

映像配信サービスを対象としたマルチメディア品質推定モデル

富永 聡子^{†a)} 山岸 和久[†] 林 孝典[†] 高橋 玲[†]

Multimedia Quality Estimation Model for Video Streaming Services

Toshiko TOMINAGA^{†a)}, Kazuhisa YAMAGISHI[†], Takanori HAYASHI[†],
and Akira TAKAHASHI[†]

あらまし IP ネットワークのブロードバンド化や符号化技術の進展に伴い、高品質な映像配信サービスを提供できる環境が整ってきた。映像配信サービスを快適な品質で提供するためには、ユーザが体感する品質 (QoE: Quality of Experience) に基づいたサービス設計・管理が重要となる。本論文では、HDTV 解像度の映像配信サービスに対する QoE 評価法を構築するため、オーディオ及び映像の個別メディア品質評価値から、これらを統合したマルチメディア (オーディオビジュアル) 品質を推定する計算モデルを提案する。まず、提案モデルの入力となる個別メディア品質を評価する方法について検討した。その結果、マルチメディア信号を用いて個別メディア品質を評価する際には、個別メディア間に相互作用が存在するため、評価者には各メディアを単独に提示・評価する必要があることが分かった。次に、個別メディア品質とマルチメディア品質の対応関係を回帰分析により導出し、品質推定モデルを構築した。更に、提案モデルを実映像配信システムや未知コンテンツを用いた品質推定に適用し、マルチメディア品質が高精度に推定できることを検証した。

キーワード ユーザ体感品質、マルチメディア品質、オーディオ品質、映像品質、HDTV

1. ま え が き

IP ネットワークのブロードバンド化や符号化技術の進展に伴い、高品質な映像配信サービスを提供できる環境が整ってきた。映像配信サービスを快適な品質で提供するためには、音声・音響メディアや映像メディアに対するユーザ体感品質 (QoE: Quality of Experience) [1] を適切に設計・管理することが重要である。

メディア品質を定量化する手法は、人間の知覚品質を視聴覚心理評価実験により定量化する主観品質評価が基本である。しかしながら、主観品質評価の実施には多大な時間・労力を要し、映像配信サービスの品質評価・設計の効率化や品質監視・管理へ適用するためには、映像配信サービスにかかわる物理的な特徴量 (例えば、メディア信号、ビットレート、IP パケット転送品質等) から QoE を推定する客観品質評価技術の確

立が必要となる。

従来、音声・音響メディアや映像メディア等、個別メディア品質に対する客観評価法の検討が盛んに行われている。客観品質評価技術は、利用シーンや入力情報の違いごとに技術検討が進められ、①プランニングモデル、②メディアレイヤモデル、③パケットレイヤモデル、④ビットストリームモデル、⑤ハイブリッドモデルに大別される [2], [3]。それぞれ、①ネットワークやアプリケーション/端末の品質設計・管理パラメータ情報、②メディア信号情報、③パケットヘッダ情報、④符号化後のビットストリーム情報 (ペイロード情報)、⑤前記情報を組み合わせた情報、を入力としてユーザ体感品質を推定するモデルである。この中で、プランニングモデルとメディアレイヤモデルの一部は国際標準化された技術として確立している。具体的には、音声メディアに対するプランニングモデルの代表例として、IP 電話サービスの QoE を推定する E-model [4] が挙げられる。E-model は、端末要因・環境要因・ネットワーク要因などに関する複数パラメータを入力し、R 値を出力する。また、E-model はパケット損による音声品質劣化と遅延時間の相互作用を正確に評価できないという問題点があるため、こ

[†] 日本電信電話株式会社サービスインテグレーション基盤研究所，武蔵野市

NTT Service Integration Laboratories, NTT Corporation,
3-9-11 Midori-cho, Musashino-shi, 180-8585 Japan

a) E-mail: toshiko.tominaga@lab.ntt.co.jp

れを改善した方法も提案されている [5]。音響メディアレイヤモデルの代表例としては、電話帯域音声信号の符号化やパケット損失の影響を考慮した客観品質評価法である PESQ (Perceptual Evaluation of Speech Quality) [6]、これを広帯域音声信号の評価に拡張した Wideband-PESQ [7]、音響信号の符号化品質を評価対象とした PEAQ (Perceptual Evaluation of Audio Quality) [8] 等がある。一方、映像メディアに対するメディアレイヤモデルには、SDTV 映像の符号化品質を評価対象とした ITU-T 勧告 J.144 [9] が存在する。また、携帯端末や PC を想定し、QCIF～VGA 映像の符号化及び IP パケット損などのネットワーク品質劣化を考慮した映像品質を評価可能な ITU-T 勧告 J.246 [10]、J.247 [11] をはじめ、文献 [12]～[16] でも検討されている。

このように、音声・音響や映像の個別メディアに対する客観品質評価法については確立されつつある。しかしながら、個別メディアを統合したマルチメディア品質をどのように取り扱うかについては、マルチメディア品質推定モデルのフレームワークが ITU-T 勧告 J.148 [17] において記載されているものの、具体的な計算モデルの規定までには至っていない。ここで、マルチメディア品質とは、音声・音響品質と映像品質を統合したオーディオビジュアル品質のことである。マルチメディア品質推定モデルの従来の検討例としては、Hands によってマルチメディア品質推定モデルの基礎検討が行われており [18]、オーディオ品質、映像品質及び両者の相互作用を考慮したモデルが有効であると報告されている。田坂らの検討においても個別メディアの相互作用の効果が検討され、心理的距離尺度を用いた品質推定モデルが提案されている [19]。また、コンテンツやタスクの違いにより、オーディオ品質及び映像品質のマルチメディア品質への寄与度が異なる結果も報告されている [18], [20], [21]。その他、映像の劣化箇所がマルチメディア品質に与える影響の評価を検討した例 [22] や、モデルの入力としてパケットレイヤモデルによりオーディオ品質及び映像品質を推定し、それらを用いてマルチメディア品質を推定する検討も行われている [23], [24]。テレビ電話を対象としたマルチメディア品質推定モデルとして、ITU-T 勧告 G.1070 [25] がある。このモデルは、コンテンツとして人物像を想定したテレビ電話サービスに特化した品質推定モデルであるため、映像配信サービスを対象とした場合には必ずしも適用できるとは限らない。

