

# ビデオ・音声・力覚メディアインタラクティブ IP 通信における メディア適応型メディア内同期制御による QoE 向上

磯村 栄一<sup>†a)</sup> 田坂 修二<sup>††</sup> 布目 敏郎<sup>†</sup>

QoE Enhancement by Media Adaptive Intra-Stream Synchronization  
in Audiovisual and Haptic Interactive IP Communications

Eiichi ISOMURA<sup>†a)</sup>, Shuji TASAKA<sup>††</sup>, and Toshiro NUNOME<sup>†</sup>

あらまし 本論文では、ビデオ・音声・力覚メディアインタラクティブ IP 通信を対象とし、各メディアの性質を考慮したメディア内同期制御を行うメディア適応型メディア内同期制御方式を提案する。提案方式が利用者の体感品質である QoE (Quality of Experience) の向上に有効であることを示すために、3 種類のメディア内同期制御方式 (方式 1: メディア適応型バッファリング, 方式 2: Skipping & buffering, 方式 3: バッファリング) を用いて主観評価実験を行う。加えて、力覚メディアにソーススキッピングを適用し、送出レートを減らした際の影響を調査する。主観評価実験では、応答性の重要度が異なる 2 種類の作業を用いて QoE を多角的に評価する。実験の結果、メディア適応型メディア内同期制御と力覚メディアのソーススキッピングの組合せは、QoE 向上に有効であることが示された。また、作業内容の違いによって異なる QoE の評価結果が得られた。

キーワード インタラクティブ IP 通信, ビデオ・音声・力覚メディア, 多次元 QoE, メディア内同期制御, ソーススキッピング

## 1. まえがき

医療, 教育, ロボット制御, ゲームなどの分野において、力覚メディアが利用されている [1]~[3]。力覚情報をネットワークで伝送することにより、遠隔地にいる別の利用者と力覚を共有できる。

本論文では、遠隔会議システムのような実空間において、利用者同士がビデオ・音声通信によって対等なコミュニケーションを行うことのできるサービスを考える。その高機能化の方法として、実空間においてビデオ・音声と力覚メディアを組み合わせた場合 (三感インタラクティブ IP 通信) を取り上げる。三感インタラクティブ IP 通信は、より高い臨場感を実現し、マルチメディア通信サービスの新たな利用形態を創出する。

IP ネットワークは、ベストエフォート型のサービスを提供しており、QoS (Quality of Service) が保証されない。すなわち、IP ネットワークではパケット欠落や伝送遅延、遅延揺らぎなど、制御できない要因によりメディアの品質が著しく低下する可能性がある。これは、サービスの利用者が体感する品質である QoE (Quality of Experience) [4]~[6] の低下につながる。QoE の定義にはユーザの満足度だけでなく、作業の有効性、効率、楽しさまで含まれることも多い [5]。

ビデオなどの連続メディア伝送において QoE を向上させる手法に、メディア内同期制御 (intra-stream synchronization) [7], [8] やソーススキッピング [9], [10] がある。メディア内同期制御はメディアの受信側で行う制御で、連続メディアの時間構造を正しく保持する役割をもつ。連続メディアを伝送する際には不可欠な制御であり、メディア内同期制御の方法や設定値は QoE に大きく影響を与える。また、ソーススキッピングはメディアデータを送信側で計画的に破棄し、ネットワークを流れるデータ量を減らすことで、ふくそうを抑制する制御である。

<sup>†</sup> 名古屋工業大学大学院工学研究科, 名古屋市  
Graduate School of Engineering, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan

<sup>††</sup> 名古屋工業大学総合工学プロジェクト研究所, 名古屋市  
Interdisciplinary Engineering Laboratory, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan

a) E-mail: isom@inl.nitech.ac.jp

ビデオ・音声インタラクティブ通信においては、設定するプレイアウトバッファリング時間が重要である [11]。プレイアウトバッファリング時間は、メディアの出力品質と応答性との間にトレードオフの関係を生じ、QoE を最大とするプレイアウトバッファリング時間が存在する。すなわち、プレイアウトバッファリング時間が長すぎると大きなネットワーク遅延揺らぎも吸収でき、メディアの出力品質は高まるが、エンドツーエンド遅延の増加により応答性の低下につながる。プレイアウトバッファリング時間が短すぎる場合は、逆の結果となる。したがって、プレイアウトバッファリング時間には最適値が存在する。

一方、ビデオ・音声・力覚の各メディアは、伝送時のデータサイズや送出レート、許容遅延などが異なる [8], [12]。具体的には、力覚メディアは遅延への耐性が低く、ビデオ・音声はメディアデータの欠落に対する耐性が低い。そのため、力覚メディアにも最適なプレイアウトバッファリング時間は存在するが、その値はビデオ・音声のものよりも小さくなる。このようなメディアの性質の違いを考慮したメディア内同期制御を行うことで、QoE の向上が期待できる。しかし、従来の研究にはこの点に着目し、議論されたものはない。

例えば、ビデオ・音声・力覚メディアの伝送を扱った文献 [8] では、3 メディアに対して同値のバッファリング時間を設定し、QoE を評価している。その結果、バッファリングによるエンドツーエンド遅延の増加は、被験者に提示される反力増加の原因となり、力覚メディアの操作性が低下している。また、従来の研究では、三感インタラクティブ通信の応答性に関する議論も十分でない。文献 [8] では、力覚メディアにより遠隔にあるタンバリンをたたく作業を行っている。しかし、タンバリンは片方の端末にのみ設置されており、もう片方の端末からタンバリンを一方向的にたたき続ける。この作業では、被験者間の指示や会話がないう。応答性が QoE に与える影響を調査するには、本論文で対象としているような、被験者間で対等なインタラクションが必要となる作業が重要である。

力覚メディアの性質を考慮し、メディア内同期制御での応答性を高めるためにプレイアウトバッファリング時間を短く設定することが考えられる。しかし、この方法では一般的に力覚情報の欠落が発生しやすくなる。この欠落は、受信側において遅延揺らぎによりランダムに発生するため、出力品質の低下につながる。そこで、本論文では送信側で情報転送量を規則的に減

らすソーススキッピングによる QoE の向上も考える。

ソーススキッピングに関する研究として、文献 [9], [10] がある。文献 [9], [10] では、CG による仮想環境を用いたシステムにおいて、力覚メディアに Dead reckoning を適用している。しかし、文献 [9], [10] では、ビデオや音声扱われておらず、力覚メディアのソーススキッピングがこれらの出力品質に与える影響については、定量的な評価が行われていない。

本論文の新規性は、メディアの性質を考慮したメディア内同期制御を提案したことと、その QoE を多次元的かつ定量的に評価し、提案方式の有効性を立証したことである。メディアの受信側で行う制御では、各メディアの性質を考慮したメディア適応型メディア内同期制御 [13] を提案し、3 メディアに同値のバッファリング時間を設定した場合 [8] よりも高い QoE が得られることを多次元的に示す。また、メディアの送信側で行う制御として、力覚メディアのソーススキッピングが QoE に与える影響も多次元的に評価する。同時に、アプリケーションレベル QoS [6] の評価も行う。QoE を評価する主観評価実験では、作業内容の違いが QoE 評価に与える影響を考慮し、特徴の異なる 2 種類の作業を用いる。多次元 QoE の評価結果から、作業ごとに適切なプレイアウトバッファリング時間を定量的に調査するとともに、適切なメディア内同期制御方式を明らかにする。

