



映像品質評価法

の国際標準化動向

Current Status and
Trend of International Standardization
of Video Quality Assessment

堀田裕弘

富山大学大学院理工学研究部

映像品質評価法の分類

地上デジタル放送やDVD、携帯端末などの普及とともにデジタル映像は我々の身近なものとなってきた。地上デジタル放送では映像素材の多くが従来のSDTVからHDTVへと移行しつつあり、この映像品質の大幅な向上は、白黒テレビからカラーテレビへ移行した時代を彷彿とさせる大きなインパクトを視聴者に与えている。一方、電波を利用した放送サービスだけでなく、IPネットワークを利用した映像配信サービスも普及が目覚ましい。光ファイバや次世代ネットワーク（NGN）などの広帯域ネットワークの実現により、HDTVクラスの映像もネットワーク経由で安定に配信できる時代がすぐそこまで来ている。

通常のHDTVの情報量は非圧縮の場合、約1ギガビット/秒であるため、情報量の削減の観点からMPEG-2やH.264などの映像符号化技術が必要となる。しかしながら、この映像符号化が、映像の品質を低下させてしまう原因となる。また、映像の品質をリアルタイムに評価することも要求されるようになり、人間を介さずに自動的に評価する技術への要求が高まってきている。従来は、テレビジョン放送を前提とした符号化方式の性能評価を行うことが典型的な目的であったが、近年ではインターネットや3Gモバイルによる高速データ通信網などの普及により、パケット交換ネットワーク上で一定のエラー率が発生するという前提のもとでさまざまな解像度、フレームレート映像の受信品質を測定するなど、要求される品質やフォーマットの幅が広がっており、それぞれのアプリケーションにとって最適な映像品質の評価方式を提供することが求められている。

一般的に、映像品質をコンピュータ上で取り扱うためには、その映像品質を数値化して表現することが必要である。これは人間が映像コンテンツを観視することで受ける印象を数値化することに相当する。映像品質を数値化する手法には、大きく分けて「主観的評価」と「客観

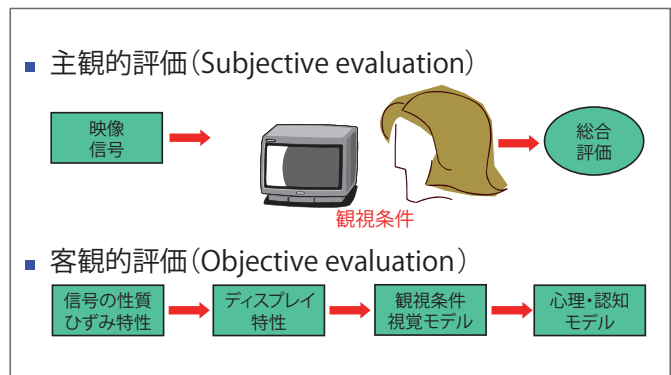


図-1 評価方法の分類

的評価」の2種類が存在する(図-1)。主観的評価(以後、主観評価と略す)法は、表示デバイスに映し出された映像を人間が観視し、その品質を主観的に評価する方法である。この方法は、映像をストレスなく提示するための表示デバイスや再生装置、20人程度の評価者を集めることなど、多大な時間と労力を必要とする。観視条件に関しては、ITU-R(国際電気通信連合無線通信部門)の勧告BT.500として評価の実施方法などが決められている¹⁾。

この負荷を軽減するための手段として、人間のかわりにコンピュータがこの評価プロセスを代行する方法が、客観的評価(以後、客観評価と略す)法である。これは、評価を行いたいデジタル化された映像素材などをコンピュータに入力し、さまざまな映像信号処理を経て、主観評価により得られた評点を模倣する数値を出力するものである。したがって、客観評価を行うコンピュータ・ソフト(あるいはハード的な装置)には、人間が映像を見てその品質を評価する処理プロセスが何らかの信号処理手段として考慮されていることになる。



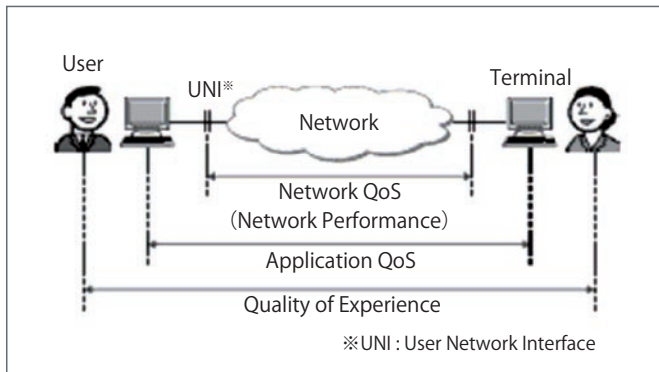
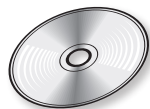