本論文では、HDTV 解像度の映像配信サービスに対する品質評価法を構築することを目的として、オーディオ及び映像の個別メディア品質評価値から、これらを統合したマルチメディア品質を推定するマルチメディア品質推定モデルについて検討する。前述のマルチメディア品質推定モデルに関する既存検討の結果を踏まえ、(1) 評価対象となる映像解像度やコンテンツ/タスクにより計算モデルが変化しているが、今回評価対象とする高品質な映像配信サービスにおいてどのような品質推定モデルを構築することが適切であるか、(2) 従来検討では個別メディア品質を独立に制御して品質推定モデルを検討しており、実映像配信システムで発生する品質劣化パターンとは必ずしも一致しない場合があるが適用可能であるか、(3) モデル構築時とは異なる評価コンテンツ・評価条件に対しても統一的な品質推定モデルが有効であるかどうか、という点が課題となる。

以下、まず 2. では、本論文で前提とするマルチメディア推定モデルの構成要素について述べる。3. では、マルチメディア品質推定モデルの入力値となるオーディオ品質及び映像品質を求めるための主観品質評価法の検討を行う。具体的には、一つの評価実験内で個別メディア品質とマルチメディア品質の主観評価値を同時に求める方法とメディア品質ごとに評価する方法を比較し、提案モデルの構築に適した主観品質評価法を明確にする。次に 4. では、個別メディア品質とマルチメディア品質の対応関係を重回帰分析により導出し、マルチメディア品質推定モデルを構築する。更に 5. では、構築したマルチメディア品質推定モデルをモデル構築時とは異なる実映像配信システムあるいは未知コンテンツに対する品質推定に適用し、その有効性を検証する。最後に 6. でまとめ及び今後の課題について述べる。

2. 前提とするマルチメディア品質推定モデル

マルチメディア品質推定モデルとして、ITU-T 勧告 J.148 に記載されている図 1 のフレームワークに基づいて検討を行う。

J.148 では、オーディオ信号及び映像信号からオーディオ品質 A_q 及び映像品質 V_q を推定するとともに、オーディオ信号と映像信号の遅延差を算出する。次に、これらの値とタスクを考慮したマルチメディア品質統合関数により、マルチメディア品質を推定している。

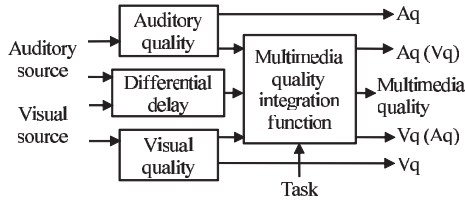


図 1 J.148 マルチメディア品質推定モデル
Fig.1 J.148 Multimedia quality estimation model.

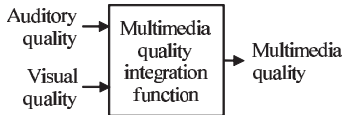


図 2 本論文で取り扱うマルチメディア品質推定モデル
Fig.2 Multimedia quality estimation model.

マルチメディア品質統合関数は、マルチメディアサービスにおける品質判断においてオーディオ及び映像の知覚・認識過程を統合するものである。J.148 で示されるマルチメディア品質統合関数は、マルチメディア品質、映像品質が考慮されたオーディオ品質 $A_q(V_q)$ 、及びオーディオ品質が考慮された映像品質 $V_q(A_q)$ の三つの品質評価値を出力する。

J.148 で示されている構成要素のうち、オーディオ信号及び映像信号の遅延差については、映像配信サービスを対象とした場合、十分な受信バッファによりオーディオと映像のメディアレベルのずれを吸収できる場合が多い。このため、本論文では同期ずれについては品質劣化要因の対象外とする。

以上のことを考慮し、図 2 に示すフレームワークで検討することとした。マルチメディア品質推定モデルの入力はオーディオ及び映像の個別メディア品質評価値であり、出力はマルチメディア品質である。今回の検討では入力となる映像品質やオーディオ品質は主観評価値を用いて定式化及び有効性検証を実施する。この理由は、個別メディアの客観品質評価値を入力として用いる場合、入力時点で個別メディアに対する主観品質評価値の推定誤差が含まれるため、マルチメディア品質統合関数で考慮すべき要因を適切にとらえることができない可能性を回避するためである。以下では、マルチメディア品質統合関数をマルチメディア品質推定モデルという。

インサービス品質管理への適用を考えた場合、入力は主観品質評価値ではなく客観品質評価値を用いてもかまわない。例えば、音声入力には E-model や PEAQ

の適用が考えられ、映像入力には J.246 や J.247 等の客観品質評価技術が適用可能となる。個別メディアの客観品質評価値を入力とした場合には、前述のとおり、客観品質評価値に含まれる誤差がマルチメディア品質推定に与える影響の程度を検証する必要があるが、この点は今後の課題として取り扱うこととする。

3. マルチメディア品質に対する主観品質評価法

この章では、オーディオ品質、映像品質及びマルチメディア品質を求める主観品質評価法の検討を行う。

マルチメディア品質の主観評価法として、双方向映像通信サービスを対象とした主観評価法である ITU-T 勧告 P.920 [26] では、Appendix に評価用紙が記載されている。この評価用紙では、評価者はマルチメディア品質、映像品質、及びオーディオ品質の三つの評価を順次答える形式となっている。

この評価法で得られるオーディオ品質及び映像品質の評価値は、図 1 に対応づけると $A_q(V_q)$ と $V_q(A_q)$ である。本論文では、マルチメディア品質推定モデルの入力として、 A_q 及び V_q を想定している。

そこで、オーディオ信号のみ、映像信号のみ、マルチメディア信号に対する品質評価をそれぞれ独立して実施する場合と、マルチメディア信号の提示後、オーディオ品質、映像品質及びマルチメディア品質の三つの評価を一つの実験内で同時に得る場合とで、評価特性の違いが見られるかどうかを明らかにするため主観評価実験を実施した。

3.1 主観評価実験条件

主観評価実験は、評価法 A：オーディオ信号と映像信号で構成されるマルチメディア信号を提示し、オーディオ品質、映像品質、及びマルチメディア品質の三つの評価を同時に実施する方法（ITU-T 勧告 P.920 Appendix 記載の評価方法）、評価法 B：オーディオ信号のみを提示してオーディオ品質を評価する実験、映像信号のみを提示して映像品質を評価する実験、マルチメディア信号を提示してマルチメディア品質を評価する実験の三つを実施する方法、の 2 通りを実施した。以下に各評価法の特徴を述べる。

評価法 A は、一度に三つの観点：オーディオ品質、映像品質、及びマルチメディア品質で品質評価を行うため、効率的に評価を行える長所がある。一方、評価者はそれぞれの体感品質を同時に評価するため、各評価結果間の対応関係を意識して採点してしまう危険が

表 1 評価環境

Table 1 Subjective assessment test environment.

