

## 同期型遠隔応援のためのメディア遅延要件に関する検討

椎名亮太<sup>1</sup> 藤原稔久<sup>1</sup> 福井達也<sup>1</sup> 小野央也<sup>1</sup> 谷口友宏<sup>1</sup>

**概要**：スポーツなどの試合観戦において、物理的に離れた遠隔拠点にいるユーザ間で応援を行う遠隔応援が新たな観戦スタイルとして提案されている。遠隔拠点間での一体となった同期型応援は、デバイスやネットワークを含めた音声/映像のメディア遅延の影響を大きく受ける。本研究では、End-to-Endにおけるメディア遅延が遠隔応援の体感に与える影響を実験的に評価した結果を報告する。

**キーワード**：同期型遠隔応援、リアルタイム遠隔拠点間通信、メディア遅延要件、体感品質評価

### A Study of Media Delay Requirements for Synchronous Remote Cheering Systems

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

**Abstract**: Remote cheering between users at physically distant locations has been proposed as a new style of watching sports games. Synchronized cheering between remote locations is greatly affected by media latency of audio/video including devices and networks. In this study, we experimentally evaluated the relationship between the end-to-end media delay and the quality of the experience of remote cheering.

**Keywords**: Synchronous remote cheering, Real-time communications, Media latency requirements, Quality of experience evaluation

#### 1. はじめに

近年、COVID-19の流行に伴い、ワークスタイルやイベント開催などの形態が多様化し、オンライン環境でのワークやイベントの開催が浸透しつつある。例として、スポーツ観戦についても、新たな観戦スタイルが提案されている[1,2]。これまで、現地で観戦するか、映像を通して視聴することが主流であった観戦スタイルから、遠隔地から視聴するユーザも参加可能な参加型の観戦スタイルが浸透しつつある。YAMAHAの提案する「Remote Cheerer powered by SoundUD」[1,2]では、遠隔地にいるユーザ同士がアプリケーションを介して繋がり、自身の声援を実際の会場のスピーカーに対して反映することができる。プロスポーツ競技では、会場において、観戦者同士が声やリズムを合わせ、同期的な応援を行う場面がしばしば見受けられる。上述の[1]では、実際の声援や拍手などに対応した疑似信号をユーザ装置から送信し、集約した信号を非同期で試合会場などにおいて流す仕組みになっている。したがって、実際の競技場で実施されているような同期的な応援を、会場の観客や、ネットワーク越しの観客同士で実現することは難しい。ネ

ットワーク越しに音声や映像を共有し、リアルタイムでの同期応援を行う場合、拠点間で生じるデバイスやネットワークにおける遅延を含んだEnd-to-Endでのメディア遅延が問題になる。

一方、メディア遅延に対して、体感品質を評価し、遅延要求を評価する検討は様々なアプリケーションで実施されてきた。例えば、[3]ではIPTVを対象とした、遅延に対するユーザの体感品質を評価し、遅延条件について論じている。また、[4]ではテレビ電話システムにおける音声遅延と映像遅延のそれぞれの観点で影響を主観評価によって評価している。このようなユーザの体感品質に伴う遅延要件は、対象とするアプリケーションに応じて異なる。また、同じアプリケーションであっても、扱う音声や映像、ユーザ行動等の条件によっても、要求条件は異なることが想定される。

そのため、本稿では、スポーツやイベント等での同期型遠隔応援アプリケーションを想定した、同期行動におけるメディア遅延の影響を評価した。同期型の応援をいくつかのモデルに分類し、分類したモデルに対し、ユーザ主観評価を行った結果を報告する。

<sup>1</sup> 日本電信電話株式会社  
Nippon Telegraph and Telephone Corporation

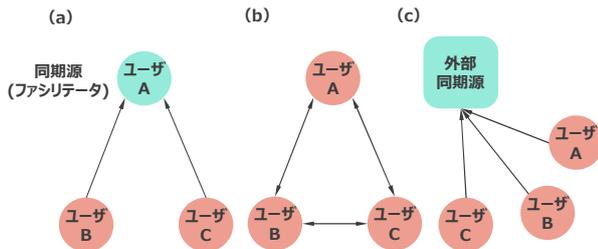


図 1 同期型遠隔応援におけるモデル

Figure 1 Models in Synchronous Remote Cheering.