図-2 QoS と QoE の定義

QoS と QoE

ここでは、評価方法の考え方として、ネットワーク主導であるか、あるいは、ユーザ主導であるかの違いを表す QoS と QoE について述べる²⁾。従来、通信サービスの品質を表す用語として QoS (Quality of Service) が広く用いられている。ITU-T では、ネットワーク性能やアプリケーション性能としての QoS 目標値などが議論されている。QoS がカバーする領域は多岐にわたり、OSI (Open Systems Interconnection) 参照モデルのさまざまなレイヤで提供される“サービスの性能”を表す用語として用いられている。通信サービスを楽しむユーザが体感する品質も QoS の1つの側面として捉えられてきたが、NGN に関する議論が進められる中で、QoS をネットワーク性能 (Network Performance) に対応付けて用いるケースが多くなってきたため、ITU-T は 2007 年 1 月、「ユーザ体感品質」を表現する用語として、新たに QoE (Quality of Experience) を定義した。QoS は通信事業者やサービス提供者から見たサービス品質の尺度であるのに対し、QoE はユーザから見たサービス品質の尺度と言える。今後は、人間の知覚・認知特性を考慮した品質を QoE と呼び、ネットワーク性能やアプリケーション性能としての QoS などと区別されることになる (図-2)。



レイヤによる客観的な品質評価法分類

次に、レイヤを基本とした客観的な品質評価について考える。客観的な品質評価法は、さまざまなメディアに対する主観品質を、それぞれの物理量を用いて推定する方法である。利用シーンや入力情報の違いにより、表-1 に示す 5 種類の客観的な品質評価モデルに大別される³⁾。

●メディアレイヤモデル

メディアレイヤモデルは、音声・映像メディア信号を用いて品質劣化を定量化することで主観品質を推定する。符号化器のパラメータ・チューニングやサービス品質の実力把握などに利用できる。メディアレイヤモデルは、原信号と劣化信号を直接比較することで主観品質を推定する。

●パラメトリックパケットレイヤモデル

パラメトリックパケットレイヤモデルは、RTP や RTCP などのパケットのヘッダ情報のみを用いて主観品質を推定する。メディアレイヤモデルのようにメディア信号を復号しないことから処理負荷が非常に軽く、サービス提供中に実現品質を把握するインサービス品質管理への適用が期待されている。メディア情報を用いず品質を推定するため、特に音声・映像メディア品質のコンテンツ依存性を考慮した評価は本質的に困難である。このため、コンテンツ属性を仮定したり、符号化方式などの各種システム情報を事前に得ておく必要がある。

●ビットストリームレイヤモデル

ビットストリームレイヤモデルは、パケットのヘッダ情報に加え、ペイロード情報 (復号前の符号化ビット系列情報など) を用いて主観品質を推定する。ペイロード情報までを利用することにより、パラメトリックパケットレイヤモデルでは扱えなかったコンテンツ依存性を考慮した評価を実現するモデルとして近年注目を集めている。メディアレイヤモデルを利用するためには音声や映像がユーザに提示されるインタフェース部でメディア信号を取得する必要があるのに対して、ビットストリー

	メディアレイヤモデル	パラメトリックパケットレイヤモデル	ビットストリームレイヤモデル	パラメトリックプランニングモデル	ハイブリッドモデル
入力情報	メディア信号	パケットのヘッダ情報	ペイロード情報	品質設計/管理パラメータ	左記情報の組合せ
主な評価用途	・パラメータ最適化 ・サービス実力値把握	・インサービス品質管理	・インサービス品質管理	・品質設計	・インサービス品質管理

