

# 音声・ビデオ IP 伝送における QoE 監視技術の研究

田坂 修二 名古屋工業大学大学院工学研究科教授

## 1 はじめに

近年、IP ネットワークの急速な普及と高速化により、これまで放送や専用線システムで提供されていた音声・ビデオ伝送サービスが IP ネットワークによるものに置き換わりつつある。しかし、IP ネットワークは、基本的にはベストエフォートサービスしか提供しないため、パケットの欠落や伝送遅延が起り、その QoS (Quality of Service) は低下する可能性がある。この QoS の低下は、QoE (Quality of Experience) の低下を引き起こす。

本調査研究は、音声・ビデオ IP 伝送における QoE 監視技術の開発を行うものである。QoE の測定・評価とリアルタイム推定・監視との 2 段階での研究実施に加えて、QoE 監視技術の QoE 向上への応用までの研究を行った。前者 2 段階の研究成果は、QoE 監視技術の QoE 向上への応用の研究の中に集約される。そのため、本報告では、QoE 向上への応用の観点から、成果を記述する。

本報告では、まず、第 2 章で QoE 監視技術の QoE 向上への適用例として、QoE ベースビデオ出力方式 SCS を提案する。第 3 章では実験方法を説明し、第 4 章で実験結果と考察を述べる。

## 2 QoE 監視技術の QoE 向上への応用 : SCS

ビデオの IP 伝送においては、パケットの欠落や伝送遅延による QoE の低下を抑えるため、誤り補償とフレームスキップによる出力制御が行われることが多い。

ビデオの誤り補償は、ネットワーク上での情報の欠落や同期はずれによる情報の廃棄に対処し、失われた符号化情報を他の情報から補完するものである。これにより、ビデオフレーム内の一部の画像が欠落しても、ビデオ出力を中断することなく継続することができる。そのため、ビデオの時間品質の低下を防ぐことができる。しかし、補完された画像は、元の画像情報に比べて画質が劣化している場合が多いため、空間品質が劣化する。また、空間品質の劣化したフレームが、次フレームにおける情報補完のための参照フレームとして用いられた場合には、その空間品質劣化が GOP (Group Of Pictures) 単位で伝播する問題が生じる。

一方、フレームスキップ方式は、パケット欠落によってビデオの空間品質が乱れたビデオフレームの出力を行わないものである。このフレーム出力の一時中断は、次の正常な I フレームの出現まで継続する。すなわち、この方式では、情報の欠落によって空間的構造が乱れたビデオフレームは出力せず、高い空間品質を持つフレームのみを出力することができる。しかし、ストリームがフリーズするため、ビデオの時間品質の低下が起こってしまう。

以上より、誤り補償はビデオの時間品質を保持し、フレームスキップはビデオの空間品質を保持することが分かる。従来では、これら 2 方式は別個の問題として取り上げられてきた。

まず、誤り補償によって引き起こされるビデオの空間品質劣化を評価した研究は、主に画像符号化の研究者によって多く発表されている。例えば、文献 [1] は、H. 264 符号化ビデオにおいてマクロブロック毎に重要度を定め、それをもとに行う誤り補償が空間品質に及ぼす影響を、PSNR (Peak Signal to Noise Ratio) を用いて評価している。また、文献 [2] では、H. 264 ビデオ符号化する際、FMO (Flexible Macroblock Ordering) を用いた場合と用いない場合を比較し、FMO の有効性を PSNR によって評価している。これらの研究は、ビデオの空間品質のみを客観的に評価するという目的には意味を持つ。しかし、実際に提供されるサービスは、音声も含めたマルチメディアサービスであるにもかかわらず、これらの研究では音声品質は考慮されていない。また、誤り補償方式を適用せず、フレームスキップ方式によって時間品質を劣化させた場合と比較すると QoE がどのようになるのかは明らかではない。

一方、フレームスキップによって引き起こされるビデオの時間品質の劣化を評価した論文は、主に通信分野で発表されている。例えば、筆者は、文献 [3] において音声・ビデオ IP 伝送を対象として QoE リアルタイム推定方法を提案している。そして、機械的に測定可能なアプリケーションレベル QoS パラメータの値か

ら QoE パラメータの値を推定している. 文献 [1] や文献 [2] と異なる点は, ビデオの時間品質だけでなく, 同時に引き起こされる音声品質の劣化も考慮し, QoE の定量的な評価を行っている点である. しかし, ビデオの誤り補償は行っておらず, ビデオの空間品質は評価されていない.

