

■3群（コンピュータネットワーク）- 5編（通信品質）

1章 QoE：アプリケーションの品質

（執筆著：高橋 玲）[2011年1月 受領]

■概要■

アプリケーション／サービスとしての通信品質は、ユーザが体感する品質（これを QoE: Quality of Experience と呼ぶ）によって表現される。従来通信サービスの品質は、いわゆる「通信の三品質」と呼ばれる、①伝送品質、②接続品質、③安定品質で記述されてきたが、QoE はより広範囲な概念である。QoE の主な構成要素としては、ネットワーク伝送品質に端末などでのメディア処理品質を加えたメディア品質（例えば、音声や映像の良し悪しなどの快適性）、接続品質に相当するサービスの可用性、安定品質に相当する信頼性、機器の操作性・機能性などが挙げられる。特徴的な点は、ネットワークが関係する品質だけでなく、例えば映像配信サービスにおけるリモコンの操作性など、端末のみに起因する要因も QoE に影響を与えるということである。

本章ではメディア品質に着目し、その評価法について記述する。メディア品質は、人間が音を聞いたとき、あるいは映像を観たときに知覚・認知する「主観品質」により表現される。主観品質を評価する基本的な方法は、主観品質評価法（Subjective Quality Assessment Method）と呼ばれる視聴覚心理実験である。主観品質評価法は、評価の普遍性・再現性確保の観点から国際標準化が進んでおり、評価すべき品質要因に応じて多くの方法が規定されている。一方、評価の効率化、リアルタイム化などの観点から、物理的な特徴量に基づいて主観品質を推定する技術の開発も進められている。これを「客観品質評価法（Objective Quality Assessment Method）」と呼ぶ。客観品質評価法は、その適用領域に応じて様々なアプローチに分類される。

【本章の構成】

本章では、音声の主観品質評価法（1-1 節）、音声の客観品質評価法（1-2 節）、映像の主観品質評価法（1-3 節）、映像の客観品質評価法（1-4 節）、及びこれらの国際標準化動向について述べる。

■3群 - 5編 - 1章

1-1 主観品質評価

(執筆者：林 伸二) [2009年3月 受領]

1-1-1 はじめに

通信システムにおける通話音声の品質を評価する場合、その利用者が人であることから手段は究極的には主観的とならざるを得ない。音声の主観品質評価は人の聴覚・知覚・心理機構を手段として評価する官能試験の一種を意味する。評価の主目的は、そのシステムが利用者を満足させるか否か明らかにすることである。また、通信システム利用者の志向、要望を汲み取ったり、新規に開発するシステムの設計に生かしたり、網やその構成機器の規格適合性を評価したり相互比較を行うなど、様々な目的で主観品質評価は行われる。評価結果は、聴取者が受け取った音声の総合的品質評価であり、音源である発声者の声質、表現力から收音環境、端末機器、通信システムの性能、聴取者の聴力、評価能力まですべてを含んだ総合評価となってしまう。そこで、極力不安定な要因を排除し再現性良く、目的とする品質要因を評価する方法が開発されてきた^{1,2)}。それらは通信システムや機器の発達につれ、時間・空間などの環境変化によらず一定の評価ができるよう国際的に標準化されてきた。本章では現在までに確立されてきた主観評価法について基本的考え方を紹介するとともに、今後通信システムの構成要素を主観評価する場合に有用となるとと思われる点に触れる。

1-1-2 標準系

主観品質評価の信頼性、安定性のために通話系の単純化、要素への分割、環境の整理が行われた。すなわち、基本的通話モデルの構築と送信系、伝送系、受信系への分離である。基本的通話モデルの一つとして正調通話特性 (Ortho Telephonic Response: OTR³⁾) が挙げられる。OTR は、人の通話の基本が図 1・1 のように、自由空間で 1 m の距離で対面会話を行うことであるととらえ、これを電話系で極力実現した特性である。このようにして設定された標準系 (Reference System) に対し、評価対象である系を比較評価し、あるいは、同一の枠組み内で並列的に評価することで、対象の要素的变化のみを評価することができ、試験の信頼性が高まる。標準系の定義は評価目的や標準化の経緯などにより唯一ではないが、以下の系が使用されている。詳細は ITU-T の Telephony Handbook⁴⁾を参照されたい。

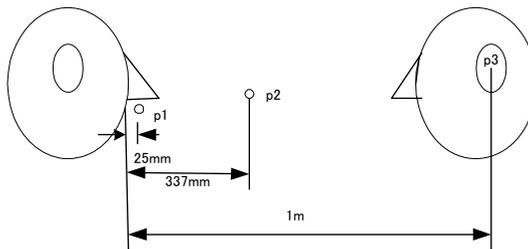


図 1・1 自由空間中の 1 m 標準会話システム

(1) ARAEN (Appareil de Référence pour la détermination de l'AEN)