表-1 レイヤによる評価法の分類

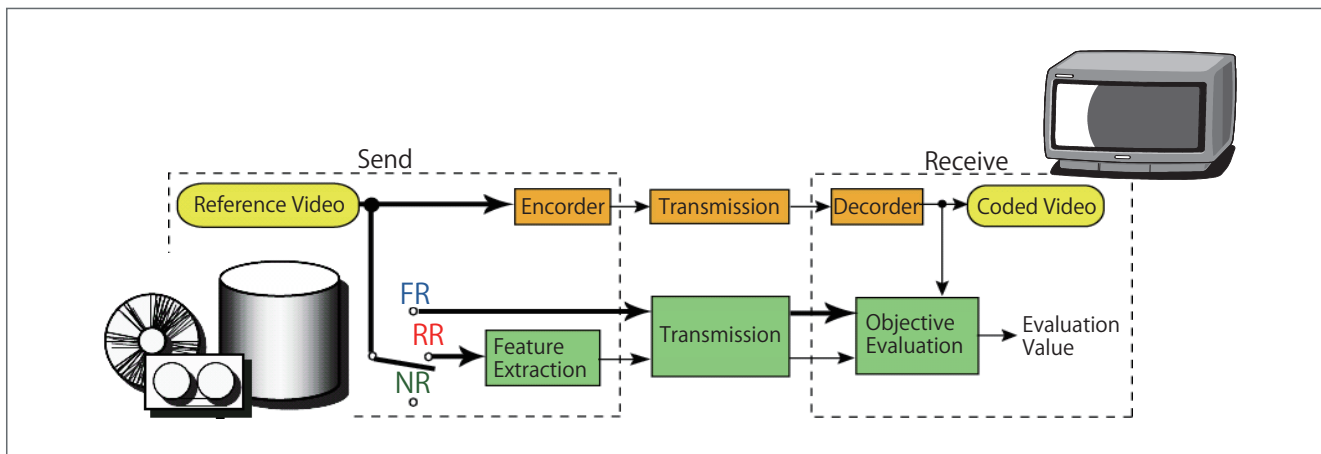


図-3 客観的な品質評価のフレームワーク

ムレイヤモデルではパケット情報を収集できれば推定可能となるため、装置構成が単純化できるメリットもある。ITU-Tでは、IPTVを含めた映像配信サービスを対象として本技術の標準化を行う予定であり、この勧告は暫定的にP.NBAMS（Non-intrusive Bit-stream-layer Assessment of Multimedia Streaming）と呼ばれている。

● **パラメトリックプランニングモデル**

前述までの客観的な品質評価モデルが、メディア信号やパケット情報などを入力として主観品質を推定するモデルであったのに対し、パラメトリックプランニングモデルは、ネットワークや端末の品質設計・管理パラメータ（たとえば、符号化ビットレートやパケット損失率など）を入力として主観品質を推定する。評価対象となる符号化方式やシステムごとに主観的な品質評価特性をあらかじめ求め、データベース化しておく必要がある。本モデルはサービス設計段階において机上で効率的に品質設計ができる点が特徴である。

● **ハイブリッドモデル**

ハイブリッドモデルは、これまでの4種類のモデルを組み合わせて構成することで、簡易かつ精度良く主観品質を推定するアプローチである。インサービス品質管理などにおいて取得可能な情報を最大限利用するためのモデルである。たとえば、メディアレイヤモデルのみでは評価対象の品質劣化が符号化によるものかネットワークで発生するパケット損失によるものかを判断することが困難な場合があるが、ビットストリーム情報やパケット情報を補足的に利用することにより、品質推定精度を改善することが期待される。逆に、パラメトリックパケットレイヤモデルに対して受信信号の特徴量を補足情報として使い、コンテンツ属性を考慮した管理を実現するこ

とも考えられる。

参照情報の有無によるメディアレイヤモデルの分類

前 述のメディアレイヤモデルでは、音声・映像メディア信号を用いて品質劣化を定量化することで主観品質を推定するが、参照情報（映像では無圧縮の映像情報）をどのように利用するかで、その評価のフレームワークが異なる。客観的な評価モデルとして、ITU-T勧告J.143により規定された以下の3つのフレームワーク（図-3）について述べる⁴⁾。

- Full Reference (FR) フレームワーク
- No Reference (NR) フレームワーク
- Reduced Reference (RR) フレームワーク

● **Full Reference (FR) フレームワーク**

FR型は、送信側のリファレンス映像SRC（Source Reference Circuit,あるいはSource Video Sequence）、受信側の劣化映像PVS（Processed Video Sequence）双方のベースバンド信号をモデルに入力するフレームワークである。映像の送信側、受信側の両方の映像が必要であるため、伝送系でのリアルタイムの監視には向かず、エンコーダの符号化性能評価や伝送エラー時のデコーダでのエラーコンシールメント性能など、伝送路に相当するHRC（Hypothetical Reference Circuit）に包含されるシステム・伝送系の性能評価に用いられる。SRC、PVSを画素単位の精度で解析することができるため、すべてのフレームワークの中で最も高い精度で主観品質を推定することができる。

● **No Reference (NR) フレームワーク**

NR型は、PVSのベースバンド信号の解析のみから主観品質を推定するフレームワークである。受信側の映

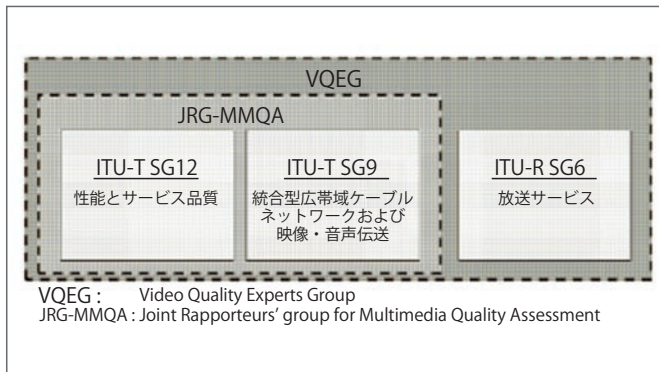


図-4 映像の品質評価法に関する国際標準化機関

像のみで品質評価ができるため、伝送系のリアルタイム監視を簡易な構成で実現できるというメリットがある。一方、SRC の情報をリファレンスとして利用することができないため、他のフレームワークに比べて精度の面で劣る点の問題となっている。一般に、FR 型では SRC と PVS の差分信号に対して、視覚特性や心理特性などさまざまな知覚モデルを適用することにより主観品質推定のパラメータとなる映像特徴量を抽出しているが、NR 型ではこの差分情報を利用することができないため、PVS のみからこれに相当する情報を抽出しなければならない。この点が NR 型の精度改善に対する最も大きな課題である。こうした課題を克服するためのアプローチとして、復号映像の解析により、当該映像のビットストリームに記述された DCT 係数の復元を試みる手法や、符号化により発生するブロックひずみの境界を検出する手法などが提案されている。