Video monitor	32-inch CRT monitor SONY BVM-D32E1WJ
Viewing distance	3H (H : picture height) about 110 cm
Illumination intensity	About 30 lx
Listening equipment	Stereo loud speakers ONKYO GX-77M
Ambient noise at receiving side	Hoth noise at 35 dB (A)

表 2 符号化条件

Table 2 Codec conditions.

Video codec	H.264 High profile level 4.0
Group of picture (GOP)	M = 3, N = 15
Frame rate	30 fps
Video resolution	1440 × 1080
Audio codec	MPEG-2 audio AAC-LC

表 3 コンテンツ条件

Table 3 Contents conditions.

	Video	Audio
1	European market (No.7 [27])	BGM (No.7 [27])
2	Harbor scene (No.12 [27])	BGM (No.30 [27])
3	Crowded crosswalk (No.30 [27])	Pop [28]
4	Green Leaves (No.23 [27])	Piano [28]
5	Weather report (No.39 [27])	Instrumental [28]
6	Street car (No.10 [27])	BGM (No.10 [27])
7	Woman's speech	Woman's voice
8	Woman with toy block	Woman's voice

ある。

評価法 B は、マルチメディア品質、オーディオ品質及び映像品質のうち、一つの品質要因のみに着目して評価するため、評価者の着眼点が明確になるという長所がある。一方、着目する品質要因を変えて評価実験をする必要があり、評価実験の実施時間が長くなるという欠点がある。

実験条件は以下のとおりである。オーディオ符号化方式は MPEG-2 AAC、映像符号化方式は H.264 とし、オーディオは 2 チャンネルのステレオスピーカ再生、映像は 32 インチ CRT モニタ表示とした。評価環境及び符号化条件を表 1 及び表 2 に示す。評価用映像は ITU-R 勧告 BT.1210 [27] より選んだ 6 映像にテレビ会議を模擬した 2 映像を加えた 8 映像であり、オーディオは映像付属の BGM 若しくは音質評価用音源 (SQAM) [28]、及び会話音声である (表 3)。本評価実験では、オーディオ/映像符号化データを 188 Byte の MPEG-2 TS (Transport Stream) パケットで多重化し、一つの IP パケットに七つの MPEG-2 TS パケッ

表 4 品質パラメータ条件

Table 4 Quality parameter conditions.

Video Quality	Video bitrate	14, 10, 6 Mbit/s
	Audio bitrate	192 kbit/s
	Random packet loss ratio (%)	14 Mbit/s: 0, 0.010, 0.030, 0.070 10 Mbit/s: 0, 0.015, 0.040, 0.100 6 Mbit/s: 0, 0.020, 0.080, 0.160
Audio Quality	Video bitrate	14 Mbit/s
	Audio bitrate	192 kbit/s
	Random packet loss ratio (%)	0, 0.100, 0.200, 0.600

トを格納して転送する映像配信システムを用いた。IP パケットサイズは 1356 Byte である。品質パラメータは、ビットレートとパケット損失率とした (表 4)。先行検討において、オーディオ品質及び映像品質がパケット損から受ける影響は、提示時間 (10 秒) 内に発生する劣化知覚回数に大きく依存することが報告されている [29], [30]。オーディオ及び映像のパケット損失率は、各メディアの品質を良い条件から悪い条件まで広範囲に変化させることを考慮し、10 秒間のパケット損発生回数が 4 段階、すなわち、0 回、1 回、2~5 回、6~8 回となるように、各ビットレートに応じて算出・設定した。ここで、オーディオ符号化ビットレートはいずれの映像符号化ビットレートに対しても 192 kbit/s とし、映像符号化ビットレートが 14 Mbit/s の条件を用いてパケット損失率を設定した。この条件では、映像符号化ビットレートがオーディオ符号化ビットレートに比べて約 70 倍大きいため、約 10 個の IP パケット内 (約 70 個の TS パケット内) に一つのオーディオ TS パケットを含む IP パケットが存在する。このため、オーディオに対するパケット損失率は、映像符号化ビットレートが 14 Mbit/s の場合に設定されたパケット損失率の約 10 倍の設定値となっている。また、オーディオと映像に異なるパケット損失率を与えた理由は、任意の映像品質とオーディオ品質からマルチメディア品質を推定可能なモデルを検討するためである。個別メディア品質が独立に変化するケースは、オーディオパケットと映像パケットが独立の IP パケットストリームとして転送され、ネットワーク内での IP パケット優先制御を適用したり、端末/アプリケーションの誤り訂正/劣化隠ぺい処理等により個別メディア品質を独立に制御したりする場合等が挙げられる。

表 4 の条件で作成されたオーディオ信号及び映像信号をいったん VTR に記録し、様々な品質のオーディオ信号と映像信号を編集して組み合わせることで、個

表 5 5 段階妨害尺度
Table 5 Five-grade impairment scale.

評点	カテゴリー
5	劣化が全く認められない
4	劣化が認められるが気にならない
3	劣化がわずかに気になる
2	劣化が気になる
1	劣化が非常に気になる

別メディア品質が独立に制御されたマルチメディア信号を用意した。評価条件数は、ビットレート、パケット損失率及びコンテンツ種類を組み合わせた全条件のうち、オーディオ品質評価、映像品質評価及びマルチメディア品質評価に対し、各 24 条件とした。評価コンテンツの提示方法は二重刺激妨害尺度法 (DSIS 法: Double Stimulus Impairment Scale Method) [31] を用い、評価尺度はいずれの実験に対しても 5 段階妨害尺度 (表 5) とし、品質を平均オピニオン評点 (DMOS: Degradation Mean Opinion Score) で表した。これは、評価対象が HDTV 解像度の高品質な映像であり、軽微な品質劣化についても評価可能とするためである。評価者は、いずれの実験に対してもオーディオ及び映像に関する業務に従事していない非専門家の 20 代男女合計 24 名とした。なお、評価法 B の実験実施においては、評価順序による品質評価への影響を避けるため、実験日ごとに着目する品質要因 (マルチメディア品質、オーディオ品質、映像品質) の評価順序を変化させた。