次に, 筆者は, 文献 [4] において, ビデオの空間・時間品質及び音声品質を考慮に入れて QoE を評価している. 文献 [4] では, ビデオの空間・時間品質と音声品質から高い精度の QoE 推定を実現している. これは, QoE 保証実現への大きな一歩である. しかし, 文献 [4] では QoE 向上技術に関する検討は行われていない.

そこで, 筆者は, 本研究において, QoE 監視技術の QoE 向上への応用を考え, 受信側でのビデオ出力制御のために, ビデオの誤り補償とフレームスキップを切り替える方式 SCS(Switching between error Concealment and frame Skipping)を提案した [5]. SCS では, 1 ビデオフレーム内における欠落したスライスの割合を誤り補償率 (単位: %) とし, この誤り補償率と閾値とを比較して, ビデオの誤り補償とフレームスキップとを切り替える.

図 1 に SCS 動作のフローチャートを示す.

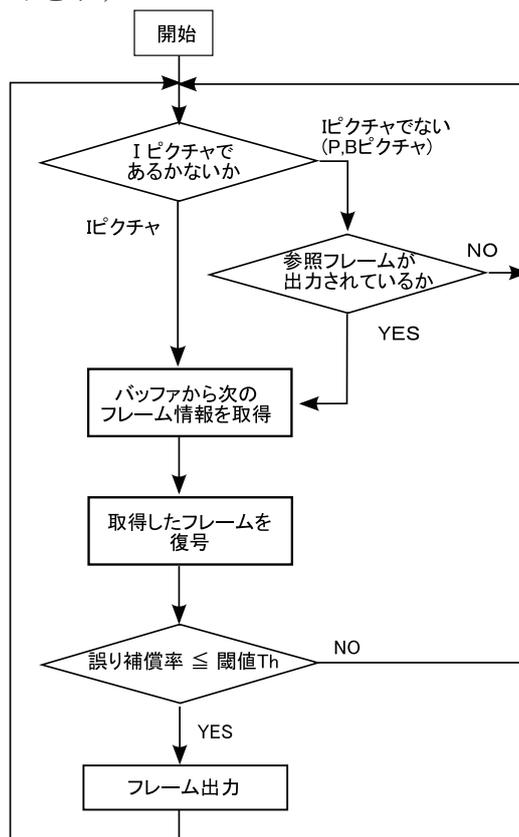


図 1 : SCS 動作のフローチャート

SCS では, 1 ビデオフレームにおいて誤り補償率と閾値とを比較し, 誤り補償とフレームスキップとを切り替える. 誤り補償率は, 1 ビデオフレームにおいて誤り補償された割合を示し, 式 (1) で定義される.

$$\text{誤り補償率} = \frac{\text{誤り補償されたスライス数}}{\text{1ビデオフレーム全体のスライス数}} \times 100[\%] \quad (1)$$

図 1 に示されるように, まず, 受信されたスライスを格納するバッファから, 次に出力する予定のフレームを構成するスライスを取得し, 復号する.