● Reduced Reference (RR) フレームワーク

RR 型は、SRC 側から所定の情報量に相当する映像特徴量を抽出（図-3の“Feature Extraction”のブロックに相当）した後、これを映像回線とは別に用意されたデータ回線を用いて受信側に伝送し、受信側でこれをリファレンス情報として用い主観品質を推定する方式である。なお、受信側は、他のフレームワークと同様に PVS ベースバンド信号の情報を得ることができる。前述のとおり、NR 型では SRC の情報がまったく参照できないため精度に問題があったが、RR 型では情報量の制限があるとはいえある程度のリファレンス情報を得ることができる。NR 型と同じく伝送系のリアルタイム監視が可能で、リファレンス情報の取得による精度の向上が期待できる意味で、RR 型は、FR 型と NR 型の中間のフレームワークと呼ぶことができる。

一方、RR 型の実用上の課題として以下の点が挙げられる。まず、RR 方式では、データ回線を介してリファレンス情報を取得するため、リファレンス情報取得まで

の遅延が生じてしまう。これは、品質評価尺度の出力や障害の検出に遅延が生じる原因となる。また、通常のマルチメディアアプリケーションでは、符号化やデータ伝送による遅延を測定する手段がないため、送信側から送られてくる各フレームの特徴量が PVS のどのフレームに相当するのかを照合するフレーム同期機構が必要となる。しかも、伝送エラーなどによる映像遅延の変動に常時追従する必要があるため、安定した運用のためには高度な実装が必要となる。

VQEG (Video Quality Experts Group)

映像の主観・客観的評価技術の国際標準化は、ITU-T (International Telecommunication Union -Telecommunication standardization sector) および ITU-R (同-Radio communication sector) において行われている。特に通信と放送の連携を視野に両分野の専門家が共同で品質評価技術の標準化に取り組む必要があることから、ITU-T/R を横断し品質評価の専門家グループ VQEG (Video Quality Experts Group)^{5), 6)} が設立され、各組織の緊密な連携の下、国際標準化が進められている（図-4）。

VQEG は、ITU (International Telecommunication Union : 国際電気通信連合) での標準化を前提に、複数の機関が提案する客観的な品質評価方式を公平な条件のもとで評価し、各提案方式での主観品質の推定精度を議論することを目的としたテスト実施機関である。VQEG での方式評価テストにあたっては、テストプランという実施手順書が作成され、前提とする客観的な評価フレームワーク、符号化パラメータや伝送エラー率などの伝送条件、主観評価実験の実施方法、各提案方式の性能評価のためのデータ解析方法などが詳細に定義されている。

● VQEG におけるこれまでの活動概要

VQEG のこれまでの活動概要について述べる。VQEG の活動目的は公正な方式評価テストの実施とその結果の出力であり、ITU の関連 Study Group に対して標準方式を勧告するなどの作業は行わない。これは、VQEG の所掌はあくまで方式評価テストの実施とその結果の出力までであり、テストの結果をどのように評価し、複数の提案方式のうちどれを勧告化するのか（あるいは、しないのか）は各標準化団体の判断にゆだねるという方針に基づくものである。

VQEG では、主に、メディアレイヤモデルに関して、標準テレビ方式を前提とした Full Reference フレームワーク型の方式評価テストを 2 回 (FRTV, FRTV-II)