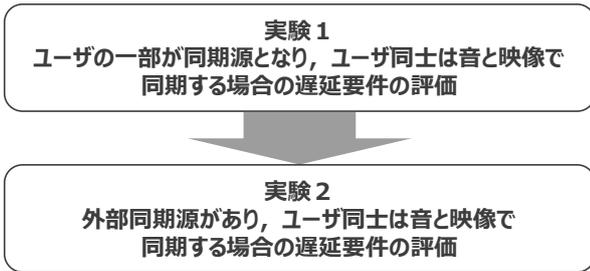


図 2 遅延要件の評価の流れ

Figure 2 Flow of Experiments to Evaluate Delay Requirements.

## 2. 主観評価実験

### 2.1 同期型応援の分類と実験内容

ユーザ同士が同期的に応援を行うケースは、いくつかのモデルに分類できる。図 1 には同期型遠隔応援におけるモデルを示す。同期型遠隔応援のモデルは、ユーザ内に同期源を持つ内部同期型と、ユーザ外に同期源を持つ外部同期型に分類される。前者の内部同期型は、さらに図 1(a)に示すような固定同期型と図 1(b)に示す自然同期型に分類され



図 3 実験に用いた応援方法

Figure 3 Style of Cheering Used for Experimental Evaluation.

る。固定同期型ではユーザの中に明確な固定の同期源ユーザ（ファシリテータ）を定め、その他ユーザはファシリテータに同期する。自然同期型では、ユーザ内で明確な同期源を定めることなく、ユーザ同士は、自分以外のどちらかのユーザに対して同期する。一方、外部同期型は、内部同期型のようにユーザの一部が同期源になるのではなく、ユーザの外部に同期源を設定し、映像や音声などの共通の外部同期源に対して、全ユーザが同期する。本稿では、内部同期型と外部同期型のモデルのうち、具体的な同期源が存在する固定同期型と外部同期型に着眼し、遅延要件の評価を行った。特に、固定同期型ではユーザの中にファシリテータが存在し、ファシリテータは応援のテンポやタイミングを調整することから、他ユーザに比べて遅延に対する感度が高くなると考えられる。そのため、固定同期型の評価では、ファシリテータとそれ以外のユーザのそれぞれの視点で評価を行う必要がある。また、外部同期型では、同期源に対して、ユーザ同士が音を発することなく映像のみで同期するのか、映像と音で同期するのかといったように、同期方法によってユーザ同士の遅延体感が異なると考えら