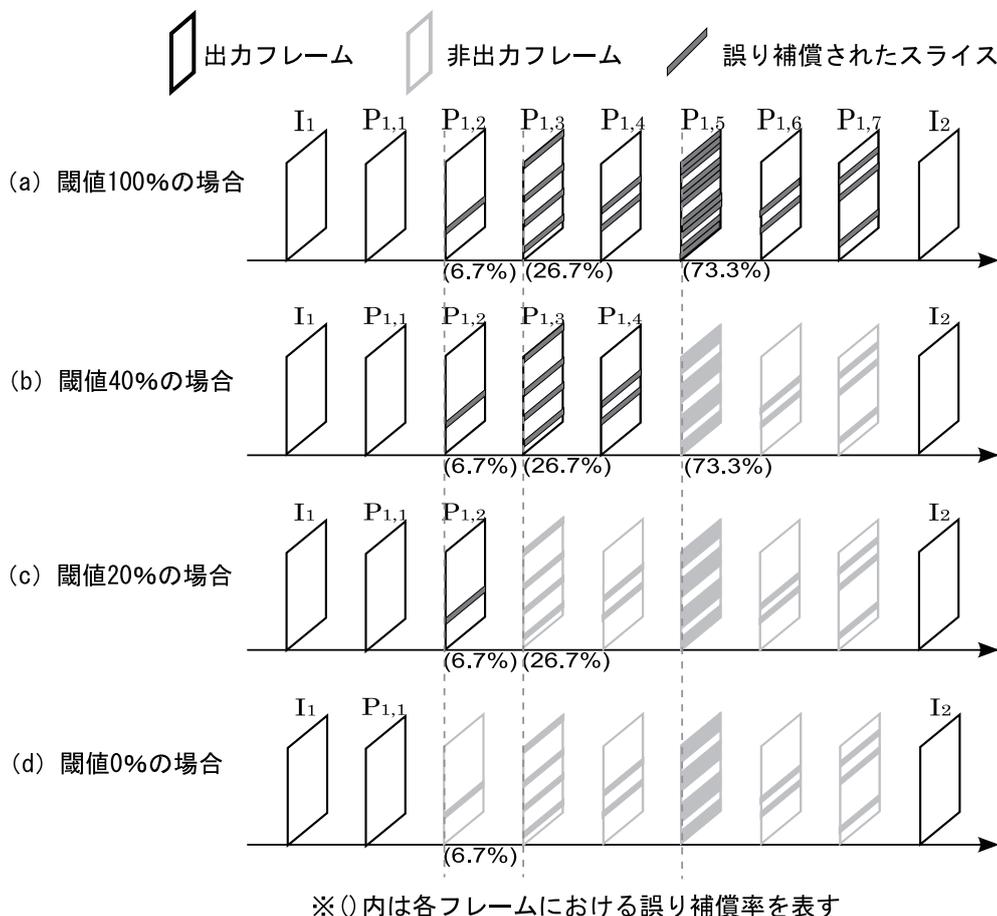
そして, 復号されたフレーム内で欠落のあったスライスに対して誤り補償を行う. その際, 誤り補償方式は特定のものに限定せず, 誤り補償された割合が算出できるものであれば何でもよい.

求めた誤り補償率が予め設定された閾値より大きい場合, このフレームの出力を取り消し, 非出力フレームとする. 誤り補償率が閾値以下の場合でも, 復号対象フレームの動き予測参照フレームが非出力フレームであると, 同様に出力を取り消し, 非出力フレームとする.

誤り補償率が閾値以下で, 且つ対象フレームの動き予測参照フレームが出力フレームである場合に, フレームを出力する. この操作を 1 フレーム毎に行い, 誤り補償とフレームスキップとを切り替える.

スライスグループマップタイプをインタリーブとして、閾値を 100%、40%、20%、0%と設定した場合の動作例を図 2 に示す。図中の実線の平行四辺形が出力フレームを示し、淡い実線の平行四辺形が非出力フレームを表す。

誤り補償されたスライスは、グレーの小さな平行四辺形で示す。  
 ま図中の () 内の数値は、そのフレームの誤り補償率を指す。



※ () 内は各フレームにおける誤り補償率を表す

図 2：閾値別の動作例

SCS の研究において、まず、多数のコンテンツ及びピクチャパターンの組み合わせに対して、4 種類の閾値を適用して、QoE を定量的に評価した。そして、コンテンツタイプやピクチャパターンによって、最適な QoE を導く閾値が異なることを明らかにした。従って、SCS が高効率で動作するためには、適切な閾値の選択が肝要である。この選択のために、QoE 推定による監視方式を用いて、それが有効であることを示した [5]。なお、この評価においては、FMO は用いられなかった。

H. 264 のビデオ圧縮符号化技術においては、FMO のスライスグループマップタイプとして、インタリーブをはじめ他に 6 種類のものを用いることができる。出力されたビデオは、このマップタイプによって画像劣化の現れ方が異なる。

本研究では、QoE 向上技術を更に発展させるため、SCS において FMO を用いた場合の検討も行う。インタリーブ、フォアグラウンド/レフトオーバ (以下、フォアグラウンドと呼ぶ) の 2 種類のマップタイプを取り上げる。そして、これらのマップタイプが QoE にどのような影響を及ぼすかを調査する。また、マップタイプの違いが SCS の閾値設定に与える影響も調べる。

### 3 実験方法

本実験では、SCS における閾値を変化させて QoE を評価する。そして、求めた QoE 評価結果より、閾値の変化が QoE に及ぼす影響を考察する。以下に実験システムと QoE 評価実験方法を示す。

#### 3-1 実験システム

図 3 に実験ネットワーク構成を示す。全ての回線は、100Mbps の Ethernet である。ルータとして、Alcatel Lucent 社製の RS3000 を用いる。このネットワークにおいて、音声とビデオは、メディア送信端末からメディア受信端末へ伝送される。その際、メディア受信端末では、伝送遅延の揺らぎを吸収するため、1 秒間のプレイアウトバッファリング制御を行う。