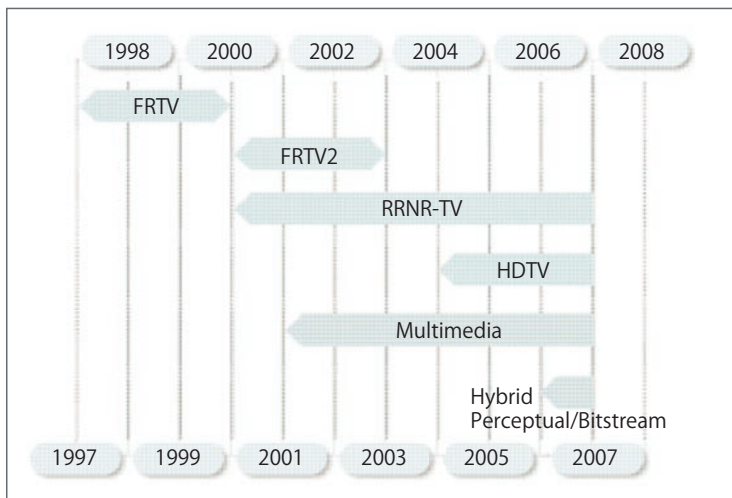


図-5 VQEGプロジェクトのロードマップ

実施しており、その結果としてITU-T勧告J.144およびITU-R勧告BT.1683として標準化されている。これらの勧告は符号化歪を評価対象としており、IPネットワークを介した映像配信サービスなどに適用するためにはパケット損失歪による品質劣化をいかに定量化するかが課題となる。その後、テレビジョン伝送の2次分配でのReduced ReferenceフレームワークおよびNo Referenceフレームワークによる評価を前提としたRRNR-TVテストおよび、マルチメディアアプリケーションにおける映像フォーマットを前提としたMultimediaテストが計画され、さらに2004年にはHDTVテスト、2006年にはHybrid Perceptual/Bitstreamテストが実施予定に追加された。現在、Multimediaテストが実施中であり、2008年前半にはテスト完了の予定となっている。また、Multimediaテスト完了後には、RRNR-TV、HDTV、Hybrid Perceptual/Bitstreamの各テストが順次実施される予定である。VQEGの各プロジェクトのロードマップを図-5に示す。

Multimedia テストプラン

VQEGでは、PCやPDAなどを用いたマルチメディアサービスにおける映像品質を客観評価する技術の検討を現在進めており、2008年中には勧告化予定である。また、音声・映像の個別メディア品質の相互作用とメディア間同期（いわゆるリップシンク）を考慮し、マルチメディア品質を推定する客観品質評価モデルに対する要求条件が、ITU-T勧告J.148にまとめられており、上述の個別メディアの客観的な品質評価技術の進展に合わせて詳細が検討されることになっている。Multimediaテストは、VQEGにとってはテレビ

ジョン伝送以外のアプリケーションを想定した初めてのテストであり、当初はRRNR-TVテストの後に実施される予定であった。しかし、FRTV-IIテストの完了が予想以上に伸びた上、マルチメディアアプリケーションにおける客観的な画質評価技術への強い要望からRRNR-TVと順番を入れ替えて実施することとなった。

● Multimedia テストにおける画質劣化モデル

Multimedia テストプランで前提とする劣化モデルを図-6に示す。同図は、映像伝送を簡略化したモデルで、送信側のリファレンス映像をSRC (Source Video Sequence), 受信側の劣化映像をPVS (Processed Video Sequence) と呼ぶ。また、伝送路に相当する部分はHRC (Hypothetical Reference Circuit) と呼ばれ、この部分に、エンコード、デコードおよび伝送中のエラー、ジッタ等のすべての劣化要因がモデル化されている。よって、HRCには単純な符号化のみの劣化モデルもあれば、伝送エラーモデルが含まれる場合もある。マルチメディアアプリケーションでは、基本的に圧縮符号化時にフレームレートが削減されるため、SRCとPVSではフレームレートは異なる。なお、Multimediaテストでは、SRCは標準テレビ方式で使用されているD1以上の解像度を持つ映像をダウンコンバートして作成するため、フレームレートは30fps (frame per second) ないし25fpsのいずれかとなる。客観的な評価モデルには、SRC、PVSの両方が入力される。ただし、No Referenceフレームワークでは、PVSのみが入力され、SRCの情報は使用しない。