以下では 2. で本論文の三感インタラクティブ通信システムを紹介する。3. ではメディア内同期制御を説明し、本論文で用いる制御方式を示す。4. に力覚メディアのソーススキップの方法を記述する。5. で実験システムを概説し、6. では作業内容と多次元 QoE の評価方法を説明する。7. で実験結果と考察を示す。最後に 8. で本論文のまとめと今後の課題を述べる。

## 2. 三感インタラクティブ通信システム

本論文で対象としている三感インタラクティブ通信システムにおけるメディア端末の機能を図 1 に示す。各メディア端末には、触覚インタフェース装置 (PHANToM Omni)、Web カメラ、ヘッドセットがそれぞれ接続され、3 メディアを入出力する機能をもつ。本論文で用いる PHANToM Omni (以降、PHANToM と記す) とは、ペン型の触覚インタフェース装置であり、スタイラス (ペンの部分) にかかる力を入出力することができる。なお、図 1 では紙面の都合上、メディア端末 2 を簡略化して記しているが、メディア端末 1, 2 は同じ構成及び機能をもつ。ビデオのキャプチャ

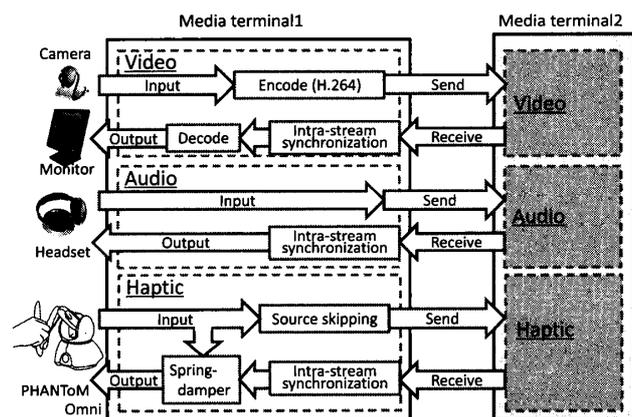


図 1 メディア端末の機能  
Fig. 1 Functions of media terminals.

は Web カメラによって行い、ビデオ符号化方式には H.264 を用いる。音声は、ヘッドセットのマイクロホンでキャプチャする。PHANToM からはスタイラスの位置情報を入力として得る。入力した位置情報は、ソーススキッピングにより送信または破棄される。

本論文では、アプリケーションレベルにおけるメディアデータの処理単位として MU (Media Unit) を用いており、ビデオ 1 フレーム、音声 320 サンプル、当該時刻における PHANToM の位置情報をそれぞれ 1 MU とする。また、3 メディアは別々の UDP ストリームとして相手端末へ伝送される。

本システムでは、いずれのメディアも二つのメディア端末間で双方向に伝送される。受信した MU は、メディア内同期制御により時間構造が保持される。メディア内同期制御の処理は各メディアごとに独立して行う。

PHANToM から被験者に提示される反力の計算式は、反力  $\mathbf{F}$  [N]、自端末と相手端末におけるスタイラス位置の差分  $\mathbf{x}$  [mm] 及び弾性係数  $k$  [N/mm] を用いて  $\mathbf{F} = k\mathbf{x}$  と表される [14]。エンドツーエンド遅延が増加すると  $\mathbf{x}$  の値が大きくなりやすくなるため、反力が増加する。反力は、スタイラス同士が引き合う方向に働く。なお、本論文では、弾性係数に 0.1 を固定値として設定している。力覚 MU を受信していない場合、新たな反力は加わらない。

### 3. メディア内同期制御

メディア内同期制御は、単一メディアにおける時間構造を保持する役割をもつ。本論文では、Skipping とプレイアウトバッファリング [7] を扱う。

#### 3.1 Skipping

Skipping では、受信した MU に対し、順序制御を

したのち、これを直ちに出力する。二つ以上の MU を同時に受信した場合、その中から最も新しい MU のみを出し、残りを破棄する。Skipping における順序制御では、MU に付与されたシーケンス番号から、受信した MU が既に出力された MU よりも古いものであると判断された場合、受信した MU を破棄する。

#### 3.2 プレイアウトバッファリング

プレイアウトバッファリングは、受信した MU をバッファに格納し、出力すべき時刻 (目標出力時刻) に出力する制御である [8]。

最初に到着した MU ( $MU_1$ ) の目標出力時刻  $O_1$  は、 $MU_1$  の到着時刻  $A_1$  とプレイアウトバッファリング時間  $T_b$  を用いて式 (1) のように表される。

$$O_1 = A_1 + T_b \quad (1)$$

また、 $i$  ( $i \geq 2$ ) 番目の MU ( $MU_i$ ) の目標出力時刻  $O_i$  は、 $MU_1$  の発生時刻  $B_1$  と  $MU_i$  の発生時刻  $B_i$  を用いて式 (2) のように表される。

$$O_i = O_1 + (B_i - B_1) \quad (2)$$

なお、MU の発生時刻  $B_1$ ,  $B_i$  は、MU のヘッダに記録されて送信される。 $MU_i$  の到着時刻  $A_i$  が  $O_i$  よりも前であれば、 $MU_i$  を  $O_i$  までバッファリングしたのち、出力する。

また、目標出力時刻を越えて受信した MU に対しては、直ちに出力する場合と破棄する場合が考えられる。本論文のプレイアウトバッファリングでは、受信した MU に対して Skipping と同じ順序制御を行い、破棄されなかった場合のみ直ちに出力する。

#### 3.3 メディア適応型メディア内同期制御方式

ビデオ・音声・力覚メディアは、それぞれ異なる性質をもつ [8], [12]。そのため、各メディアにおける最適なメディア内同期制御は異なり得る。そこで、メディアの性質を考慮したメディア内同期制御を行うメディア適応型メディア内同期制御方式 [13] を用いる。この方式として、各メディアに適したプレイアウトバッファリングを行う場合 (メディア適応型バッファリング) と各メディアに応じてプレイアウトバッファリングか Skipping のいずれかを行う場合が考えられる。そこで、本論文では、次節で定義するメディア適応型バッファリングと Skipping & buffering を扱う。

#### 3.4 比較方式

主観評価実験では、メディア内同期制御方式として、次に示す 3 方式を用いる。

#### [方式1] メディア適応型バッファリング

ビデオ, 音声, 力覚メディアに対してプレイアウトバッファリングを行い, 各メディア別にそれぞれ適したプレイアウトバッファリング時間を設定する方式.

#### [方式2] Skipping & buffering

ビデオと音声に対しては, プレイアウトバッファリングを行い, 力覚メディアには, Skipping を行う方式.

#### [方式3] バッファリング

ビデオ, 音声, 力覚メディアに対してプレイアウトバッファリングを行い, 全メディアに同値のプレイアウトバッファリング時間を設定する方式.

なお, 方式1と方式2でもビデオ・音声に対しては同値のプレイアウトバッファリング時間を用いる. これは, リップシンクを考慮すると, ビデオ・音声間の同期品質は高い方がよいと考えられるためである.

方式1と方式2では, ビデオ・音声の出力品質を維持しつつ, 力覚メディアの操作性を高く保つことができる. しかし, MUを受信してから出力するまでの時間がメディアごとに異なる場合があり, メディア間の同期誤差が方式3に比べて大きくなり得る. 一方, 方式3ではプレイアウトバッファリング時間が短い場合, ビデオや音声の出力品質が低下する可能性があり, プレイアウトバッファリング時間が過度に長い場合, 力覚メディアの操作性が低下してしまう[8]. なお, Skipping & bufferingにおけるQoE評価は, 文献[13]で行われている. また, 本論文ではインタラクティブ通信を対象としている. そのため, いずれの3方式においても, ビデオ・音声のプレイアウトバッファリング時間が増加した場合に応答性が低下する.