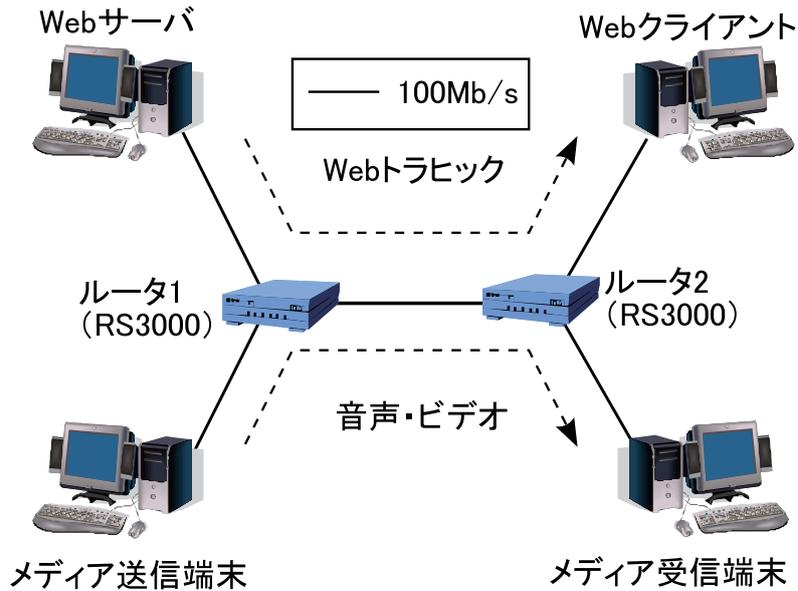


図 3 : 実験ネットワーク構成

音声とビデオの仕様を表 1 と表 2 に示す。表中の MU とは、メディア同期の処理単位である。1MU は、音声では 20 ミリ秒分のサンプリングデータに相当し、ビデオでは 1 フレームに相当する。ビデオの符号化には H. 264 (JM13.2) を用いる。

表 1 : 音声・ビデオの仕様

音符号化方式	Linear PCM 24kHz 16bit 1ch
音声平均 MU レート [MU/s]	50
音声平均ビットレート [kb/s]	384
ビデオ符号化方式	H. 264 (JM13.2) 320×240
ビデオ分割スライス数	15
ビデオ平均 MU レート [MU/s]	30
ピクチャパターン	I IPPPP IPPPPPPPPPPPPP
スライスグループマップタイプ	インタリーブ フォアグラウンド
スライスグループ数	5
再生時間 [s]	10

SCS は、第 2 章で述べた手順に従って動作する。その際の閾値は、第 2 章の例と同じ、100%、40%、20%、0%とする。

ピクチャパターンの違いとして、I, IPPPP, IPPPPPPPPPPPPPP の 3 種類を設定する。

負荷として、図 3 のネットワークにおいて、Web サーバから Web クライアントへ Web トラフィックを伝送する。Web クライアントには、Web サーバの性能評価に用いられる WebStone を使用する。Webstone に設定するクライアント数は、20, 30, 40, 50, 75, 100 の計 6 種類とする。

ビデオの誤り補償方式には、JM13.2 に実装されているものを用いる。I フレームにおける誤り補償では、欠落情報を周辺の情報から補間する。P フレームにおける誤り補償には、前出力フレームから欠落部分を複写する方式 (Frame Copy) を用いる。

コンテンツの選択には、VQEG のテストプランを参考にした。本検討では、2 種類のコンテンツタイプを選び、その各々に対して 2 個のコンテンツを用意して、計 4 個のコンテンツを用いた。

インタリーブとフォアグラウンドとでは、表 2 から分かるように、ピクチャパターンが同じならば、同程度の平均ビットレートを示す。このことから、マップタイプが異なっても、同じピクチャパターンならば公平な比較ができることが分かる。

表 2 には、参考のため、全てのコンテンツに対して ITU-T P.910 で提唱されている TI (Temporal Information) 値を示しておく。これは、動きの程度を表し、数値が大きいほど動きが大きいものである。なお、この TI 値は、シーンチェンジを除いて算出したものである。