● テストの手順とILG

方式評価テストの基本は、あるSRC/PVSのセットに

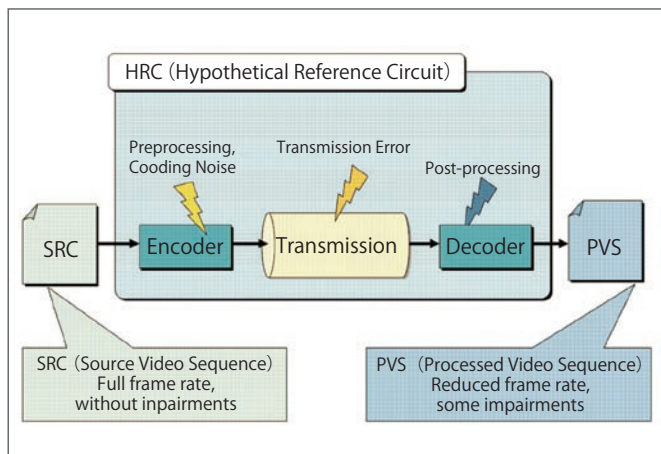


図-6 Multimedia テストにおける画質劣化モデル

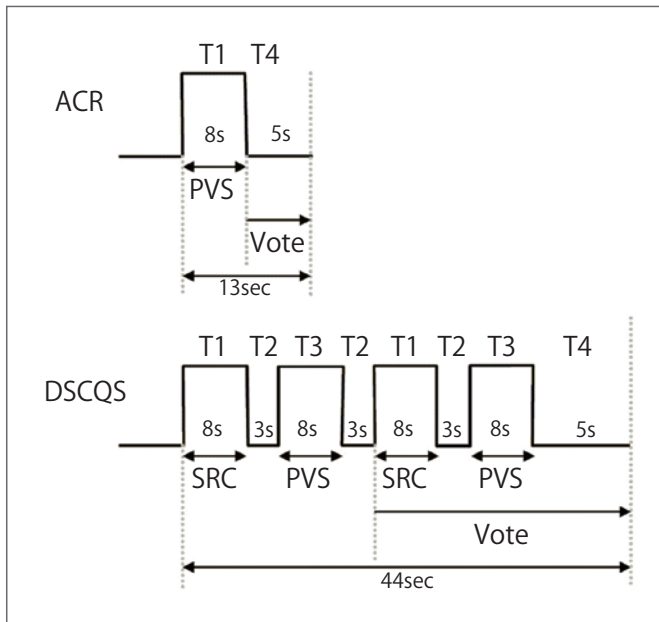


図-7 ACR-HRR法とDSCQS法の所要時間の違い

対する主観評価値と、各機関の提案する客観的な評価方式が出力する客観評価値とがどれだけの相関を持つかを調査する点にある。ただし、SRC/PVSの特徴やHRCの劣化条件をあらかじめ提案機関が知っている場合、それらの条件に対して提案方式を最適化できるため、既知の映像や劣化条件にのみ特化した方式が高い成績を上げてしまう恐れがある。しかし、これでは正しい結果とは言えないため、VQEGではテストの公正さを維持するために、方式提案に関係せず各提案方式の評価を行う独立した第三者機関ILG (Independent Laboratory Group) を組織している。ILGは提案機関が特定のテスト映像や劣化条件に対して提案方式を最適化し、本来の性能以上の成績を上げることを避けるため、テスト映像の選定や劣化条件の決定、および主観評価実験を提案機関には非公開で独自に実施する役目を担う。ただし、Multimediaテストでは、提案方式の数やテスト画像数が膨大な量になり、ILGの処理能力を超えるとの理由から、この手続きの一部は提案機関が分担することとなった。ILGは、PVSの作成や主観評価実験の実施を行うほか、各機関の提案方式の出力した客観評価値のダブルチェックを行うなど、テスト進行の至る所で信頼性を保つための役割を担っている。また、主観・客観的な評価データがすべて出そろった後のデータ処理や性能評価のための各種指標の導出についても責任を負っている。

● 主観的な評価フレームワーク (ACR-HRR法)

主観的な評価法として、ACR-HRR (Absolute Category Rating with Hidden Reference Removal) 法を採用している。ACR法とは、ITU-T勧告P.910に

規定された1重刺激法であり、5段階の品質評価尺度を用いて評価を行う。1重刺激法とは、被験者にPVSのみを提示し、その映像の絶対的な品質を評価する主観評価実験法を意味する。これに対して、評価の際SRCとPVSを組にして提示し、両者の相対的な品質差を評価する方法を2重刺激法と呼ぶ。

Hidden Reference Removalとは、1重刺激の評価映像の中にSRCを被験者にはリファレンスだとは明示せずPVSと同様に提示し、SRCのMOS (Mean Opinion Score ; 1つの映像に対する評価値の平均値) とPVSのMOSの差分を2重刺激法におけるDMOS (Differential MOS) と同等の尺度として導出する手法である。MOS (SRC) をSRCのMOS, MOS (PVS) をPVSのMOSとすると、ACR-HRR法におけるDMOSは以下の式で求められる。

$$DMOS = MOS(PVS) - MOS(SRC) + 5 \quad (1)$$

上式より、DMOSは、最も品質が悪い場合には最小値1に、最も品質が高い場合には最大値5を示す(例外的にMOS(PVS) > MOS(SRC)となる場合には、DMOSは5以上になる)。今回のテストプランで2重刺激法ではなくACR-HRR法を用いているのは、ACR-HRR法では、評価すべきPVSの総数に対する映像の提示回数を2重刺激法に比べて大幅に減らすことが可能である。

図-7は、ITU-R勧告BT.500-11に規定されている代表的な2重刺激法であるDSCQS (Double Stimulus Continuous Quality Scale)法Variant IIとACR-HRR法の1PVS提示あたりの所要時間の比較を示している。なお、同図における比較の条件はMultimediaテストプランおよび勧告の設定値に従い、SRC、PVSの提示時間(T1, T3)を8秒、DSCQS法におけるSRCとPVSの間のグレー表示期間(T2)を3秒、最後のPVS表示終了から次の画像の表示までの期間(T4)を5秒としている。DSCQS法では、1回のPVSの評価のために必ずリファレンスとしてSRCを提示しなければならない。通常、主観評価実験では、同じSRCに対して複数のHRC (符号化条件または伝送エラー条件) を適用するため、その都度同じSRCを提示する必要がある。一方、1重刺激法であるACR-HRRではSRCはHidden Referenceとして1回のみ提示すればよい。また、ACR法では、PVSの提示は1回のみと規定されており、DSCQS法のように繰り返しの提示を行う必要がない。このように、ACR-HRR法では、2重刺激法を持つ冗長性を排除することで主観評価実験の試験時間を3分の1以下に短縮することが可能となっている。なお、ACR-HRR法とDSCQS法の主観評価値の相関を調