3.2 実験結果及び考察

主観評価実験で得られたオーディオ品質 $DMOS_A$ 、映像品質 $DMOS_V$ 、及びマルチメディア品質 $DMOS_{MM}$ について、評価法による品質評価特性の違いを見るため、メディアごとの同じ条件に対する品質評価値を対応づけたものを図 3 に示す。横軸は評価法 A で得られた DMOS、縦軸は評価法 B で得られた DMOS を示しており、括弧内に評価法を記している。

オーディオ品質評価特性 (図 3(a)) については、評価法 A の DMOS が評価法 B の DMOS より高い傾向がある。危険率 5% で有意差検定を行った結果、両者に有意な差が認められた。特に評価値の差が大きい条件は、品質が比較的低い領域であり、オーディオ品質劣化が 2 回から 6 回程度発生している条件である。このときの映像条件は、コンテンツの偏りはなく劣化の発生回数は 0 回から 6 回であり、オーディオ品質の劣

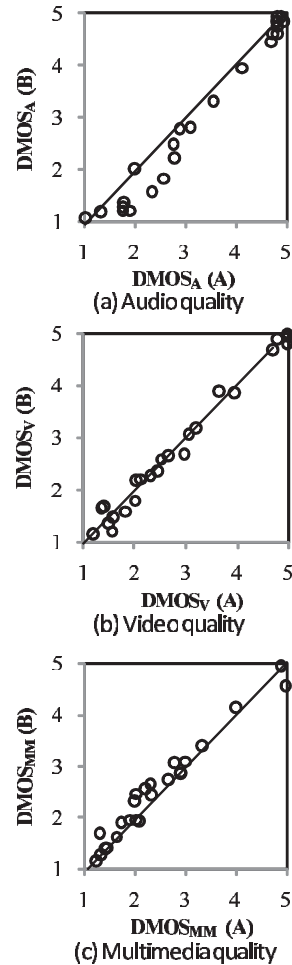


図 3 評価法 A 及び B で得られた DMOS の対応関係
Fig. 3 DMOS relationships between methods A and B.

化発生回数とは無関係であった。逆に、両者の差がない条件としては、オーディオ品質劣化が 0 回若しくは 8 回がほとんどであり、オーディオ品質評価がしやすい条件といえる。この結果、マルチメディア信号を提示した際のオーディオ品質評価は、品質が特に良いあるいは悪い場合を除き映像品質の影響を受けることが分かった。

映像品質評価特性 (図 3(b)) については、同様に危険率 5% で有意差検定を行ったが評価値に有意な差はなかった。また、マルチメディア品質評価特性 (図 3(c)) についても有意差検定を行ったが、映像品質評価特性と同様、評価値に有意な差は見られなかった。この結果、今回の HDTV 映像の品質評価においては、マルチ

メディア信号提示時にオーディオ品質を評価する場合には、映像が提示されるとその影響を受け、オーディオと映像の品質を分離して評価することができないことが分かった。

以上のことから、HDTV 映像を対象としたマルチメディア品質モデルを構築するための個別メディア品質に対する主観評価値を正しく求めるためには、単一メディアを提示する評価環境において品質評価を行う主観品質評価法を用いることが適切であることが分かった。

4. HDTV を対象としたマルチメディア品質推定モデルの構築

3. で明らかになった主観品質評価特性を踏まえ、マルチメディア品質推定モデルの構築に用いるオーディオ及び映像の品質評価値は、それぞれ独立した実験を実施して個別メディア及びマルチメディアの主観品質評価値を得ることとする。

4.1 主観評価実験条件

マルチメディア品質に対するオーディオ品質及び映像品質の影響を明らかにするため、個別メディアの品質を独立に変化させたコンテンツを作成し、評価実験を実施した。

具体的には、まずオーディオ品質条件及び映像品質条件のそれぞれについて、個別メディア品質を独立に制御し、様々な品質の組合せでマルチメディア品質を評価した。これは、3.1 で示したとおり、任意の映像品質とオーディオ品質からマルチメディア品質を推定可能なモデルを検討することが必要と考えたためである。実験条件は表 1 から表 4 と同じである。評価条件数は、モデル構築用として主観評価法比較の場合より多くしている。具体的には、オーディオ及び映像のビットレート、パケット損失率及びコンテンツ種類を組み合わせた条件のうち、オーディオ品質評価及び映像品質評価に対し各 40 条件、マルチメディア品質評価に対し 136 条件である。

4.2 回帰分析を用いたマルチメディア品質推定モデルの作成

主観評価実験の実験条件におけるオーディオ品質と映像品質の分布を図 4 に示す。この図から、それぞれのメディア品質はまんべんなく分布していることが分かる。

図 2 で示したフレームワークに基づき、映像配信サービスに対するマルチメディア品質推定モデルとし

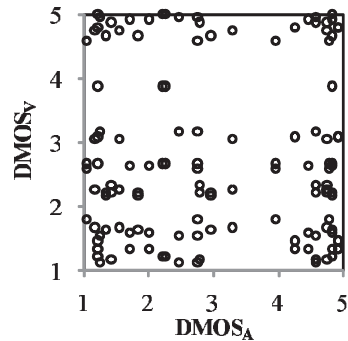


図 4 実験条件に対する $DMOSA$ と $DMOSV$ の分布

Fig. 4 Distribution of $DMOSA$ and $DMOSV$.

表 6 各パラメータ選択時の回帰分析結果

Table 6 Multiple regression analysis results for parameters.

推定パラメータ	相関係数 (r)	回帰式 の有効性	RMSE
$DMOSA$	0.508	NA	0.846
$DMOSV$	0.752	NA	0.647
$DMOSA DMOSV$	0.951	NA	0.302
$DMOSA, DMOSV$	0.915	NG	0.392
$DMOSA, DMOSA DMOSV$	0.963	NG	0.262
$DMOSV, DMOSA DMOSV$	0.963	OK	0.261
$DMOSA, DMOSV, DMOSA DMOSV$	0.964	NG	0.257

て次式 (1) を基準モデルとする。

$$DMOS_{MM} = \alpha DMOSA + \beta DMOSV + \gamma DMOSA DMOSV + \delta \quad (1)$$