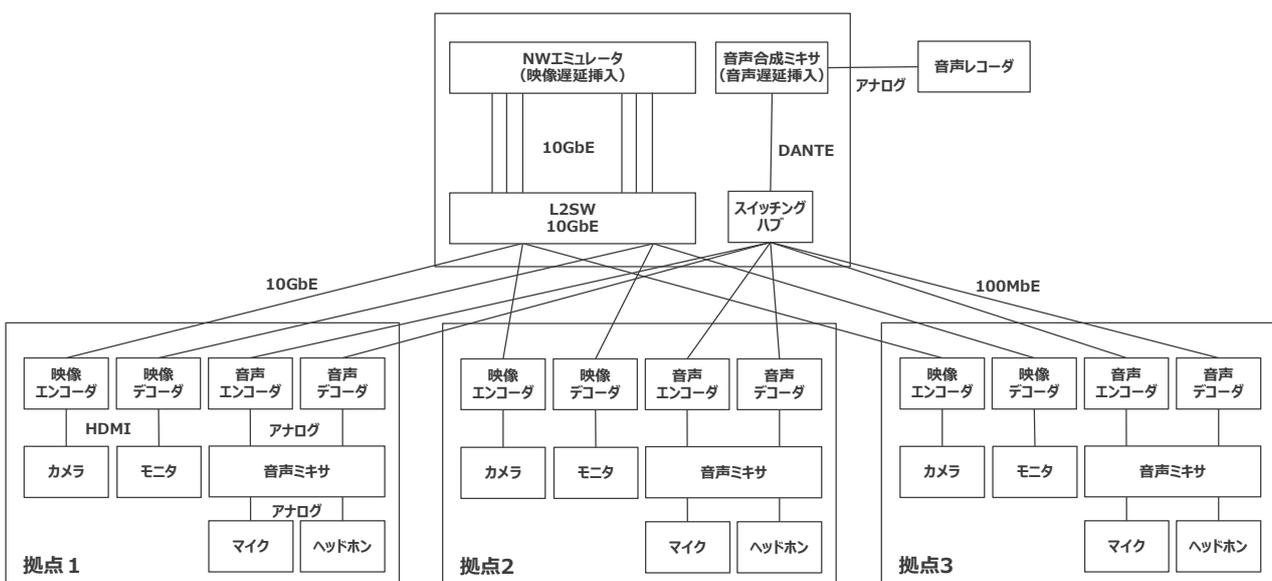


図 4 実験構成

Figure 4 Experimental Setup.



図 5 各ユーザ拠点 (防音室) の機器配備の様子  
Figure 5 Deployment of equipment at each user site.

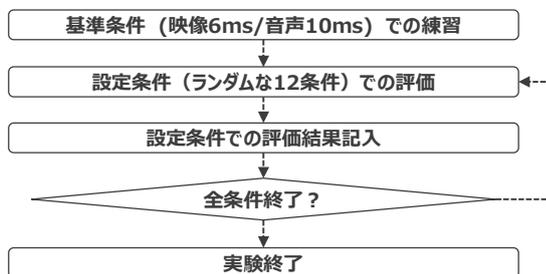


図 6 各実験の手順と流れ

Figure 6 Procedure and Flow of Each Experiment.

れる。

上記を踏まえ、図 2 に示すような二つの実験を実施した。実験 1 では、ユーザの一部が同期源となる固定同期型のモデルの評価を行う。ユーザの中からファシリテータを一人選択し、コール型の応援を実施する。図 3(a) に示すように、ファシリテータは自身のタイミングで「ニッポン」とコールし、その他のユーザはファシリテータのコールに合わせて、外部同期源の音楽に対して、各ユーザは音声と映像の両方によって同期する。図 3(b) に示すように、外部同期源の音声に合わせて、各ユーザはモニタに映る他のユーザを確認しながら応援治具を用いてリズムをたたく。

## 2.2 実験環境

図 4 に実験構成を示す。防音された 3 つの拠点をを用意し、それぞれの拠点には、自拠点映像を撮影するためのカメラ (OMRON SENTECH STC-HD853HDMI)、他拠点映像を表示するモニタ (IODATA LCD-GC271UXB)、そして自拠点音声の集音用のマイク (AKG P420)、他拠点の合成音声を出力するヘッドホンに配備した。ここで、カメラからの High-Definition Multimedia Interface (HDMI) の出力信号は、映像エンコーダ

		End-to-End 映像遅延 (ms)									
		6	20	25	30	40	50	75	100	125	150
End-to-End 音声遅延 (ms)	10	○		○			○	○	○	○	○
	20		○								
	30				○						
	40					○					
	50						○				
	100								○		

図 7 実験における End-to-End 遅延パラメータ

Figure 7 End-to-End Delay Parameters in the Experiment.



図 8 評価の様子と評価指標

Figure 8 Evaluation Status and Evaluation Indicators.

によって 10GbE 信号にフレーム化されたのち、集約拠点へと送信される。集約拠点では各拠点の映像信号を収集し、network (NW) エミュレータで遅延を付与したのち、各拠点へと映像を配信する。各拠点では、映像デコーダで HDMI 信号に変換されたのち、自拠点以外の映像がモニタに表示される。

一方、音声信号は、マイクによって集音され、音声ミキサを介して音声エンコーダに入力される。音声エンコーダではアナログ信号からデジタル信号に変換され、Dante プロトコルにより 100MbE のネットワークを介して集約拠点まで伝送される。集約拠点では音声合成ミキサにおいて、各拠点の音声信号が合成される。各拠点の音声信号、及び合成された音声信号は、音声レコーダを用いて録音した。合成された音声信号は、スイッチングハブを経由して各拠点に配信される。各拠点では音声デコーダにおいて、アナログ信号へと変換されたのち、音声ミキサを介してヘッドホンにて信号が出力される。ここで、ヘッドホンからの出力信号は、自拠点以外の拠点の音声合成した信号とした。また、開放型のヘッドホンを利用することで、自拠点の音声のループバックを不要とした。