### 4. 力覚メディアのソーススキッピング

本論文における力覚メディアのソーススキッピングの動作を説明する. PHANToMでは, 操作精度を高めるため1kHzの力覚情報更新が推奨されている[14]. したがって, 相手のPHANToMに力覚を伝える場合, 1秒間に1000個の力覚MUを転送する必要がある. しかし, これはネットワークふくそう発生の原因となり得る. また, 本論文のシステムのように力覚メディアを実空間で扱う場合は, 仮想空間ほどの高い操作精度は要求されないと考えられる. したがって, 力覚メディアの用途によっては, ネットワークふくそうを緩和できるソーススキッピングにより, QoE向上が期待できる. 本論文のシステムでは1秒間に発生する1000個の力覚MUに対し, メディアの送信側で力覚MUの送信

と破棄を交互に繰り返すことで, 力覚MUの送出レートを500MU/sに下げている(注1). 力覚メディアの適切なMUレートは, 作業によって異なると考えられる.

### 5. 実験システム

実験システムの構成を図2に示す. 実験システムは, メディア端末2台, 負荷送信端末2台, 負荷受信端末2台, ルータ2台(Cisco2811)からなる. ルータ間は10Mbit/sの全二重Ethernet回線で接続され, その他の回線は100Mbit/sの全二重Ethernetである. 10Mbit/sの回線がボトルネックとなる.

メディアの仕様について, 表1, 表2に, それぞれビデオ・音声のもの, 力覚メディアのものを記す. ビデオキャプチャ時の解像度は800×600画素である. ビデオ1フレームは符号器によって15の方形スライスに分割され, 各々はIPパケットとして相手に転送される. ビデオスライスが欠落した際は, FFmpegによる誤り補償を行う. 力覚メディアのMUレートは,

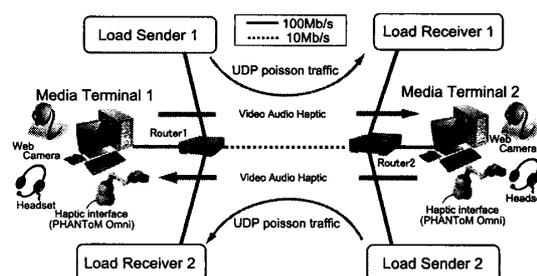


図2 実験システムの構成

Fig. 2 Experimental system configuration.

表1 ビデオ・音声の仕様

Table 1 Specifications of video and audio.

	ビデオ	音声
符号化方式	H.264 (x264)	Linear PCM
符号化パラメータ	800 × 600 画素	16 kHz 8 bit 1ch
ピクチャパターン	IPPPP	-
スライス数	15	-
符号化ビットレート [kbit/s]	約 2048	128
MU レート [MU/s]	25	50

表2 力覚メディアの仕様

Table 2 Specification of haptic.

MU レート [MU/s]	1000	500
符号化ビットレート [kbit/s]	320	160

(注1): 本論文では簡単化のため Dead reckoning は用いない. Dead reckoning には, 遅延揺らぎが増加した場合, 予測した力覚情報と実際の力覚情報との誤差が大きくなり, 操作性が著しく低下する問題点があるためである[9].

ソーススキッピングを行わない場合の 1000 MU/s とソーススキッピングを行う場合の 500 MU/s を用いる。

背景トラフィックとしてランダム (ポアソン) トラフィックを想定し、この実験システムを構築する。負荷送信端末 1, 2 は、UDP 負荷トラフィックを、それぞれ負荷受信端末 1, 2 に向け送信する。負荷送信端末では、1472 バイトの UDP データグラムを指数分布に従う間隔で発生させ、送信する。これにより、ルータ間のボトルネックリンクにおいてふくそうを発生させる。実験環境としては、UDP を利用したインタラクティブサービスが集まったものに対応する。実験で発生させる負荷トラフィックの平均ビットレートは、6.0 Mbit/s<sup>(注2)</sup> である。なお、TCP と UDP の混合トラフィックを用いた実験を行ったが、UDP トラフィックのみの場合と同様の傾向であった [15]。しかし、多様なネットワーク環境による実験は、今後の課題として挙げられる。

## 6. 評価方法

### 6.1 作業内容

三感インタラクティブ通信における作業特性には、PHANToM の利用方法や利用頻度、操作速度の違い、ビデオ・音声の重要度の違いなどがある。

システムの利用者間でインタラクションがあるインタラクティブ通信では、応答性の善しあしが QoE に影響を与える重要な要因である [11]。応答性が重要な作業では、メディアの出力品質を考慮しただけでは QoE 向上が困難な可能性がある。そこで本論文では、特に応答性の影響を調査するため、オブジェクト移動作業とカスタネットたたき作業の 2 種類を用いる。

両作業とも被験者は二人一組となり、一方が操作役、他方が指示役となって作業を進める。被験者はそれぞれ別の部屋で作業を行い、各被験者に同じ作業領域、機材が与えられる。作業中は、両方の被験者はスタイラスを握る。操作役の被験者がスタイラスを動かすと、それに応じて指示役側のスタイラスが引っ張られるように動作する。このとき、指示役の被験者はスタイラスの動きを操作者に委ね、PHANToM を主体的に操作することはしない。なお、両作業ともに被験者は各メディアの劣化がない状態で練習を行い、作業の流れ及び装置の操作に慣れた段階で実験を開始している。

#### (作業 1) オブジェクト移動作業

相手側の PHANToM を遠隔操作し、指定された実物のオブジェクトを移動する作業を行う。オブジェクトは 3 種類用意し、それぞれ円形 (直径 3 cm)、正方

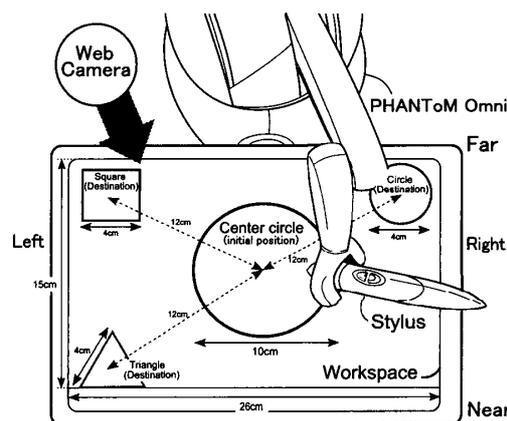


図 3 オブジェクト移動作業：作業領域及び機材配置  
Fig. 3 Task1: Workspace and layout.