ただし、 $1 \leq DMOS_{MM} \leq 5$ 。

ここで、 $DMOSA DMOSV$ はオーディオ品質と映像品質の相互作用を考慮した項であり、 $\alpha \sim \delta$ は定数である。

マルチメディア品質推定モデルにおける各係数を求めるため、表 6 に示す全 7 通りの推定パラメータの組合せに対し、得られた評価値を用いて回帰分析を行い推定精度の比較を行った。これは、対象とするマルチメディアサービスによりオーディオ及び映像のマルチメディア品質に対する寄与度が異なることが考えられるため、それぞれの場合において最適なモデル式を確認することが重要であると考えたためである。そこで、各組合せに対する相関係数、回帰式の有効性及び平均二乗誤差 (RMSE) を求めた結果を表 6 に合わせて示す。ここで、回帰式の有効性とは、回帰式及び偏回帰係数が危険率 5% で有意であること、パラメータ間の多重共線性がないこととした。ここで、“NA” は

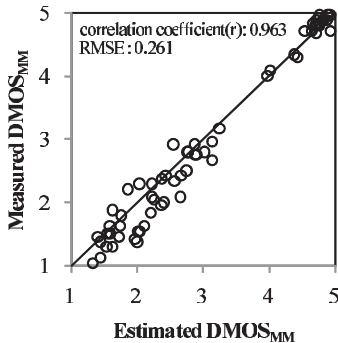


図5 マルチメディア品質推定モデルの品質評価精度

Fig. 5 Accuracy of multimedia quality estimation model.

単回帰分析のため回帰式の有効性は確認しているが多重共線性は発生しない場合，“NG”は回帰式の有効性が確認できなかった場合，“OK”は回帰式の有効性が確認できた場合を示している．この結果，モデルの選択基準を「回帰式の有効性が確認できたもののうちRMSEが最も低いこと」とすると，得られたパラメータは $DMOS_V$ ， $DMOS_A$ $DMOS_V$ ，及び定数項であり，式 (2) で与えられる．

$$\begin{aligned} DMOS_{MM} = & 0.155DMOS_V \\ & + 0.133DMOS_A DMOS_V \\ & + 0.905 \end{aligned} \quad (2)$$

得られたモデル式の品質推定結果を図5に示す．マルチメディア品質の実測値と推定値との相関係数は0.963となり，高精度で品質推定できることが分かった．また，評価コンテンツは絵柄の精細度や動き量，有音率が変化したものを用いたが，コンテンツの違いによらず，式 (2) で統一的に推定できることが分かった．更に，オーディオ品質と映像品質のマルチメディア品質への寄与度を比べると，映像品質の方がオーディオ品質に比べてマルチメディア品質に対する寄与度が大きいことが分かった．

4.3 心理的距離尺度を用いた場合のマルチメディア品質推定モデル

マルチメディア品質モデルの検討として，田坂らは心理的距離尺度を用いてモデルの作成を行っている [19]．これは，カテゴリ尺度は主観品質評価には多く用いられているが，順序尺度であるため QoE 評価尺度としては心理的距離尺度を用いる方が適切であるとしているためである．心理的距離尺度は，カテゴリ尺度

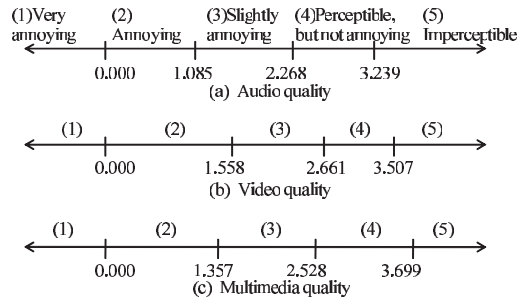


図6 心理的距離尺度によるカテゴリ境界値

Fig. 6 Boundary of each category by psychological scale.

によって得られるデータにカテゴリ判断の法則を適用して算出する．DMOS を構成する全評価者の評点分布が標準正規分布であることを仮定した場合に適用可能である．カテゴリ判断の法則に基づき算出した各カテゴリの境界値を図6に示す．この結果，いずれの品質に対してもカテゴリ尺度は心理的距離では等間隔ではないことが分かる．また，得られた心理的距離尺度と DMOS の対応関係を見ると， $DMOS = 1$ あるいは $DMOS = 5$ 付近では心理的距離尺度との線形性が保たれていないことが分かる (図7)．ここでは，カテゴリ尺度を用いることによって生じる品質推定誤差が，マルチメディア品質推定の誤差として蓄積されていないことを確認する．

まず，カテゴリ尺度で得られた品質評価値 $DMOS_A$ ， $DMOS_V$ 及び $DMOS_{MM}$ をそれぞれ心理的距離尺度による品質評価値 PS_A ， PS_V 及び PS_{MM} に変換する．これらを用いて表6と同様に回帰分析を行った結果を表7に示す．この結果，得られた心理的距離尺度によるマルチメディア品質推定モデルを式 (3) に示す．

$$\begin{aligned} PS_{MM} = & 0.196PS_A + 0.390PS_V \\ & + 0.108PS_A PS_V + 0.085 \end{aligned} \quad (3)$$

心理的距離尺度によるマルチメディア品質推定モデルで得られた品質推定値と実測値の対応関係を図8に示す．この結果，両者の相関係数は0.961であり，カテゴリ尺度を用いた場合の相関係数0.963と同程度であることが確認できた．

以上より，カテゴリ尺度を用いたマルチメディア品質推定モデルも心理的距離尺度を用いたモデルと同様の推定性能をもつことが確認でき，特に評価尺度の線形性が保たれないことによる誤差は見られないこと

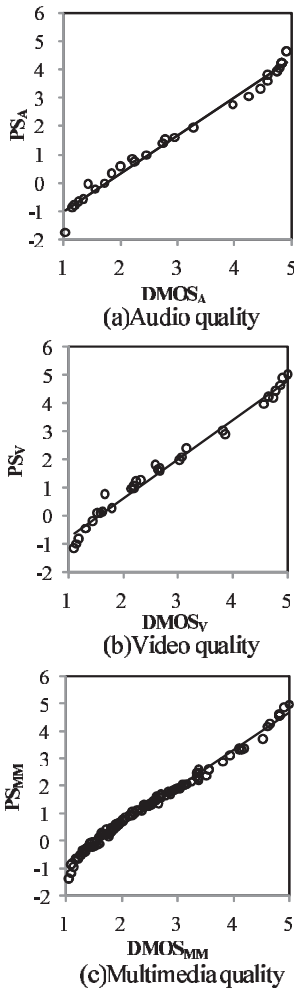


図 7 心理的距離尺度と妨害尺度の対応関係
Fig. 7 Psychological scale versus DMOS.

表 7 心理的距離尺度を用いた場合の回帰分析結果
Table 7 Multiple regression analysis results using psychological scale.