図 5 には各拠点の機器配備の様子を示す。他拠点映像を表示するモニタを二つ図 5 に示すように配備し、自拠点以

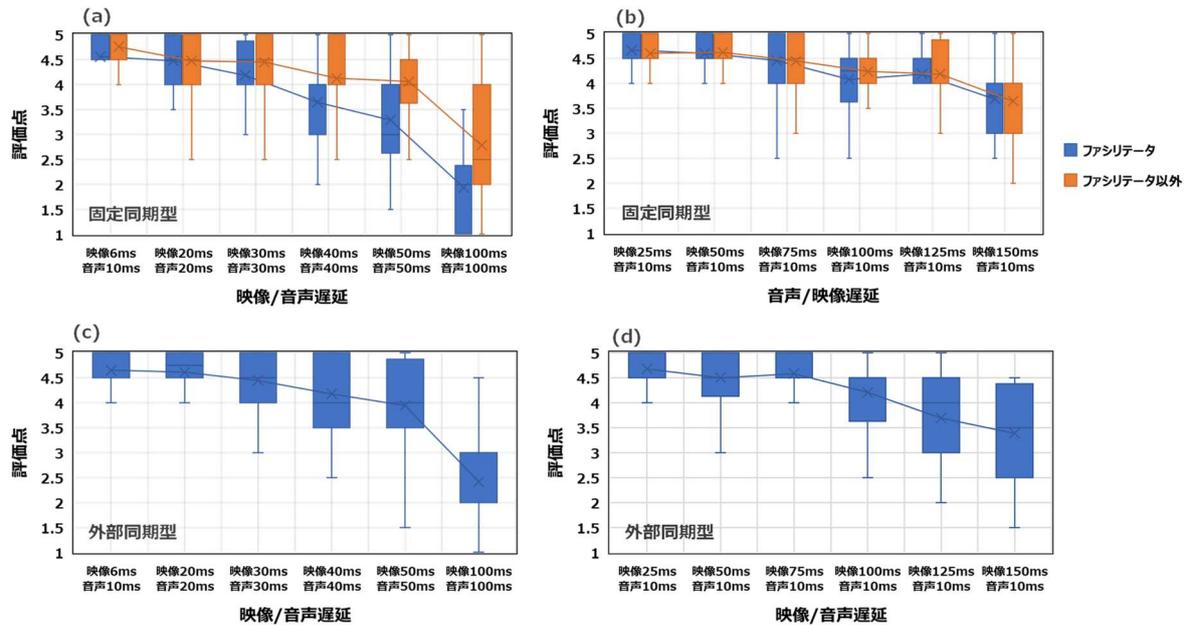


図 9 映像/音声遅延に対する体感品質評価の結果

Figure 9 Evaluation Results of Experience Quality for Video and Audio Delay.

外の映像をそれぞれ表示している。自拠点の映像はモニター間に配備されたカメラを用いて撮影した。尚、ユーザとマイク間の距離は約 1m とし、ユーザとカメラ/モニター間の距離は約 2m として設置した。

### 2.3 評価とパラメータ

本実験では、10~20 代の男女 30 人に対して、それぞれの拠点に 1 人を割り当て、3 人 1 組で実験を行った。

図 6 には、各実験の手順と流れを示す。実験では、より現実に近い(NW を経由しない場合と近い)遅延値として、本実験システムの最小遅延である映像 6ms/音声 10ms を基準条件とし、設定条件である遅延値と基準条件を比較することで評価を行った。まず、基準条件において、練習を行った後、遅延値を変更した設定条件での評価を行った。ここで、本実験に用いた設定条件は図 7 に示す。設定条件は各実験および各組に対してランダムに設定した。各設定条件での試行が終了するたびに被験者は評価結果を記入し、すべての設定条件について評価した。尚、実験 1 については、1 組 (3 人) につきファシリテータを 2 人選定し、ファシリテータを変更して実験 1 を 2 回実施した。また、実験 2 では外部同期源として[5]の音源を利用した。