査する予備実験が行われたが、両者の相関は依然高く、ACR-HRR法は十分に実用的な主観評価法であるといえる。

●映像フォーマットと映像の劣化条件 (HRC)

アプリケーションの実行環境、映像の再生環境および端末の表示性能のカテゴリー分類として高解像度 PC、低解像度 PC、PDA/携帯電話の3種類を想定し、これらのカテゴリーに対応する解像度として、VGA、CIF、QCIFの3つを規定している。なお、PCおよびPDA/携帯端末での表示を前提としているため、各解像度ともSRC、PVSはプログレッシブ構造とし、色空間はYUV4:2:2に統一されている。また、HRCすなわち映像の劣化条件として、符号化劣化、伝送エラー(パケット損)劣化、符号化の後処理、デインタレース(インタレース走査の画像をプログレッシブ走査に変換する処理)による品質劣化が含まれている。

前述のとおり、各解像度のSRC作成はD1以上の解像度を持つコンテンツを縮小処理することにより作成される。一般に、これらの元素材はインタレース構造であるため、縮小処理に加えデインタレース処理を行い、プログレッシブ構造に変換する必要があるが、動きのあるシーンにおいては、この処理により視覚的な劣化が発生する可能性がある。

●客観的な評価モデルの性能評価

主観的・客観的な評価フレームワークに基づき主観評価値と客観評価値の組が得られた後のデータ処理の主眼は、提案モデルによる客観評価値がいかに主観評価値を近似できているかを解析する点にある。そのため、まず主観評価値を近似する回帰曲線を求め、その後、その近似特性における主観評価値の推定精度を求める。

主観評価値のスクリーニングにより不適切な主観評価値を除外した後、式(1)に基づきPVSごとのDMOSを求め、客観評価モデルの出力した客観評価値との間で回帰分析を行う。ACR-HRR法では、5段階のカテゴリー尺度により主観評価値が与えられるが、一般に、最高点と最低点の付近では評点が圧縮される傾向にある。つまり、主観評価値と客観評価値の関係をプロットした場合、本来直線上に分布すべき関係が主観評価値の最大・最小値の付近でクリップされるような形状を示す場合がある。こうした傾向を反映するためには、ロジスティック関数や高次の多項式による近似が有効であることが経験的に知られている。

Multimediaテストでは、3次の多項式を用いることが規定されている。すなわち主観評価値の予測値をDMOS_pとするとき、

$$DMOS_p = c_3x^3 + c_2x^2 + c_1x + c_0 \tag{2}$$

により近似を行う。なお、この回帰曲線には、当然ながら単調な特性を持たなければならない。

この主観評価値の近似特性を求めることにより、提案方式による最終的な主観評価値の推定値DMOS_pが得られる。この値が実際の主観評価値といかに一致するかが提案方式の性能を左右することになる。Multimediaテストでは、以下の3つの指標とその95%信頼区間により提案方式の性能比較を行う。

- Pearson相関係数：提案方式による主観評価値の推定値DMOS_pと実際の主観評価値DMOSの線形性を求めるために用いられる。
- RMSE：これはRoot Mean Square Errorの略で、DMOS_pの二乗推定誤差の平方根により定義される。
- Outlier Ratio：これは、各PVSのDMOS推定誤差Perror(i)が95%信頼区間を越えるような「外れ値」の割合を示す指標である。

●VQEG京都会合の成果

最後に2008年3月に京都で開催されたVQEG京都会合(NTTがホスト役)で主に議論された、Multimedia品質評価の報告書素案(ver.1.1)について述べる。Multimedia Phase-Iテストは、評価用映像材料における2つの評価で構成されており、1つの評価が人間の観察者による主観評価であり、もう一方が映像品質の客観的な計算モデルである。およそ、40の主観評価実験が、評価モデルの妥当性チェックを行うために必要なデータを提供するために実施された。主観評価実験は、広範囲にわたり、また、圧縮と伝送エラーの両方が含む映像系列で主観評価実験が行われた。モデル提案者は、異なった映像の解像度(VGA、CIF、QCIF)とモデルタイプ(FR、RR、NR)をカバーする最大13個の異なったモデルまで提出することができた。具体的には、3つのFRモデル、7つのRRモデル、そして3つのNRモデルである。RRモデルに関しては、以下の7つのビットレートが想定されていた。

- QCIF : 1kbit/s, 10kbit/s
- CIF : 10kbit/s, 64kbit/s
- VGA : 10kbit/s, 64kbit/s, 128kbit/s

結果として、以下に示す合計13の機構がMultimedia Phase I.の主観的なテストを実行した。

- モデル提案機関：NTT(日本), Opticom(ドイツ)
Psytechnics(イギリス)
SwissQual(スイス), Yonsei大学(韓国)
- ILG : Acreo, CRC/Nortel, IRCCyN, France