形 (一辺 3 cm)、正三角形 (一辺 3 cm) の形をしている。オブジェクトの厚さはいずれも 5 mm 程度で、重さは約 1.5 g である。図 3 に作業 1 における作業領域と機材配置を示す。Web カメラは上から見下ろすように配置し、撮影範囲は作業領域全体である。

作業の手順を示す。作業が始まる前に三つのオブジェクトを中央の円内に配置する。作業が始まると、指示役の被験者がオブジェクトの形と移動先を、音声で相手に伝える (例えば、丸を三角へ)。操作役の被験者は、相手側に配置されたオブジェクトを、ビデオで位置関係や状況を確認しながら、PHANToM によって指定された場所まで移動させる。オブジェクトの移動は、オブジェクトの側面とスタイラスの先を接触させながら押し進めるように行う。移動が完了したら役割を交代し、作業を続ける。これを 30 秒間続ける。

#### (作業 2) カスタネットたたき作業

PHANToM を用いて楽器をたたくことによる掛け合い作業を行う。両被験者には、図 4 に示す作業領域が与えられる。カスタネットは実物で、スタイラスの真下に固定されている。カメラの撮影範囲は、カスタネットとスタイラス、被験者の前腕が映る程度である。

まず、指示役の被験者がカスタネットをたたく回数を 1 から 3 回の中からランダムに選択し、相手に音声で伝える。他方の被験者は、両被験者のカスタネットが鳴るように PHANToM を操作する。このとき、操作側の被験者はスタイラスを約 15 cm の高さから振り

(注2)：実験システムを用いてメディアデータと負荷トラフィックを同時に伝送したところ、6.0 Mbit/s 未満の負荷ビットレートでは、メディアの劣化がほとんど生じなかった。一方、6.0 Mbit/s よりも大きい負荷ビットレートでは、ネットワーク遅延が大きくなりすぎてしまい、力覚メディアの操作性が著しく低下した。そのため、本論文では、実験環境として適切だと考えられる 6.0 Mbit/s の負荷ビットレートを用いている。

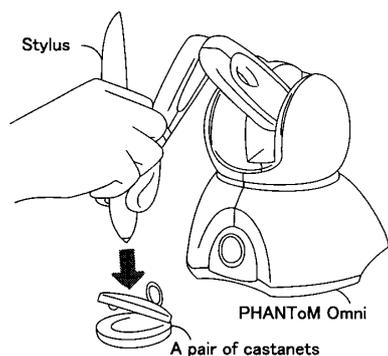


図4 カスタネットたたき作業：作業領域及び機材配置  
Fig. 4 Task2: Workspace and layout.

下ろす操作を行う。自端末におけるカスタネットの音は、ヘッドセットのマイクロホンによって集音され、相手の被験者に伝わる。指示された回数をたたいたのち、役割を交代し、作業を続ける。この流れを30秒間続ける。楽器をたたく間隔は、実験前に被験者に提示する。

オブジェクト移動作業は、ビデオにより相手側の様子を確認する必要がある。また、PHANTOMの操作速度は遅く、細かい位置調整が重要である。オブジェクト1個を移動するのに要する時間は約10秒であり、スタイラスの移動速度は30から60mm/s程度である。オブジェクト移動作業では、指示などの被験者間におけるやり取りが少なく、応答性は重要ではない。

一方、カスタネットたたき作業は、PHANTOMの精密な操作は要求されない代わりに、これの操作速度が速い。カスタネットを1回たたくために要する時間は0.7から1秒程度である。そのときのスタイラス移動速度は300から500mm/sである。また、この作業では被験者間におけるやり取りの回数が多いことに加えて、音声の遅延が自端末と相手端末におけるカスタネット音のずれとなり被験者に知覚される。そのため、カスタネットたたき作業は応答性が重要な作業である。

## 6.2 QoE 評価方法

QoEの評価はSD (Semantic Differential) 法[16]を用いて多次元的に行う。SD法は対義語の組合せで表される評価対を複数用いて、対象を複数の視点から評価する手法である。実験で用いる評価対は、文献[17],[18]を参考に選定した。評価対は、ビデオ、メディア間同期品質、力覚メディア、音声、インタラクティブ性、総合品質に関する項目に分類される。選定した評価対を以下に示す。

[ビデオ]

(v1) 空間品質：映像が崩れた-まとまった

(v2) 時間品質：映像がぎこちない-滑らかな

(v3) 状況把握：状況が把握しにくい-把握しやすい  
[メディア間同期品質]

(s1) ビデオ-力覚メディア間同期品質<sup>(注3)</sup>：

PHANTOMの動きとずれた-一致した  
[力覚メディア]

(h1) 安定感：操作が不安定な-安定な

(h2) 滑らかさ：動作がぎこちない-滑らかな

(h3) 操作感：操作が重い-軽い

[音声]

(a1) 自然さ：音声が不自然な-自然な

(a2) 楽器音の同期品質<sup>(注4)</sup>：

カスタネット同士の音がずれた-一致した

[インタラクティブ性]

(i1) 応答性：応答性が悪い-良い

(i2) コミュニケーションの自然さ：

やり取りがぎこちない-スムーズな

(i3) 快適さ：不快な-快適な

(i4) 作業難易度：作業が難しい-簡単な

[総合品質]

総合品質：不満な-満足な

各評価対に対し、被験者は与えられた刺激の評価を、5段階の評定尺度法[19]により行う。最も良いカテゴリは5であり、上記に示した評価対の右側の形容詞(動詞)に相当する。一方、最も悪いカテゴリは1であり、上記に示した評価対の左側の形容詞(動詞)に相当する。カテゴリ3は“どちらでもない”を表し、カテゴリ2, 4はそれぞれ、カテゴリ1, 5の形容詞(動詞)に“やや”を付与したものとする。

本論文におけるQoE尺度は、心理的尺度[19]と呼ばれる距離尺度[20]を用いている。QoEの評定尺度として一般的に用いられているものにMOS (Mean Opinion Score)がある。これは、ある刺激に対する全被験者の評定の算術平均にあたる。しかし、MOSは順序尺度であり、厳密には評価値の大小比較のみを行うことができる。一方、距離尺度は、数値間の等間隔性が保証されるため、評価値同士の比較を行うことができる。そのため距離尺度は、人間の主観を順序尺度よりも正確に表現することができる。

距離尺度は、系列カテゴリ法により算出でき

(注3)：予備実験により、ビデオと音声間、音声と力覚メディア間、ビデオと力覚メディア間同期品質のうち、QoEに最も支配的であったビデオと力覚メディア間同期品質を選定した。

(注4)：カスタネットたたき作業(作業2)でのみ用いている。

る [19], [20]. 系列カテゴリ法では, 評定尺度法とカテゴリ判断の法則を組み合わせて用いる. 本論文では, SD 法により得られた各評価対の評点に対して, カテゴリ判断の法則を適用し, 各評価対の距離尺度を算出している. 更に, 得られた各距離尺度値の妥当性を確認するために, Mosteller の適合度検定 [21] を適用する. 検定により距離尺度の妥当性が確認できた場合, これを心理的尺度として QoE 尺度に用いる.

更に, 効率を表す尺度として, 各作業における作業効率を用いる. 作業 1 (オブジェクト移動作業) における作業効率は, 被験者一人当たりが作業時間内に移動できたオブジェクトの数を, 作業 2 (カスタンネットたたき作業) においては, 被験者一人当たりが指示を出せた回数を, それぞれ用いる.

主観評価実験で用いたプレイアウトバッファリング時間の組合せを以下に示す.

[方式 1] メディア適応型バッファリング

ビデオ・音声: 20, 40, 60, 100, 150, (300) ms

力覚メディア: 10 ms (固定)<sup>(注5)</sup>

[方式 2] Skipping & buffering

ビデオ・音声: 20, 40, 60, 100, 150, (300) ms

力覚メディア: Skipping (固定)

[方式 3] バッファリング

ビデオ・音声・力覚メディア

: 20, 40, 60, 100, 150 ms

なお, 300 ms のプレイアウトバッファリング時間は応答性の調査が重要となる作業 2 でのみ適用する. また, 方式 3 の力覚メディアに対して設定する 100 から 150 ms のバッファリングは力覚の操作性を低下させると予想されるが, メディアの劣化傾向を連続的に示すために, これらを用いる.