推定パラメータ	相関係数 (r)	回帰式 の有効性	RMSE
PS_A	0.506	NA	1.176
PS_V	0.750	NA	0.902
$PS_A PS_V$	0.882	NA	0.643
PS_A, PS_V	0.922	OK	0.529
$PS_A, PS_A PS_V$	0.882	NG	0.643
$PS_V, PS_A PS_V$	0.939	OK	0.469
$PS_A, PS_V, PS_A PS_V$	0.961	OK	0.377

が分かった．以降は，カテゴリー尺度によるモデル式を用いてマルチメディア品質推定モデルの検証を行うこととする．

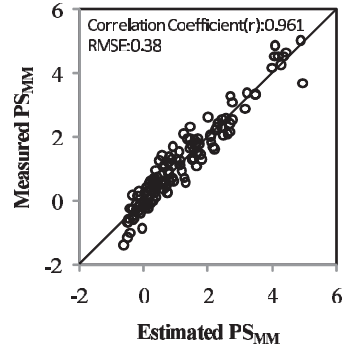


図 8 心理的距離尺度を用いたマルチメディア品質推定モデルの品質評価精度

Fig. 8 Accuracy of multimedia quality estimation for psychological scale data.

5. マルチメディア品質推定モデルの検証

カテゴリー尺度で得られたマルチメディア品質推定モデルの品質推定精度の検証を行う．検証では，実映像配信システムや未知コンテンツに対する適用を考慮した．

モデル構築時は個別メディア品質のマルチメディア品質に対する影響を把握するため，映像信号とオーディオ信号を分離し，それぞれの品質の変動範囲を考慮した条件で主観評価試験を行った．しかしながら，実映像配信システムでは，オーディオ及び映像符号化情報を MPEG-2 TS パケット化し，両者を同一 IP パケット内に混在して伝送する方法がある．このような場合，パケット損が発生すると，両メディアの品質劣化が独立して生じるとは限らない．つまり，オーディオ情報が含まれる TS パケットと映像情報が含まれる TS パケットが IP パケット内に混在する場合，その IP パケットが損失すると劣化の再生時間差はあるが，時間軸上の近い範囲でオーディオと映像の品質劣化が生じる．また，特に HDTV を対象とするような高ビットレートの映像配信サービスでは，オーディオビットレートが映像ビットレートに比べ，極端に小さいことから，IP パケット損が生じる場合には，映像品質劣化の方が音声品質劣化に比べ多く発生する可能性が高い．そこで，まず，実映像配信システムを用いた品質劣化パターンで，本モデルが適用可能かどうかを検証する．

次に，コンテンツ条件について検証を行う．モデル構築時に使用したコンテンツは，8 種類のうち 6 種類を標準動画像とした．標準動画像は映像品質評価では一般的に使われているが，マルチメディア品質を評価

するという観点においては、オーディオと映像が必ずしも意味的に関連づけられたものにはなっておらず、十分ではない可能性がある。そこで、コンテンツ条件の検証では、映像配信サービスを想定し独自に製作した各種オーディオ付き映像を用意し、本モデルの適用可能性を検証した。

二つの検証実験の実施に際しては、表 1 に示す評価環境のうちオーディオの受聴手段をスピーカからヘッドホン (SENNHEISER HD25-1) に変更した。これは、パケット損によるオーディオ品質劣化と映像品質劣化がほぼ同時に発生し、オーディオ品質劣化の検出が難しくなる可能性があるため、マルチメディア品質評価時の評価感度の低下を抑えるためである。

5.1 実映像配信システムに対するモデルの品質推定精度検証

実映像配信システムで発生するオーディオ及び映像の劣化パターンに対するモデルの品質推定精度検証を行う。符号化装置条件及びコンテンツ条件は、表 2 及び表 3 と同じである。パケット損失率については、モデル構築時と同じ実験の枠組みでモデルの検証を行うため、表 4 と品質劣化の変動範囲は同一にし、変動範

囲内でのパケット損失率の設定値を変化させた。検証用品質パラメータ条件を表 8 に示す。評価条件数は、ビットレート、パケット損失率及びコンテンツ種類を組み合わせた全条件に対し、オーディオ品質評価、映像品質評価及びマルチメディア品質評価に対し、各 72 条件である。

実験により得られた $DMOS_A$ と $DMOS_V$ の分布を図 9 に示す。IP パケット内で構成される映像パケット数はオーディオパケット数より多く、映像品質劣化

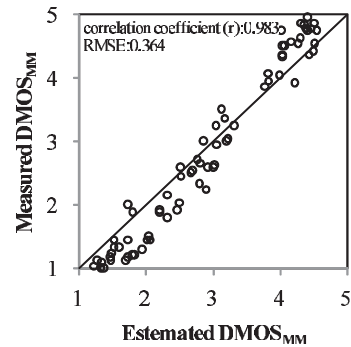


図 10 実映像配信システムに対する品質推定精度
Fig. 10 Accuracy of multimedia quality estimation for system dependency.

表 8 検証用品質パラメータ条件

Table 8 Quality parameter conditions for model qualification.

Video	Codec	H.264
	Resolution	1440 × 1080
	Bitrate	14, 10, 6 Mbit/s
Audio	Frame rate	30 fps
	Codec	MPEG-4 AAC-LC
Random packet loss ratio (%)	Bitrate	192 kbit/s
	14 Mbit/s:	0, 0.015, 0.035, 0.070
	10 Mbit/s:	0, 0.020, 0.050, 0.100
	6 Mbit/s:	0, 0.030, 0.070, 0.160

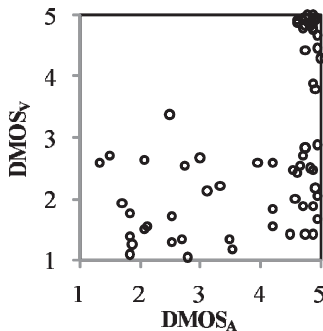


図 9 検証条件に対する $DMOS_A$ と $DMOS_V$ の分布
Fig. 9 Distribution of $DMOS_A$ and $DMOS_V$ for qualification condition.