Telecom, FUB, Verizon

- PVSの実行と主観評価実験の実施：KDDI, Symmetricon

主観評価実験を実行する前に、モデル提案者は、コンピュータで計算可能な客観評価モデルを提出した。

次に、各モデルに関する性能評価の結果について示す。

FRモデルの結果：

- VGA：平均した相関が0.79～0.82で、PSNRが0.77、モデル相関の最高値は0.939。
- CIF：平均した相関が0.72～0.84で、PSNRが0.76、モデル相関の最高値は0.923。
- QCIF：平均した相関が0.76～0.84で、PSNRが0.69、モデル相関の最高値は0.943。

各提案モデルの性能に関して統計解析により、3つのグループに分けることができた。

VGA：第1グループ (OPTICOM, Psytechnics, Yonsei)

第2グループ (NTT, PSNR)

CIF：第1グループ (OPTICOM, Psytechnics)

第2グループ (NTT, Yonsei)

第3グループ (PSNR)

QCIF：第1グループ (OPTICOM, Psytechnics, NTT)

第2グループ (Yonsei)

第3グループ (PSNR)

しかしながら、1番目と2番目のグループの差は非常に小さく、また、第1グループにおいても、あるテスト条件下ではその性能が低いことが指摘されている。すべての提案モデルは、空間的と時間的に調整された後のPSNRよりも良い性能を示していたことより、VQEGとしていくつかのモデルは、勧告案として提案できるとしている。

RRモデルの結果：

すべてのRRモデルは、PSNRよりも良い結果を得た。

- VGA

PSNR：相関係数の最低0.765、平均0.503、最高0.89

モデル：相関係数の最低0.803、平均0.622、最高0.930

- CIF

PSNR：相関係数の最低0.763、平均0.705、最高0.856

モデル：相関係数の最低0.782、平均0.677、最高0.904

- QCIF

PSNR：相関係数の最低0.689、平均0.367、最高0.819

モデル：相関係数の最低0.791、平均0.645、最高0.886

NRモデルの結果：

NRモデルの性能は、FR・RRモデルよりも低性能となった。VGAとCIFモデルでは、十分な性能が得られ

なかったので勧告案として記述するには難しいと思われる。また、QCIFに関しては、平均した相関関係は、0.70と0.63であった。

今後の課題と展望

映像品質評価法の国際標準化動向としてVQEGに関連する事項について述べてきた。VQEGプロジェクトのロードマップにもあるように、品質評価の標準化作業が開始されてからはや10年が経過している。しかしながら、その経過した時間に見合った成果は、排出されていないと思われる。この最大の原因としては、著作権フリーな評価用映像素材の収集が挙げられる。国際標準化を行うにあたり、品質評価に適した映像素材が必要になるだろう。また、10秒程度のビデオクリップだけでなく、より長時間の映像素材を取り扱う必要性もあると考える。

現在、これらの国際標準化とは別にして、世界レベルで映像品質評価用の装置やソフトウェアが開発・販売されている事実もあり、IEC/TC 100など他の機関でも国際標準化の作業もなされ始めている。このことより、映像品質評価に関しては、より広い視野で他の機関と協調しながら標準化作業を進める必要があるといえる。

よって、究極の映像品質の評価技術は、最終的には人間の感性や認識・理解の領域まで踏み込んで議論することにより、高次脳としての情報処理プロセスのさらなる解明とそれらの知見を映像品質評価へ適用する革新的なアプローチが、この分野の大きなブレークスルーとなることと期待したい。

参考文献

- 1) ITU-R Recommendation BT.500-11 : Methodology for the Subjective Assessment of the Quality of Television Pictures (2002).
- 2) 阿部威郎, 石橋 豊, 吉野秀明: 次世代のサービス品質動向, 電子情報通信学会誌, Vol.91, No.2, pp.82-86 (Feb. 2008).
- 3) 高橋 玲: 音声・映像サービス品質評価・推定技術及び標準化動向, 電子情報通信学会誌, Vol.91, No.2, pp.87-91 (Feb. 2008).
- 4) 杉本 修, 川田亮一: VQEG (Video Quality Experts Group) の動向と関連技術, 電子情報通信学会 Fundamental Review, Vol.1, No.3, pp.27-36 (Jan. 2008).
- 5) VQEG Webpage, <http://www.its.bldrdoc.gov/vqeg/>
- 6) 杉山賢二, 杉本 修, 岡本 淳: 高品質符号化技術とその評価技術, 電子情報通信学会誌, Vol.91, No.4, pp.267-274 (Apr. 2008).

(平成20年4月2日受付)

堀田裕弘(正会員)
horita@eng.u-toyama.ac.jp

1986年長岡技術科学大学大学院工学研究科修士課程電気・電子システム工学専攻修了。現在、富山大学大学院理工学研究部(工学)教授・工学部副学部長を併任。工学博士。メディア品質評価, ITS, 感性情報処理などの研究に従事。