各コンテンツのシーン内容を表 3 に示す。

表 2 : 各コンテンツのビデオ符号化ビットレートと TI 値

コンテンツ名	ピクチャパターン	平均ビットレート [kb/s]		TI 値
		インタリーブ	フォアグラウンド	
sport 1	I	2592.713	2659.274	16.652
	IPPPP	1243.395	1252.870	
	IPPPPPPPPPPPPP	1040.902	1044.835	
sport 2	I	2241.426	2315.071	55.773
	IPPPP	1228.215	1243.100	
	IPPPPPPPPPPPPP	1076.502	1085.263	
music video 1	I	1979.334	1973.633	12.924
	IPPPP	846.506	838.270	
	IPPPPPPPPPPPPP	678.828	674.535	
music video 2	I	1658.756	1686.084	48.597
	IPPPP	790.520	799.802	
	IPPPPPPPPPPPPP	690.626	688.885	

表 3 : 各コンテンツのシーン内容

コンテンツ名	シーン内容
sport 1	コーチの動きを真似るように複数の人が室内トレーニングをしている。
sport 2	レーシングカーが走っている。シーンチェンジ有り。
music video 1	一人の男性アーティストがウクレレを弾いている。
music video 2	女性歌手がピアノを弾き、踊っている。シーンチェンジ有り。

各スライスグループマップタイプにおいて、1 ビデオフレームは図 4 と図 5 に示すようにスライスグループを設定している。

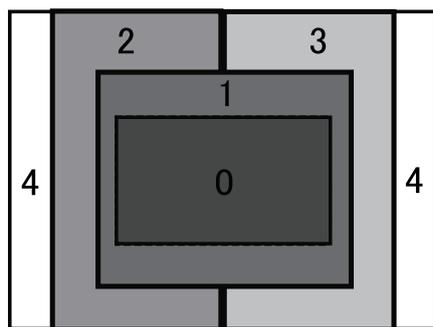


図 4：フォアグラウンド

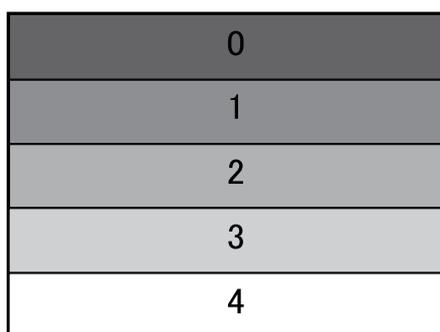


図 5：インタリーブ

本実験では、スライスグループの番号が若い順に重要度の高いスライスとして送信する。そのため、フォアグラウンドでは、画面中央の画質劣化を抑えることができる。一方、インタリーブでは、主に画面下部に帯状に直線的な空間品質劣化が起こりやすい。

### 3-2 QoE 評価実験方法

メディア受信端末で出力された音声・ビデオを記録し、これを刺激として、QoE 評価を行う。

QoE を定量的に評価する方法として、計量心理学的測定法の一つである系列カテゴリ法を用いる。系列カテゴリ法とは、評定尺度法から得られた評価値に対してカテゴリ判断の法則を適用し、評価対象（刺激）の尺度を得るものである。

評定尺度法とは、被験者の判断基準に従って、各刺激をいくつかのカテゴリに分類させるものである。本研究では、カテゴリとして、表 4 に示す 5 段階品質尺度（Absolute Category Rating：ACR）を用いる。そして、その評価値は、予めカテゴリに付与された数値を用いて算出される。

表 4：5 段階品質尺度

評価値	カテゴリ
5	非常に良い
4	良い
3	普通
2	悪い
1	非常に悪い

この評価値の平均をとったものが、平均オピニオン評点（MOS：Mean Opinion Score）である。MOS は、単一メディアの主観評価によく用いられる。しかし、各カテゴリの心理的な境界が等間隔であるという保証

がないため、評定尺度法によって得られる結果は順序のみを示す尺度（順序尺度）である。そのため、単純に平均をとるといった処理は、主観評価値を正確に表現することはできない。

そこで、系列カテゴリ法では、カテゴリ判断の法則を用いて、評定尺度法によって得られた順序尺度から距離尺度を得る。距離尺度では、尺度の等間隔性が保証されているため、多くの統計処理を行うことができる。また、カテゴリ判断の法則は、多くの仮定から成り立つため、得られた結果の妥当性を確認しなければならない。

本報告では、系列カテゴリ法により求められた尺度値に対して、Mosteller の適合度検定を行う。これにより、得られた尺度の適合性が確認できれば、これを心理的尺度と呼び、QoE パラメータとして用いる。