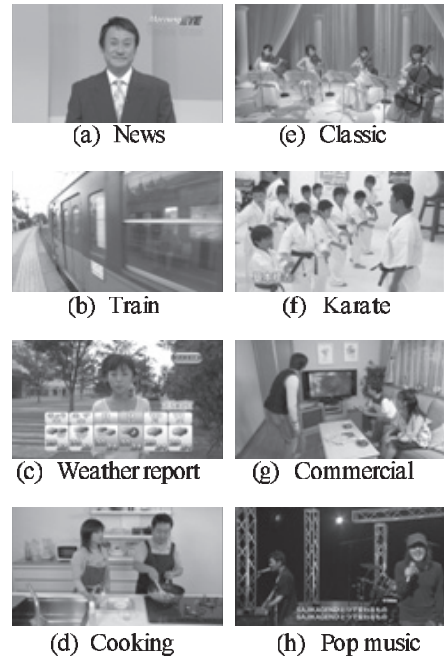


図 11 検証用コンテンツ

Fig. 11 Contents for verification.

表 9 検証用コンテンツの内容

Table 9 Contents detail.

コンテンツ名	映像	オーディオ
(a) ニュース	男性キャスター	男性の声, BGM
(b) 電車	電車, 電車内	男声ナレーション, BGM
(c) 天気予報	屋外で話す女性, 予報 CG	女性の声, BGM
(d) 料理	調理する男性と女性	男性と女性の声, 包丁で刻む音, BGM
(e) 弦楽四重奏	楽器を演奏する 4 人の女性	バイオリン, チェロ, コントラバス
(f) 空手	練習する子供たち, 指導者	男声ナレーション, 子供たちの掛け声, BGM
(g) CM	部屋の中で話す男性と女性	男性と女性の声, BGM
(h) ポップス	バンド演奏	男性の歌声, ギター, ベース, ドラム

時のみオーディオ品質も劣化している可能性が高い。このため、映像品質の高い領域でオーディオ品質が低い条件はなく、映像品質が低下する領域においてのみオーディオ品質が変動している。得られた主観評価値を式 (2) に代入してマルチメディア品質を推定した結果を図 10 に示す。マルチメディア品質の実測値と推定値との相関係数は 0.983 となり、高い相関が得られた。この結果、個別メディア品質を独立に制御して構築したマルチメディア品質推定モデルをオーディオ品質及び映像品質の $DMOS_A$ と $DMOS_V$ の分布に偏りがあるシステムの品質推定に適用しても問題ないことが分かった。

5.2 未知コンテンツに対する検証

映像配信サービスを想定した映像に対する品質推定精度の検証を行う。検証用映像として映像配信サービスを模擬した、ニュース番組、鉄道、天気予報、料理番組、弦楽四重奏、空手練習、コマーシャル及びポップスの 8 種類を選定した。各コンテンツのスナップショットを図 11 に、各コンテンツのオーディオ及び映像の内容を表 9 に示す。これらの映像とオーディオは意味的に関連づけられているものである。符号化装置条件及び品質パラメータは、表 2 及び表 8 と同じである。評価条件数は、ビットレート、パケット損失率及びコンテンツ種類を組み合わせた全条件に対し、オーディオ品質評価、映像品質評価及びマルチメディア品質評価に対し、各 72 条件である。

得られた主観評価値を式 (2) に代入してマルチメディア品質を推定した結果を図 12 に示す。マルチメディア品質の実測値と推定値との相関係数は 0.985 となり、高い相関が得られた。よって、実サービスコンテンツを考慮した未知コンテンツの場合に対しても十分な精度で推定可能であり、コンテンツが異なっても統一されたマルチメディア品質推定モデルが適用できることが分かった。

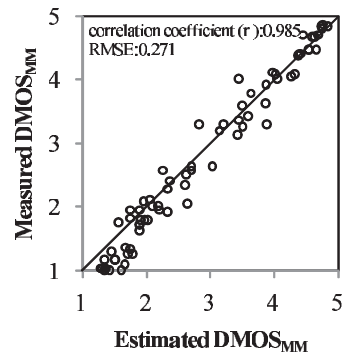


図 12 検証用コンテンツに対する品質推定精度

Fig.12 Accuracy of multimedia quality estimation for unknown contents.

6. む す び

HDTV 映像配信サービスを対象にマルチメディア品質推定モデルの構築を目的として、マルチメディア品質推定モデルについて検討を行った。

まず、マルチメディア主観品質評価法について、マルチメディア信号を視聴し、マルチメディア品質、オーディオ品質及び映像品質を同時に評価する場合と、それぞれの信号のみを提示し独立して品質を評価する場合について評価結果を比較した。この結果、マルチメディア信号提示におけるオーディオ品質評価では映像提示の影響を受けることが分かった。このことから、オーディオ品質と映像品質のそれぞれを入力とするマルチメディア品質推定モデルの場合には、オーディオあるいは映像信号のみを提示し、それぞれの品質を独立して評価することが適切であると分かった。

次に、マルチメディア品質推定モデルを構築するため、オーディオ品質及び映像品質を独立に変動させた場合の主観評価実験を上記方針に基づき実施し、回帰分析によりマルチメディア品質推定モデルを決定した。この結果、HDTV 解像度の映像配信サービスを対象

とするマルチメディア品質推定モデルは、映像品質の寄与度がオーディオ品質の寄与度よりも高いモデルが得られた。また、カテゴリー尺度により構築した本モデルと心理的距離尺度を用いたモデルとの品質評価精度を比較し、同程度の推定精度が得られていることから、カテゴリー尺度によるマルチメディア品質推定モデルは有効であることを確認した。

更にマルチメディア品質推定モデルの品質推定精度の検証として、実映像配信システムで発生する品質劣化パターンに対し適用可能であるか確認した結果、マルチメディア品質の実測値と推定値との相関係数は0.983となり、高い相関を得た。また、モデル構築時とは異なる未知コンテンツに対して検証を行った結果、マルチメディア品質の実測値と推定値との相関係数は0.985と高い相関をもち、精度良く推定できることを確認した。

今後は、今回は検討対象としなかったオーディオと映像の同期ずれについての検討や解像度が異なる映像配信サービスを対象とした場合のマルチメディア品質推定モデルへの適用可能性を検討する必要がある。また、今回の検討ではマルチメディア品質推定モデルを導く際の入力として個別メディアに対する主観品質評価値を用いたが、個別メディアに対する客観品質評価値を入力としてマルチメディア品質を推定した場合の有効性検討についても今後の課題である。