各作業の評価対象は, 先に示したプレイアウトバッファリング時間の組合せに加え, 力覚メディアのソーススキッピングの有無で 2 通りであるので, 作業 1 で 30 通り, 作業 2 で 34 通りとなる. 実験は各作業ごとに行い, 評価対象はランダムな順序で被験者に提示される. また, 行う作業の順序も被験者ペアによってランダムとした. 被験者は 10 代から 50 代の男女 42 名 (男性 16 名, 女性 26 名) である. 被験者一人当たりの評価時間は, 一作業当たり約 60 分である.

## 7. 評価結果と考察

アプリケーションレベル QoS の測定結果と QoE の評価結果を以下に示す. なお, 本章ではプレイアウト

バッファリング時間を  $T_b$  と表記している.

### 7.1 アプリケーションレベル QoS 評価結果

本論文では, アプリケーションレベル QoS パラメータとして, ビデオスライス到着率, 力覚 MU 出力レート, 力覚平均 MU 出力遅延, ビデオ・力覚メディア間及びビデオ・音声間同期平均二乗誤差の平方根を扱う.

ビデオスライス到着率は, 送信された全ビデオスライスのうち, 出力されたビデオスライスの割合である. MU 出力レートは, 1 秒間に出力された平均 MU 数である. 平均 MU 出力遅延は, MU の出力時刻から MU の発生時刻を引いた値の平均値である. メディア間同期平均二乗誤差の平方根とは, 力覚 (または音声) MU とそれに対応するビデオ MU の出力時刻の差から, それらの発生時刻の差を引いた値を二乗平均し, 平方根をとった値である.

測定結果をそれぞれ図 5 から図 9 に示す. 縦軸には各パラメータの測定値を記す. 横軸は, ビデオ・音声及び方式 3 の力覚メディアに対して設定した  $T_b$  を示す. 方式 1 (メディア適応型バッファリング) における力覚メディアの  $T_b$  は 10 ms 固定, 方式 2 (Skipping & buffering) では, 力覚メディアに対して Skipping を用いていることに注意されたい. グラフ中の凡例は, 力覚 MU レート [MU/s]-メディア内同期制御方式の形で表記されている. 力覚 MU レートについて, 1000 はソーススキッピング未適用を示し, 500 はソーススキッピングが適用されていることを示す. メディア内同期制御方式は, mab はメディア適応型バッファリング (方式 1), sb は Skipping & buffering (方式 2), buf はバッファリング (方式 3) をそれぞれ表す. グラフには作業 1 (Task1) と作業 2 (Task2) での測定結果を同時に記し, 95%信頼区間も記す.

図 5 は, ビデオスライス到着率を示している. ビデ

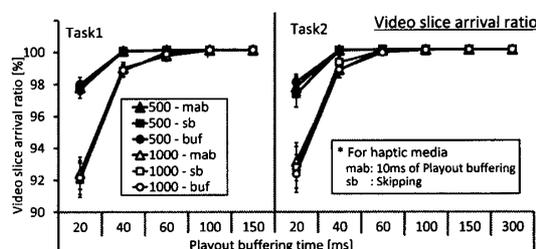


図 5 ビデオスライス到着率  
Fig. 5 Video slice arrival ratio.

(注5): 事前に実施した予備的な実験により, PHANTOM の操作感と動作の滑らかさの点から, 力覚メディアに最も適したプレイアウトバッファリング時間として 10 ms を設定した.

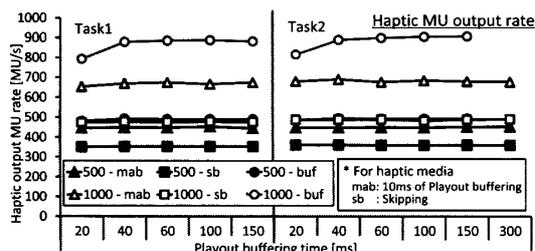


図 6 力覚 MU 出力レート  
Fig.6 Haptic MU output rate.

オの  $T_b$  が 20 ms から 60 ms のとき、力覚メディアのソーススキッピングを行った場合、ビデオスライス到着率が向上している。このことから、力覚メディアのソーススキッピングにより、ビデオの遅延揺らぎが小さくなり、出力されたビデオスライスが増加したといえる。作業内容にかかわらず、この傾向が見られる。 $T_b$  が 100 ms 以上のときスライス到着率が 100%であるため、ネットワークレベルでの欠落が発生していない。ネットワーク負荷量は  $T_b$  とは独立に一定である。そのため、 $T_b$  が 20 ms から 60 ms におけるスライス到着率の低下は、遅延揺らぎによるメディア内同期制御での MU 破棄が原因である。また、力覚 MU 送出レートを 300 MU/s として実験を行ったところ、同様にビデオスライス到着率が向上した。

力覚 MU 出力レートの測定結果を図 6 に示す。力覚 MU の出力レートも作業によって違いが見られない。また、力覚メディアのソーススキッピングを行う場合と行わない場合ともに、力覚メディアに対してプレイアウトバッファリングを行う方式 3 (buf) が最も高い値をとり、Skipping を行う方式 2 (sb) が最も低い値をとっている。比較方式間で UDP 負荷量に違いはないため、3 方式における MU 出力レートの差はメディア内同期制御の遅延揺らぎ耐性の違いによるものである。方式 3 では、力覚メディアに対して  $T_b$  を 20 ms 以上 (横軸の値と同値) に設定したため、遅延揺らぎによる MU の欠落が発生しにくい。一方、Skipping では、同時に複数の MU を受信した場合、その中から最新のもののみを出力し、残りを破棄する。そのため、わずかな遅延揺らぎが発生した場合でも力覚 MU の欠落が発生し、力覚 MU 出力レートが低下する。方式 1 (mab) では、力覚メディアに対して 10 ms のプレイアウトバッファリングを行うため、Skipping を行う方式 2 (sb) よりも MU 出力レートが高くなる。特に、力覚メディアのソーススキッピングを適用した場合、方式 1 と方

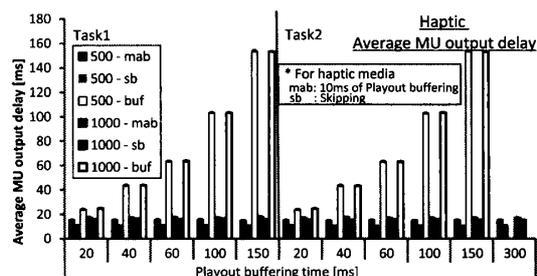


図 7 力覚平均 MU 出力遅延  
Fig.7 Haptic average MU output delay.

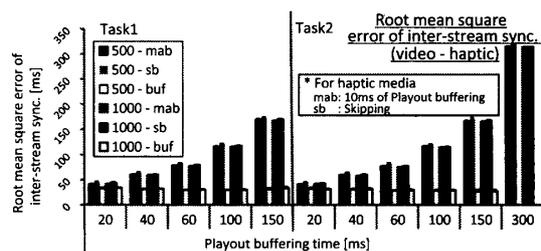


図 8 メディア間同期平均二乗誤差の平方根 (ビデオ-力覚)

Fig.8 Root mean square error of inter-stream synchronization (video-haptic).