### 3. 評価結果

図 8 に実際の評価の様子と評価指標を示す。実験では、図 8(a)に示すように、被験者は応援用のスティックメガホンを手手に持ち、ファシリテータの掛け声や音源に合わせて

とする。各設定条件での試行後に、図 8 (b)に示すアンケートを行った。アンケートは 9 段階の評価であり、1.0 から 5.0 までの評価値を被験者が選択する。図 9 には評価結果を示す。図 9(a)と(b)は固定同期型の結果を示す。固定同期型では、ファシリテータとファシリテータ以外についてそれぞれ評価している。ここで、本評価において、検知限となる閾値は、評価値の平均値で 4.0 に相当する [6]。図 9 (a) 音声と映像が同一遅延の場合、検知限はファシリテータ以外の場合、映像 100ms/音声 100ms であるのに対して、ファシリテータの場合では、映像 40ms/音声 40ms であった。これは、ファシリテータが他のユーザに比べて遅延に対する体感品質が厳しいことを意味している。また、図 9 (b)音声と映像が同一遅延でない場合を見ると、ファシリテータとファシリテータ以外で、評価値に大きな差は見られず、ファシリテータとファシリテータ以外の検知限は映像 150ms/音声 10ms であった。図 9 (a)と(b)の結果から、固定同期型の場合、映像と比べて音声の遅延の影響を強く受けると言える。また、ファシリテータ以外と比べて、ファシリテータがより音声遅延の影響を受けることが合わせて示された。

一方、図 9 (c)と(d)は外部同期型の結果を示す。図 9 (c)音声と映像が同一遅延の場合、検知限は映像 50ms/音声 50ms であった。また、図 9 (d)音声と映像が同一遅延でない場合、映像 125ms/音声 10ms が検知限であった。これらより、外部同期型のモデルにおいても、固定同期型と同様に、映像よりも音声の遅延に対して体感品質が影響を受

けることが言える。

これらの結果より、固定同期型と外部同期型を比較すると、同期型遠隔応援のモデルにおいては、固定同期型における同期源であるファシリテータが特に遅延の影響を受けやすく、遅延要件となりうる検知限は映像 40ms/音声 40ms であることが示された。

#### 4. おわりに

スポーツやイベント等での同期型遠隔応援アプリケーションを想定した、同期行動におけるメディア遅延の影響を評価した。同期型の応援をモデルに分類し、内部同期型の一つである固定同期型と、外部同期型における、ユーザ主観評価を行った。評価の結果、固定同期型と外部同期型では特に音声遅延の影響を受けやすく、検知限はそれぞれ、映像 40ms/音声 40ms、映像 50ms/音声 50ms であることが実験により示された。また、固定同期型では、同期源であるファシリテータが、遅延に対して影響を受けやすく、遅延要件が最も厳しいと言える。

#### 参考文献

- [1] 岩田貴裕, 瀬戸優樹. ファンの声援がスタジアムに届く—リモート応援システムとスポーツ界での取り組み—. 精密工学会誌, 2022, vol. 88, no. 1, pp. 12-13.
- [2] 駒崎掲, 渡邊淳司. 触覚伝送による“リモートハイタッチ”: アスリートの家族間コミュニケーションや聴覚障がい者との観戦検討. 日本バーチャルリアリティ学会論文誌, 2022, vol. 27, no. 1, pp. 2-5.
- [3] K. Yamagishi, J. Okamoto, T. Hayashi, A. Takahashi, "No reference video-quality-assessment model for monitoring video quality of IPTV services," *IEICE transactions on communications*, vol. 95, no. 2, pp. 435-448, 2012.
- [4] T. Hayashi, K. Yamagishi, T. Tominaga and A. Takahashi, "Multimedia Quality Integration Function for Videophone Services," in *Proc. IEEE GLOBECOM 2007 - IEEE Global Telecommunications Conference*, 2007, pp. 2735-2739.
- [5] “メジャーの星 cheering pep squad” . <https://dovas.jp/bgm/play8983.html>.
- [6] 大久保榮. テレビ会議教科書. VTV ジャパン株式会社, 2017. <https://vcbook.vtv.co.jp/pages/viewpage.action?pageId=1310872>.