被験者として、まず、20代の学生18人を用いた。加えて、10代から20代の学生24名にも評価を依頼した。評価対象（刺激）数は、コンテンツが4個、スライスグループの違いが2種類、切り替えの閾値が4種類、ピクチャパターンが3種類、Webクライアント数が6種類の計576（=4×2×4×3×6）個とする。さらに、評価対象とは別にダミーデータ56個を加え、計632個を被験者に提示する。各コンテンツの提示順序は、被験者によってランダムとする。また、個々のコンテンツを評価する際、スライスグループ2種類、切り替えの閾値、ピクチャパターンの種類、及びWebクライアント数の組み合わせの提示順番もランダムとする。

一つの刺激の再生時間を10秒とし、1人当りの評価時間は休憩を含めておよそ3時間であった。

## 4 実験結果

### 4-1 QoE 評価結果

系列カテゴリ法により得られた尺度値に対し、Mosteller の適合度検定を行った。その結果、有意水準5%において、系列カテゴリ法により得られた尺度値が測定結果に適合しなかった。そこで、576個の評価対象の内、推定値と実測値の誤差が大きいものから順に、刺激を一つずつ取り除いた。その結果、33個の刺激を取り除いたことで、有意水準5%において、得られた尺度値が測定結果に適合した。この尺度値をQoEパラメータである心理的尺度として用いる。

マップタイプの違いによるQoEへの影響を分かりやすくするため、図6から図9に、インタリーブとフォアグラウンドのQoE評価結果を、コンテンツ、ピクチャパターン、及び閾値の組み合わせ毎に示す。閾値を高い値に設定した場合に、マップタイプの違いがQoEへ及ぼす影響が現れるため、閾値が40%と100%の結果のみを示す。

ここで、図中の点線は、各カテゴリーの下限值を指す。今回、QoEパラメータの最小値が1となるように尺度の原点を定めた。その結果、各カテゴリーの下限值は、5.277（カテゴリー5）、4.364（カテゴリー4）、3.414（カテゴリー3）、2.187（カテゴリー2）となった。

なお、図中でデータが欠落している箇所は、検定で取り除いた刺激に対応している。

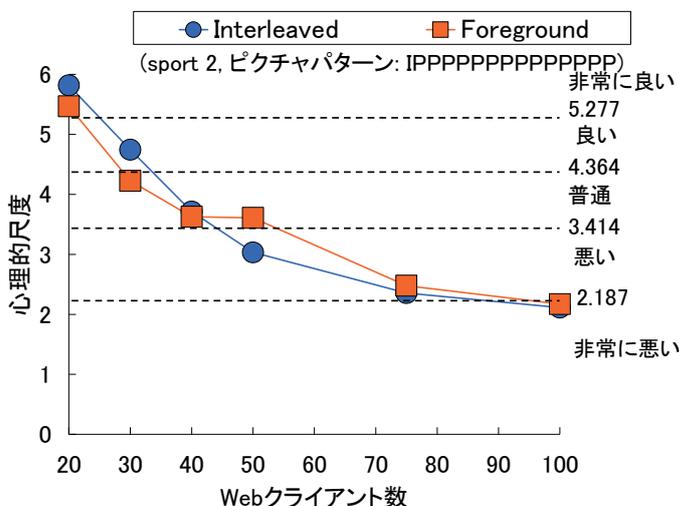


図6：心理的尺度(sport 2，閾値40%，ピクチャパターン：IPPPPPPPPPPPPP)

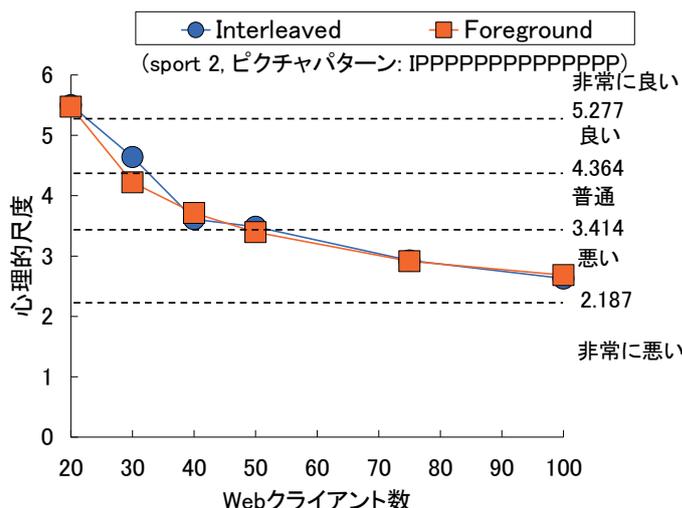


図 7：心理的尺度(sport 2 , 閾値 100%, ピクチャパターン: IPPPPPPPPPPPPPP)

図 6 から図 9 より, Web クライアント数が 20 の場合, どの図においても, 閾値やマップタイプの違いに関わらず, 心理的尺度は「非常に良い」または「良い」のカテゴリとなっている. アプリケーションレベル QoS 測定結果では, Web クライアント数 20 の場合, 閾値やマップタイプの違いによる QoS パラメータ値の差はみられなかった. このことから, Web クライアント数 20 の場合において QoE の違いがみられるものは, 測定値の揺らぎが生じたためと考えられる.