式 3 では、ほぼ同値の MU 出力レートとなっている。

図 7 は、力覚平均 MU 出力遅延を示している。方式 3 では全メディアに同値のバッファリング時間を設定するので、力覚メディアの平均 MU 出力遅延は、横軸の値に比例して大きくなっている。方式 1 や方式 2 では、MU 受信後から出力までの時間が短いため、15 ms 程度の遅延となっている。なお、ビデオ及び音声の平均 MU 出力遅延の測定結果は、力覚メディアの方式 3 とほぼ同値であったため省略する。

ビデオ・力覚メディア間同期平均二乗誤差の平方根について、図 8 より、方式 3 では 3 メディアに同値の  $T_b$  を設定しているため、同期誤差が小さい。一方、方式 1 や方式 2 では、各メディアに適したメディア内同期制御を行うため、ビデオの  $T_b$  が大きくなるほど、同期誤差が大きくなる。この誤差が主観的にはどのように感じられるかは、次節で検討する。音声・力覚メディア間同期平均二乗誤差の平方根は、ビデオ・力覚メディア間の場合とほぼ同値であったため省略する。

図 9 にビデオ・音声間同期平均二乗誤差の平方根を示す。ビデオと音声に対しては、3 方式とも、同値の  $T_b$  を設定するため、同期誤差が大きくなる。

## 7.2 QoE 評価結果

本論文の実験システムにおける QoE を特徴づける評価尺度として、ビデオ空間品質、力覚メディア操作

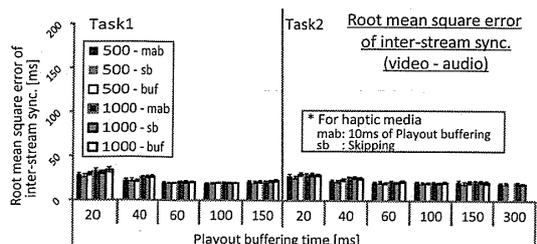
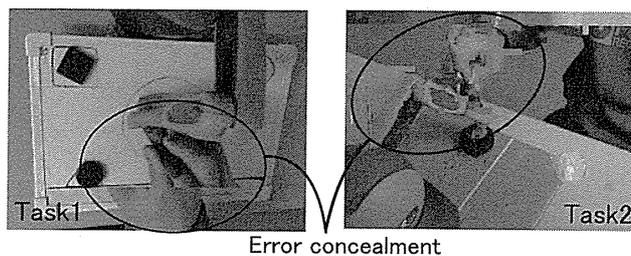


図 9 メディア間同期平均二乗誤差の平方根 (ビデオ-音声)

Fig. 9 Root mean square error of inter-stream synchronization (video-audio).



Error concealment

図 11 ビデオ誤り補償の表示例

Fig. 11 Examples of error-concealed video.

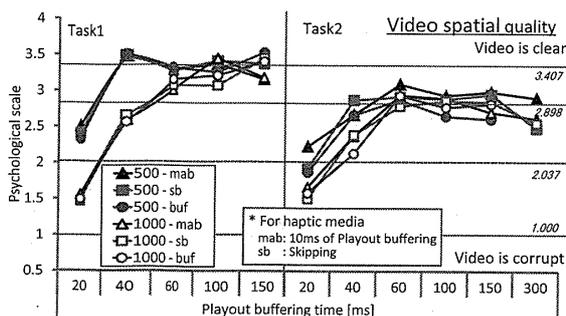


図 10 ビデオ空間品質

Fig. 10 Video spatial quality.

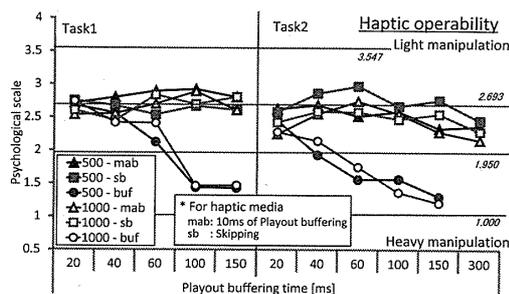


図 12 力覚メディア操作感

Fig. 12 Operability of haptic.

感, メディア間同期主観品質, 総合品質, 楽器音のずれに関するものを取り上げる。

主観評価実験により得られた評点に対して, 各評定尺度ごとにカテゴリ判断の法則を適用し, それぞれ距離尺度を算出した。得られた尺度値に対し, 有意水準 0.05 で Mosteller の適合度検定を行った。その結果, 本論文で扱う評価尺度では, その尺度値が測定結果に適合した。そこで, 算出された距離尺度を心理的尺度として用いた。心理的尺度ではカテゴリ 2 の下限値が 1.000 となるように原点を定めた。

QoE 評価結果を, 図 10, 図 12 から図 15 に示す。グラフの縦軸は心理的尺度値, 横軸にはアプリケーションレベル QoS と同様に, ビデオ・音声及び方式 3 における力覚メディアの  $T_b$  を示す。右端の数値は下から順にカテゴリ 2 からカテゴリ 5 の下限値を表す。

図 10 のビデオ空間品質において, ビデオの  $T_b$  が 20 ms から 60 ms のとき, 力覚メディアのソーススキッピングを適用することで, ビデオ空間品質の向上が確認できる。これは, 図 5 で示したとおり, 力覚メディアのソーススキッピングにより出力されるビデオスライスの割合が増加したためである。各作業における, ビデオスライス欠落時の誤り補償表示例を図 11 に示す。図 5 からはビデオスライス到着率に作業間の差が見られないが, 図 10 からはカスタネットたたき作業

を行った場合の方が, ソーススキッピングによる QoE の向上幅が小さいことが分かる。これはビデオにおける被写体の移動速度の違いが影響しているためである。カスタネットたたき作業の場合, オブジェクト移動作業に比べ, より広い範囲を撮影しており, 被写体である腕の動きも速く大きい。このことから, カスタネットたたき作業ではビデオ誤り補償による空間品質の劣化が目立ちやすく, QoE が向上しにくかったと考えられる。次に, 力覚メディアの操作感に関する QoE 尺度について考察を行う。図 12 において, 提案方式である方式 1 (メディア適応型バッファリング: mab) や方式 2 (Skipping & buffering: sb) では, 力覚 MU に生じるメディア内同期制御の遅延を小さくしているため, 作業 1, 作業 2 とともに高い操作感を達成している。一方, ビデオ・音声と同値の  $T_b$  を設定する方式 3 (バッファリング: buf) では,  $T_b$  の増加に従って反力が大きくなりやすくなる。そのため,  $T_b$  が 100 ms や 150 ms のとき, 方式 3 における操作感に関する QoE が著しく低下している。また, 方式 3 で  $T_b$  が 40 ms や 60 ms のとき, カスタネットたたき作業における QoE は, オブジェクト移動作業におけるそれと比べ小さい値をとっている。これは, カスタネットたたき作業の方が PHANToM スタイラスの操作速度が速く, 遅延による反力の増加を知覚しやすかったためである。

ビデオ・力覚メディア間同期主観品質を図 13 に示

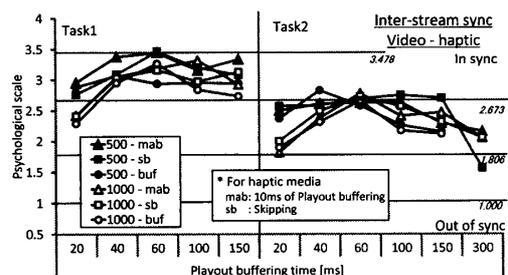


図 13 メディア間同期主観品質 (ビデオ-力覚)  
Fig. 13 Subjective quality of inter-stream synchronization (video-haptic).

