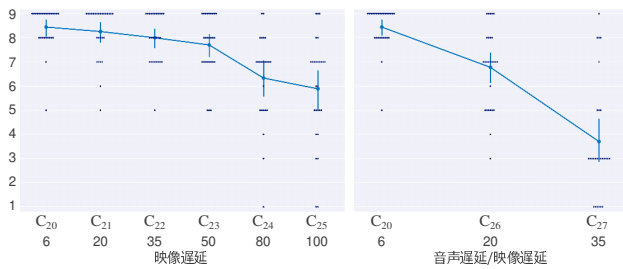


図5 実験2の主観評価結果  
(パラメータ変更前と比べて、映像と音声に遅延や違和感を感じましたか?)

Fig. 5 Subjective Evaluation Result of Experiment 2

更していない場合の試行の回答結果である。

同図より、10~20ms程度のわずかな時間であっても音楽家は遅延を知覚し、映像遅延が増加するにつれてその度合いも大きくなっていくことがわかる。また、本結果においてC<sub>20</sub>の結果と比較してt検定を行った結果、映像遅延=35ms以上の環境で統計的に有意( $p < 0.05$ )な差が得られている。以上より、今回の実験条件において遅延を知覚せずに遠隔合奏を行うための映像遅延要件は35ms以下となり、極めて小さな値になることが確認できた。また、同図より、映像遅延が80msを超えると、遠隔合奏の音声遅延要件を超過している音声遅延=20msの環境よりも強く遅延を知覚していることが確認できる。以上より、音声遅延=9msの状況下で遠隔合奏で映像情報が有用となるのは、映像遅延が80ms以下であることがわかった。

## 5. まとめ

本稿では、遠隔合奏における映像情報の必要性和遅延要件を明らかにすることを目的として、複数の演奏家による遠隔合奏実験を行った。実験によって得られた主観評価結果から、弦楽アンサンブルでは、遠隔合奏に低遅延映像情報があることにより合奏におけるタイミング同期がしやすくなることと、その遅延要件が最小で20~35msであることを明らかにした。

## 参考文献

[1] 山口恭正, “情報社会における新しい合奏活動の模索,” 情報処理学会研究報告, Vol. 2021-MUS-131, No.9, pp. 1-2, 2021.

[2] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, “An overview on networked music performance technologies,” IEEE Access, vol. 4, pp. 8823-8843, 2016.

[3] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan, “Effect of Time Delay on Ensemble Accuracy,” Proc. Int. Symp. Music. Acoust., vol. 2004, no. August 2009, pp. 3-6, 2004.

[4] Peter F. Driessen, Thomas E. Darcie, and Bipin Pillay, “The Effects of Network Delay on Tempo in Musical Performance,” Comput. Music Journal, vol. 35, pp. 76-89, 2011.

[5] A. Carôt and C. Werner, “Network music performance – problems, approaches and perspectives,” Proceedings

of the “Music in the Global Village” -Conference, vol. 162, pp. 23-10, 2007.

[6] Audinate, “Dante Controller User Guide,” April 2021, <https://dev.audinate.com/GA/dante-controller/userguide/pdf/latest/AUD-MAN-DanteController-4.5.x-v1.1.pdf>

[7] P. Ferguson, C. Chafe, and S. Gapp, “Trans-Europe express audio: Testing 1000 mile low-latency uncompressed audio between Edinburgh and Berlin using GPS-derived word clock, first with jacktrip then with dante,” 148th Audio Eng. Soc. Int. Conv., 2020.

[8] S. KAWASE, “合奏における演奏者間コミュニケーション-タイミング調整とその手がかり-,” Japanese Psychol. Rev., vol. 57, no. 4, pp. 495-510, 2014.

[9] G. Dylan Smith, Z. Moir, P. Ferguson, and G. Davies, “Low-latency Networked Music Collaborations: Does ‘Good Enough’ Do Enough Good?,” J. Netw. Music Arts J. Netw. Music Arts, vol. 2, no. 1, 2020.

[10] A. V. Elaine Chew, Alexander Sawchuk, Roger Zimmermann, Ilia Tosheff, Christos Kyriakakis, Christos Papadopoulos, Alexandre François, “Distributed Immersive Performance,” Natl. Assoc. Sch. Music, 2004.

[11] M. Bosi, A. Servetti, C. Chafe, and C. Rottondi, “Experiencing Remote Classical Music Performance Over Long Distance: A JackTrip Concert Between Two Continents During the Pandemic,” J. Audio Eng. Soc., vol. 69, no. 12, pp. 934-945, 2021.

[12] F. Schroeder, A. B. Renaud, P. Rebelo, and F. Gualdas, “ADDRESSING THE NETWORK : PERFORMATIVE STRATEGIES FOR PLAYING APART,” Int. Comput. Music Conf., no. December 2013, pp. 133-140, 2007.

[13] A. Olmos et al., “Exploring the role of latency and orchestra placement on the networked performance of a distributed opera,” Proc. 12th Annu. Int. Work. Presence, pp. 1-9, 2009.

[14] S. Delle Monache et al., “Time is not on my side: Network latency, presence and performance in remote music interaction,” Proc. 22nd Colloq. Music Informatics, pp. 152-159, 2018.

[15] Geoff Luck, Petri Toiviainen, “Ensemble Musicians’ Synchronization With Conductors’ Gestures: An Automated Feature-Extraction Analysis,” Music Perception 1 December 2006., 24 (2): 189-200.

[16] J. W. Davidson, “Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies,” Psychol. Music, vol. 40, no. 5, pp. 595-633, 2012.

[17] P. E. Keller and M. Appel, “Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles,” Music Percept., vol. 28, no. 1, pp. 27-46, 2010.

[18] R. Timmers, S. Endo, A. Bradbury, and A. M. Wing, “Synchronization and leadership in string quartet performance: A case study of auditory and visual cues,” Front. Psychol., vol. 5, no. JUN, pp. 1-9, 2014.

[19] Recommendation ITU-R BT.500-13 “Methodology for the subjective assessment of the quality of television pictures”, January 2012