図 6 及び図 7 より, コンテンツ sport 2, ピクチャパターン IPPPPPPPPPPPPPP では, 閾値 40% で Web クライアント数 50 の場合を除き, インタリーブはフォアグラウンドに比べて同等以上の心理的尺度を示すことが分かる. この理由を以下に述べる.

sport 2 のシーン内容は, レーシングカーが左右に大きく動くものである. 本実験で設定したフォアグラウンドでは画面の両端に空間品質劣化が目立つため, 左右に動きのあるコンテンツでは心理的尺度は低下したと考えられる. 今回設定しているインタリーブでは画面の両端よりも画面下部に空間品質劣化が表れることが多いため, 心理的尺度が低下しにくかったと考えられる. 図には示していないが, ピクチャパターンが IPPPP の場合においても, 同様の傾向がみられた.

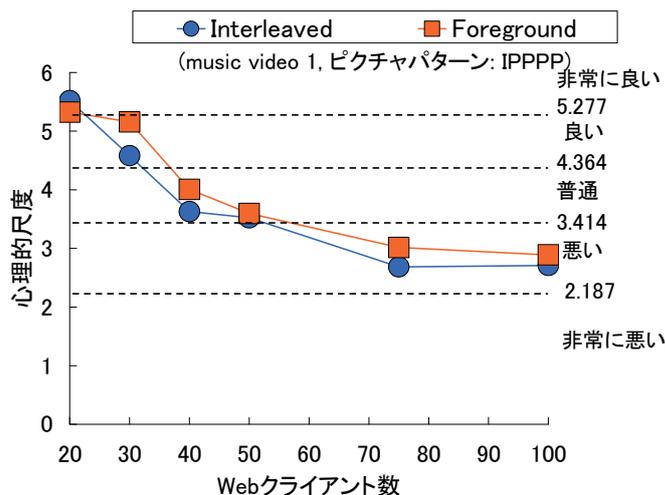


図 8：心理的尺度(music video 1, 閾値 40%, ピクチャパターン: IPPPP)



一方、フォアグラウンドの場合は、負荷が低ければ、演奏者の部分の空間品質劣化がほとんど起こらず、QoE の低下を防ぐ。そのため、Web クライアント数が 30 や 40 といった低負荷の場合では、閾値 40%として、時間品質劣化を抑えた方が高い QoE を示した。

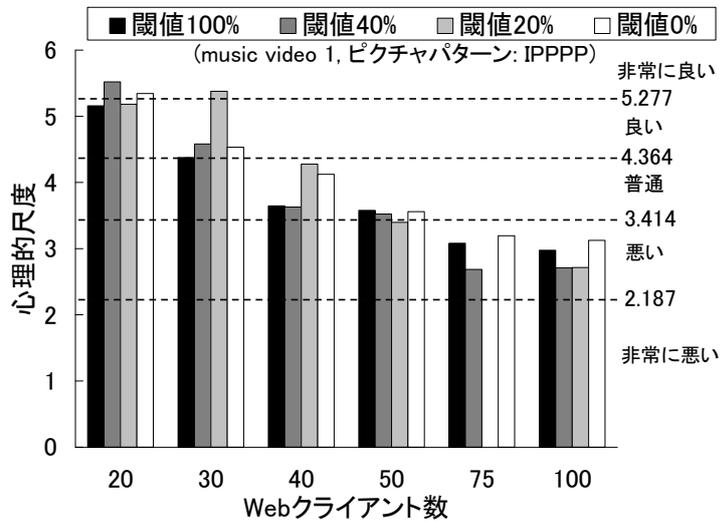


図 10 : 心理的尺度 (music video 1, インタリーブ, ピクチャパターン: IPPPP)