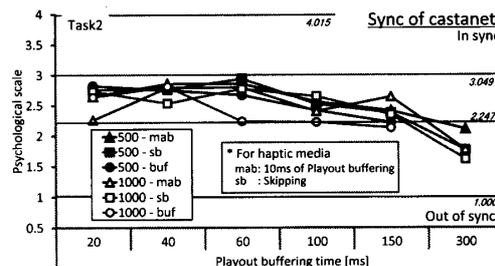


図 15 楽器音の同期品質  
Fig. 15 Synchronization of castanets.

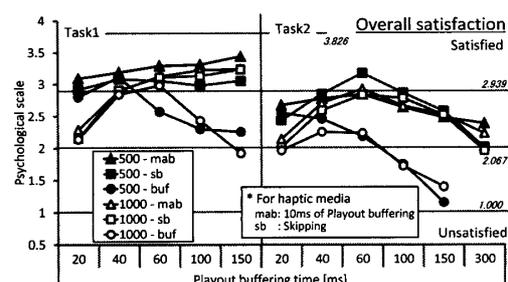


図 14 総合品質  
Fig. 14 Overall satisfaction.

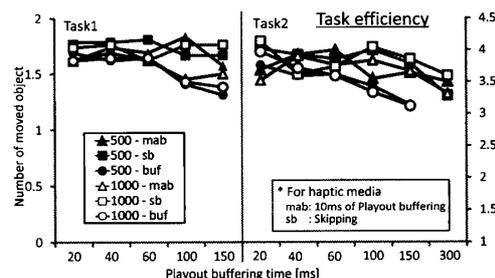


図 16 作業効率  
Fig. 16 Task efficiency.

す。図 8 に示したとおり、方式 1 と方式 2 におけるビデオと力覚メディア間の出力同期誤差はビデオの  $T_b$  増加に従って大きくなる。一方、方式 3 ではビデオと力覚メディア間の同期誤差を小さく保つことができる。しかし、方式 1 と方式 2 における同期品質に関する心理的尺度値は、方式 3 のそれと同程度の評価値となっている。このことから、被験者はビデオと力覚メディアの同期誤差を知覚しにくかったといえる。

図 14 に、総合品質に関する評価結果を示す。図 14 から、ビデオと音声の  $T_b$  が 20 ms や 40 ms の場合、両作業ともに力覚メディアのソーススキッピングを適用することで QoE の向上が確認できる。これは、力覚メディアのソーススキッピングによりビデオの出力品質が向上したことが大きい。また、力覚 MU 送出レート 300 MU/s の場合においても、総合品質の QoE は向上することを別の実験で確認した。方式 1 や方式 2 は、メディアの性質を考慮しない方式 3 に比べ、高い QoE を達成している。これは、方式 1, 2 が力覚メディア操作感の点で方式 3 よりも有効であることと、方式 1, 方式 2 で発生するメディア間の同期誤差の影響は小さく、QoE 低下の要因になりにくいことが主な原因である。各方式のピーク時の心理的尺度値に注目しても、方式 1, 2 は方式 3 に比べ高い値をとる。また、カスターネットたたき作業での、方式 1, 方式 2 に

おけるビデオ・音声の  $T_b$  が 100 ms を超える場合で QoE が低下している。これは、カスターネットたたき作業が応答性の重要な作業であることに加え、自端末と相手端末におけるカスターネット同士の音ずれも影響していると考えられる。このことは、図 15 に示す楽器音のずれに関する評価結果からも確認できる。

図 14 では、方式 1 と方式 2 では傾向に大きな違いは見られない。しかし、PHANToM の操作精度が必要となるオブジェクト移動作業の場合は、力覚の出力 MU レートを高めることのできる方式 1 がわずかに優れており、PHANToM の操作が簡単で操作速度の速いカスターネットたたき作業では、力覚 MU におけるメディア内同期制御での遅延が最小となる Skipping を用いた方式 2 が有効である。

各作業ごとの作業効率を図 16 に示す。いずれの作業においても、方式 3 の力覚メディア操作感が劣化する部分で、作業効率が低下している。また、カスターネットたたき作業では、被験者間のやり取りが多いため、応答性の低下による作業効率の低下が見られる。

### 7.3 主成分分析結果

実験により得られた多次元的な QoE に対して、QoE に影響を与える主要因を求めるために主成分分析を行った。主成分分析は作業ごとに、3 種類のメディア内同期制御方式をまとめて行った。

分析の結果、オブジェクト移動作業では第二主成分

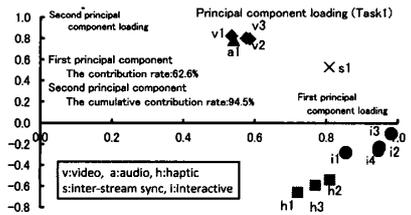


図 17 主成分負荷量 (作業 1: 第一・第二主成分)

Fig. 17 Task1: first and second principal component loading.

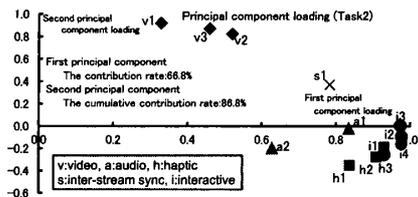


図 18 主成分負荷量 (作業 2: 第一・第二主成分)

Fig. 18 Task2: first and second principal component loading.

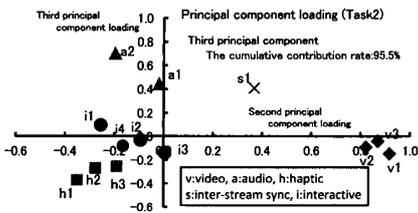


図 19 主成分負荷量 (作業 2: 第二・第三主成分)

Fig. 19 Task2: second and third principal component loading.

までで、カスタネットたたき作業では第三主成分までで累積寄与率が 90% を超えた。そこで、オブジェクト移動作業は第二主成分までの、カスタネットたたき作業は第三主成分までの主成分負荷量を扱う。オブジェクト移動作業における主成分負荷量を図 17、カスタネットたたき作業のそれを図 18、図 19 に示す。

図 17 から、オブジェクト移動作業において QoE に影響を与える主要因は、インタラクティブ性・力覚メディアの品質に関するグループ (i1-i4, h1-h3) と、ビデオと音声の出力品質に関するグループ (v1-v3, a1) の二つに分類されることが分かる。一方、図 18 に示すカスタネットたたき作業の場合では、音声に関する評価尺度 (a1, a2) がインタラクティブ性に関するグループと相関が高くなっている。したがって、オブジェクト移動作業では、音声はビデオと同様 (出力品質が重視) に評価され、カスタネットたたき作業では、インタラクティブ性に関する評価尺度と同様 (応答性が重視) に評価されたといえる。

図 19 より、カスタネットたたき作業における第三

主成分は、楽器音のずれ (a2) に関する評価尺度と強い相関をもっている。このことから、カスタネットたたき作業では楽器音のずれも QoE に影響を与える主要因となることが分かる。

## 8. むすび

本論文では、ビデオ・音声・力覚メディアの各々の性質を考慮した制御により QoE が向上することを実験により多次元的に示した。実験では応答性の違いを考慮した 2 種類の作業を用いた。