謝辞 本研究の一部は、総務省戦略的情報通信研究開発推進制度 (SCOPE) 「マルチメディア品質客観評価技術の国際標準化 (073103002)」の成果である。

文 献

- [1] ITU-T Rec. P.10, "Vocabulary for performance and quality of service," July 2006.
- [2] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities in ITU on QoE assessment of IPTV," IEEE Commun. Mag., pp.78–84, Feb. 2008.
- [3] 林 孝典, 高橋 玲, 吉野秀明, "マルチメディア通信サービスの QoE 評価技術に関する動向と課題," 信学論 (A), vol.J91-A, no.6, pp.600–612, June 2008.
- [4] ITU-T Rec. G.107, "The E-model: A computational model for use in transmission planning," April 2009.
- [5] A. Takahashi, A. Kurashima, and H. Yoshino, "Objective assessment methodology for estimating conversational quality in VoIP," IEEE Trans. ASLP, vol.14, no.6, pp.1983–1993, Nov. 2006.
- [6] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Feb. 2001.
- [7] ITU-T Rec. P.862.2, "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs," Nov. 2007.
- [8] ITU-R Rec. BS.1387, "Method for objective measurements of perceived audio quality," Nov. 2001.
- [9] ITU-T Rec. J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," March 2004.
- [10] ITU-T Rec. J.246, "Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference," Aug. 2008.
- [11] ITU-T Rec. J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," Aug. 2008.
- [12] 岡本 淳, 林 孝典, 高橋 玲, 栗田孝昭, "映像の時空間的特徴量を考慮した映像品質客観評価法の提案," 信学論 (B), vol.J88-B, no.4, pp.813–823, April 2005.
- [13] H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcast., vol.50, no.3, pp.312–322, Sept. 2004.
- [14] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process., vol.15, no.11, pp.3441–3452, Nov. 2006.
- [15] 渡辺敬志郎, 岡本 淳, 栗田孝昭, "映像コミュニケーションサービスで発生するフリーズ劣化を対象とした映像品質客観評価法," 信学論 (B), vol.J90-B, no.10, pp.1036–1044, Oct. 2007.
- [16] P.L. Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," IEEE Trans. Neural Netw., vol.17, no.5, pp.1316–1327, Sept. 2006.
- [17] ITU-T Rec. J.148, "Requirements for an objective perceptual multimedia quality model," May 2003.
- [18] D.S. Hands, "A basic multimedia quality model," IEEE Trans. Multimed., vol.6, no.6, pp.806–816, Dec. 2004.
- [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] A. Rimell and A. Owen, "The effect of focused attention on audio-visual quality perception with applications in multi-modal codec design," IEEE ICASSP 2000, vol.4, pp.2377–2380, June 2000.
- [21] S. Winkler and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," IEEE Trans. Multimed., vol.8, no.5, pp.973–980, Oct. 2006.
- [22] M.R. Frater, J.F. Arnold, and A. Vahedian, "Impact of audio on subjective assessment of video quality in videoconferencing applications," IEEE Trans. Circuits Syst. Video Technol., vol.11, no.9, pp.1059–1062, Sept. 2001.
- [23] 門田俊紀, 北脇信彦, 山田武志, "映像品質と音声品質と

の相互作用を考慮したマルチメディア品質の推定モデル”
信学技報, CQ2008-33, Sept. 2008.

- [24] M.N. Garcia and A. Raake, “Impairment-factor-based audio-visual quality model for IPTV,” IEEE QoMEX 2009, pp.1–6, July 2009.
- [25] ITU-T Rec. G.1070, “Opinion model for video-telephony applications,” March 2007.
- [26] ITU-T Rec. P.920, “Interactive test methods for audio-visual communications,” May 2000.
- [27] ITU-R Rec. BT.1210, “Test materials to be used in subjective assessment,” Oct. 1995.
- [28] European Broadcasting Union, “SQAM (sound quality assessment material) recording for subjective tests,” April 1988.
- [29] K. Yamagishi and T. Hayashi, “Non-intrusive packet-layer model for monitoring video quality of IPTV services,” IEICE Trans. Fundamentals, vol.E92-A, no.12, pp.3297–3306, Dec. 2009.
- [30] N. Egi, T. Hayashi, and A. Takahashi, “Parametric packet-layer model for evaluation audio quality in multimedia streaming services,” IEICE Trans. Commun., vol.E93-B, no.6, pp.1359–1366, June 2010.
- [31] ITU-T Rec. P.911, “Subjective audiovisual quality assessment methods for multimedia applications,” Dec. 1998.

(平成 22 年 3 月 26 日受付, 8 月 23 日再受付)



富永 聡子 (正員)

昭 62 電通大・経営工学卒。平元同大大学院修士課程了。同年日本電信電話(株)入社。以来、ファクシミリ、映像、マルチメディア通信サービスの品質評価の研究に従事。現在、NTT サービスインテグレーション基盤研究所研究主任。平 20 電気通

信普及財団賞を受賞。



山岸 和久 (正員)

平 13 東京理科大・電子工学卒。平 15 早稲田大学大学院電子情報通信了。同年日本電信電話(株)入社。以来、テレビ電話及び IPTV サービスの品質評価の研究に従事。現在、NTT サービスインテグレーション基盤研究所社員。平 19 本会学術奨励賞、

平 20 電気通信普及財団賞を各受賞。



林 孝典 (正員)

昭 63 筑波大・第 3 学群・基礎工卒。平 2 同大大学院修士課程了。同年日本電信電話(株)入社。以来、映像符号化品質、マルチメディアネットワーク品質、マルチメディアサービス品質評価の研究に従事。平 9~12 通信・放送機構にて、分散型ネットワークの高性能化技術の研究に従事。現在、NTT サービスインテグレーション基盤研究所主幹研究員。平 16 本会活動功労賞、平 20 電気通信普及財団賞を各受賞。博士(工学)。



高橋 玲 (正員)

昭 63 北大・理・数学卒。平 5 米国カリフォルニア工科大学電気工学科修士課程了。昭 63 日本電信電話(株)入社。以来、音声・オーディオ・映像品質評価技術の研究に従事。現在、サービス評価グループのグループリーダーとして研究マネジメントに従事。博士(工学)。平 4 より ITU-T SG12, 平 14 より TTC 網管理専門委員会において、音声品質の主観・客観評価法の標準化活動に貢献。平 17 より ITU-T SG12 課題 13 ラポータ。H21 より ITU-T SG12 副議長。平 16 年度情報通信技術委員会(TTC)表彰功労賞「IP 電話通話品質評価の標準化にかかわる功績」受賞。平 17 年度日本 ITU 協会賞国際活動奨励賞受賞。平 18 年度本会通信ソサイエティ論文賞受賞。平 19 年度及び平成 20 年度電気通信普及財団賞(テレコムシステム技術賞)受賞。