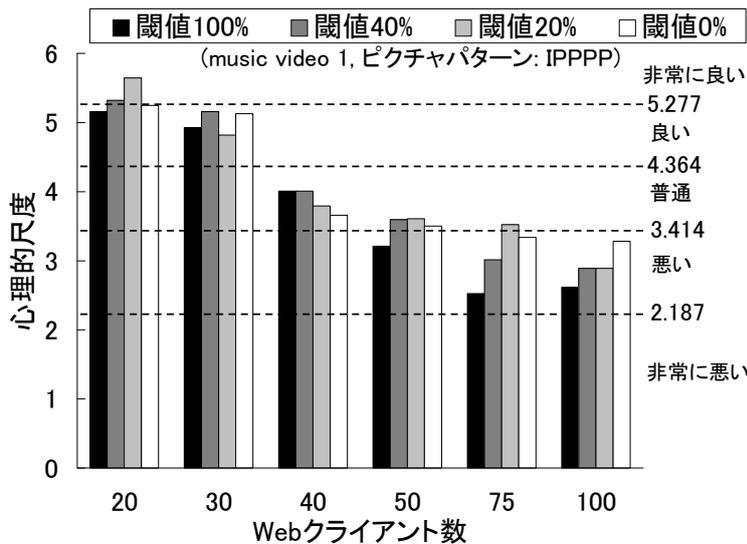


図 11 : 心理的尺度 (music video 1, フォアグラウンド, ピクチャパターン: IPPPP)

図には示していないが、文献 [5] の結果と同様に、スライスグループやコンテンツに関わらず、ピクチャパターンが I のみの時は閾値 0%、ピクチャパターンが IPPPPPPPPPPPPPPPPPPPP の時は閾値 100%が、それぞれ最も高い QoE を導いた。

なお、コンテンツタイプがスポーツの場合では、マップタイプの違いによる最適な閾値への影響は小さかった。

## 5 結論

本研究では、音声・ビデオ IP 伝送において、QoE 監視技術の QoE 向上への適用を中心課題とした。まず、QoE ベースビデオ出力方式 SCS を提案した。SCS が高効率で動作するためには、適切な閾値の選択が肝要であ

る。この選択のために、QoE 推定による監視方式を用いて、それが有効であることを示した。

次に、スライスグループマップタイプとしてインタリーブ及びフォアグラウンドの2種類を用いて QoE を評価した。その結果、スポーツのコンテンツでは2方式の差は小さかった。しかし、ピクチャパターンが IPPPP と IPPPPPPPPPPPPPP の時、画面中央で動きの少ないコンテンツにおいてでは、フォアグラウンドの有効性が確認された。

また、マップタイプの違いとして、インタリーブとフォアグラウンドを取り上げ、両方式の違いが SCS の閾値設定に及ぼす影響を調査したところ、ピクチャパターンが IPPPP で動きの小さいコンテンツでは、最適な閾値が異なることが分かった。

## 【参考文献】

- [1] S.K.Im and A.J.Pearmain.“An optimized mapping algorithm for classified video transmission with the H.264 flexible macroblock ordering” in IEEE International Conference on Image Processing, pp.1445 - 1448, July 2006
- [2] Y. Dhondt and P. Lambert. “Flexible macroblock ordering: An error resilience tool in H.264/AVC.” Abstracts of the Fifth FTW Ph.D. Symposium, (106), December 2004.
- [3] S. Tasaka and Y. Ito, “Real-time estimation of user-level QoS of audio-video transmission over IP networks,” Conf. Rec. IEEE ICC2006, June.
- [4] S. Tasaka and Y. Watanabe, “Real time estimation of user-level QoS in audio-video IP transmission by using temporal and spatial quality,” Conf. Rec. IEEE GLOBECOM, Nov. 2007.
- [5] S. Tasaka and H. Yoshimi and A. Hirashima and T. Nunome, “The effectiveness of a QoE-based video output scheme for audio-video IP transmission,” Proc. ACM Multimedia, Oct. 2008.

## 〈発表資料〉

題 名	掲載誌・学会名等	発表年月
The Effectiveness of a QoE-Based Video Output Scheme for Audio-Video IP Transmission	Proc. ACM Multimedia2008	2008年10月
音声・ビデオ IP 伝送における QoE ベースビデオ出力方式の閾値設定方法が QoE に及ぼす影響	電子情報通信学会ソサイエティ大会講演論文集 B-11-5	2008年9月
QoE ベースビデオ出力方式を用いた音声・ビデオ IP 伝送におけるスライスグループマップタイプが QoE に及ぼす影響	電子情報通信学会総合大会講演論文集 B-11-19	2009年3月
音声・ビデオ IP 伝送におけるディスパーススライスグループマップタイプが QoE ベースビデオ出力方式の閾値設定に及ぼす影響	電子情報通信学会総合大会講演論文集 B-11-22	2009年3月