主観評価実験の結果、メディア適応型メディア内同期制御と力覚メディアのソーススキッピングは、いずれの作業においても QoE 向上に有効であることが明らかとなった。また、力覚メディアの高い操作精度が必要な作業ではメディア適応型バッファリングが、力覚メディアの操作速度が速い作業では Skipping & buffering が有効であることが示された。更に、応答性が重要ではない作業の場合、ビデオや音声のプレイアウトバッファリング時間を 100 ms や 150 ms に設定しても QoE は低下しないが、楽器を使う作業のように応答性が重要となる作業では、100 ms 未満のプレイアウトバッファリング時間を設定することが望ましい。

今後の課題として、更に異なる種類の作業を行う場合の評価により、提案方式の有効適用領域を明らかにすることがある。また、様々な実験ネットワーク環境を用いて QoE の評価を行うことが考えられる。メディア適応型メディア内同期制御におけるメディア間同期誤差の許容限界やタスク依存性、メディア適応型バッファリングの力覚メディアに設定するプレイアウトバッファリング時間、作業ごとの最適能力覚 MU 送出レートや被験者属性による QoE の違いも今後の検討課題である。

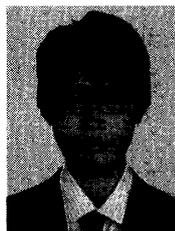
謝辞 本研究の初期段階で御助言頂いた石橋豊教授に感謝する。本研究は、平成 24 年度学術研究助成基金助成金・挑戦的萌芽研究 (課題番号 23656253) によって行われた。

## 文 献

- [1] J. Arata, H. Takahashi, P. Pitakwatchara, S. Warisawa, K. Tanoue, K. Konishi, S. Ieiri, S. Shimizu, N. Nakashima, K. Okamura, Y. Fujino, Y. Ueda, P. Chotiwan, M. Mitsuishi, and M. Hashizume, "A remote surgery experiment between Japan and Thailand over Internet using a low latency CODEC system," Proc. IEEE ICRA2007, pp.953-959, April 2007.

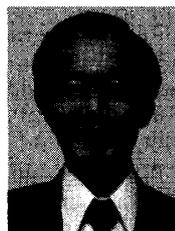
- [2] M. Mansour, M. Eid, and A.E. Saddik, "A multimedia handwriting learning and evaluation tool," Proc. Intelligent Interactive Learning Object Repositories (I2LOR), Nov. 2007.
- [3] J. Marsh, M. Glencross, S. Pettifer, and R. Hubbard, "A network architecture supporting consistent rich behavior in collaborative interactive applications," IEEE Trans. Visualization and Computer Graphics, vol.12, no.3, pp.405-416, May/June 2006.
- [4] ITU-T Recommendation P.10/G.100, Amendment 2, "New definitions for inclusion in Recommendation ITU-T P.10/G.100," July 2008.
- [5] ETSI TR 102 643, "Human Factors (HF); Quality of Experience (QoE) requirement for real-time communication services," Jan. 2010.
- [6] S. Tasaka and Y. Ishibashi, "Mutually compensatory property of multimedia QoS," Conf. Rec. IEEE ICC2002, vol.2, pp.1105-1111, April/May 2002.
- [7] 藤浦豊徳, 孫 樹春, 藤本 猛, 石橋 豊, "力覚メディア通信におけるメディア内同期アルゴリズムの出力品質比較," 信学論 (B), vol.J92-B, no.7, pp.1193-1206, July 2009.
- [8] K. Iwata, Y. Ishibashi, N. Fukushima, and S. Sugawara, "QoE assessment in haptic media, sound, and video transmission: Effect of playout buffering control," Proc. ACM International Conference on Advances in Computer Entertainment Technology (ACE'10), Nov. 2010.
- [9] K. Hikichi, H. Morino, I. Arimoto, K. Sezaki, and Y. Yasuda, "The evaluation of delay jitter for haptics collaboration over the internet," Conf. Rec. IEEE GLOBECOM 2002, vol.2, pp.1492-1496, Nov. 2002.
- [10] T. Kanbara, Y. Ishibashi, and S. Tasaka, "Haptic media synchronization control with dead-reckoning in networked virtual environments," Proc. World Multi-Conference on Systemics, Cybernetics and Informatics (SCI'04), vol.3, pp.158-163, July 2004.
- [11] S. Tasaka and N. Misaki, "Maximizing QoE of interactive services with audio-video transmission over bandwidth guaranteed IP networks," Conf. Rec. IEEE GLOBECOM 2009, Nov./Dec. 2009.
- [12] A. Marshall, K.M. Yap, and W. Yu, "Providing QoS for networked peers in distributed haptic virtual environments," Advances in Multimedia, vol.2008, Article ID 841590, June 2008.
- [13] 伊東祐一郎, 田坂修二, 布目敏郎, 石橋 豊, "音声・ビデオ・力覚メディアインタラクティブ IP 通信におけるプレイアウトバッファリング時間が QoE に及ぼす影響," 信学技報, CQ2010-64, Nov. 2010.
- [14] SensAble Technologies, Inc., "OpenHaptics Toolkit programmer's guide," Version 3.0, 2009.
- [15] 磯村栄一, 田坂修二, 布目敏郎, "ビデオ・音声・力覚メディアインタラクティブ IP 通信における QoE 監視による QoS 制御," 2011 信学ソ大 (通信), B-11-15, Sept. 2011.
- [16] C.E. Osgood, "The nature and measurement of meaning," Psychological Bulletin, vol.49, no.3, pp.197-237, May 1952.
- [17] 伊藤嘉浩, 田坂修二, "IP ネットワーク上での音声・ビデオ伝送におけるユーザレベル QoS の多次元評価," 信学論 (B), vol.J88-B, no.3, pp.689-702, March 2005.
- [18] 金森直希, 田中孝之, 田中一男, "力触覚インターフェースにおける操作感の主観評価," 機械学論 (C), vol.68, no.669, pp.1467-1473, May 2002.
- [19] S. Tasaka and Y. Ito, "Psychometric analysis of the mutually compensatory property of multimedia QoS," Conf. Rec. IEEE ICC2003, pp.1880-1886, May 2003.
- [20] 田中良久, 心理学的測定法 第二版, 東京大学出版会, 東京, 1977.
- [21] F. Mosteller, "Remarks on the method of paired comparisons: III a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed," Psychometrika, vol.16, no.2, pp.207-218, June 1951.

(平成 24 年 4 月 24 日受付, 9 月 20 日再受付)



磯村 栄一 (学生員)

平 23 名工大・工・情報卒。現在, 同大学院博士前期課程在学中。音声・ビデオ・力覚メディアインタラクティブ IP 通信の研究に従事。



田坂 修二 (正員:フェロー)

昭 46 名工大・工・電気卒。昭 51 東大学院博士課程了。工博。同年名工大・情報勤務。平 24 まで同大学院・教授。同年 4 月から名工大名誉教授及び名工大総合工学プロジェクト研究所・プロジェクト教授。昭 59~60 UCLA 客員研究員。マルチメディア通信, 無線ネットワーク, QoE/QoS の研究に従事。IEEE 会員。



布目 敏郎 (正員)

平 10 名工大・工・電気情報卒。平 12 同大学院博士前期課程了。同年バイオニア(株)入社。平 14 名工大・電気情報助手。現在, 同大学院・准教授。工博。マルチメディア転送方式, マルチキャスト通信, 無線ネットワークの研究に従事。IEEE 会員。