AEN (affaiblissement équivalent pour la netteté) 明瞭度等価減衰量を定義する標準系である。AENはこの標準系と被測定系に損失を変化しながら加え比較し、明瞭度が被測定系と等しく80%となるとき減衰量の差である。現実的な電話系に合致するよう帯域フィルタ(300-3400 Hz)と雑音付加を含む標準系はSRAEN (Système de Référence pour la détermination de l'AEN)と呼ばれる。電話系の品質がメタリック回線の損失とアナログ搬送装置の減衰ひずみ、雑音に支配されていたころに有効な品質尺度であった。通信システムの発展により明瞭性を満たすだけでは利用者の満足が得られない時代となり、省みられなくなったがARAENは標準系として用いられている。

(2) NOSFER (Nouveau Système Fondamental pour la détermination des Equivalents de Référence)

RE (Référence Equivalents) 通話等量を規定するための基本系であり、人の感じる音量感であるラウドネス (loudness) に基づいて通話品質を評価するための標準系である。ARAENが規定される以前の標準をARAENに適切な等価回路を加えたりマイクロホン位置を整合させたりしてARAENに統合したものである。通話系の品質要因に限られているとき、減衰ひずみや雑音が一定の条件を満たすならラウドネスにより系の品質を評価できるため使用された。ラウドネス評価を安定に行えるよう、標準系よりも電話系に近い周波数特性を有するIRS (Intermediate Reference System) 中間基準系⁵⁾が導入された。REは主観評価によらずとも、マスキング量を考慮に入れたラウドネス計算法⁶⁾などにより客観的に求められるようになったためNOSFER自体は使われなくなったが、電話系の周波数特性としてIRS特性が利用されている。

1-1-3 主観評価手法

通信システムの発展により音声の了解性や音量感が不満足であることは少なくなった。減衰ひずみや雑音、音量などの既知の要因による品質はITU-T 勧告 P.11⁷⁾に整理されている。一方、品質の劣化要因は多様性を増している。拡声系、マルチメディア系、広帯域系へとサービス内容が高度化すること、音声圧縮率が高く、使用条件が厳しい無線通話系と固定電話網と相互に通話が行われること、衛星通信、国際通信など空間的拡がり、IP網とPSTNがゲートウェイで接続されて、サービス品質非保証がありうること、もとよりPSTNを介さずにIPネットワークのみを用いた通話がPC端末間で独自に行われることなどである。このため、通信サービス提供者からみたQoSよりも、利用者の品質感を反映するユーザ体感品質QoE (Quality of Experience)がより重視されるようになった。表1・1に拡張サービスと技術要素、特有の品質要因を示す。FERC (Frame Erasure Concealment) は一定率以上の伝送符号誤りがあるときは符号化アルゴリズムのフレームを廃棄し、以前の正常フレームから音声合成して誤りの影響を隠蔽する、VAD (Voice Activity Detection) は音声の休止区間を検出し、音声情報の伝送率を低下する、CNG (Comfort Noise Generator) は、VADにより省略された無音区間に雑音を付加し有音・無音の切り替えによる不自然さを低減する。PLC (Packet Loss Concealment) は、パケット網でパケット廃棄されたときその区間の音声をFERCと同様の技術で合成し補うものである。

表 1・1 拡張サービスと技術要素、特有の品質要因

通話系	技術要素	品質要素
拡声系	音響結合, 音響エコーキャンセラ	エコー品質, エコーキャンセラ評価
マルチメディア系	画像との同期, 協調	映像と音の時間差
広帯域系	帯域拡張, 広帯域符号化	広帯域・高品質評価
無線モバイル網	高圧縮符号化, 背景雑音, 符号誤り, FERC, VAD/CNG	背景雑音時評価, FERC, VAD/CNG 評価
IPGW VoIP	パケットジッタ, 遅延時間, 高圧縮符号化, 背景雑音, 符号誤り PLC, VAD/CNG, QoS 非保証混在	エコー品質, 遅延時間, PLC, VAD/CNG 評価

これらの要因はおおむねデジタル処理に関するものであり、新しい処理アルゴリズムが開発される時、その品質の特徴はユーザにとって未知であることが多い。そこで、これらの要因の評価手段として、QoEを表すのに適したオピニオン評価が主に用いられる。オピニオン評価はACR (Absolute Category Rating) と呼ばれるカテゴリー評価であって表 1・2 に示すカテゴリー評点間に距離の意味は保証されず、たかだか順位を示すものである。しかし、定量的評価値を得るため、評点の平均値 MOS (Mean Opinion Score) を利用している。これは一定の条件のもとではカテゴリー間の距離が心理的評価尺度とほぼ線形な関係にあることが知られているためである。オピニオン評価などの心理的測定法については、特に音に関する心理評価という側面でまとめられた難波・桑野の書⁸⁾を参考にされたい。オピニオン評価の実施については、ITU-T 勧告 P.800⁹⁾ に整理されているのでおおむねこれに沿って概説する。拡声会議などには、電話帯域よりも広い 7 kHz 帯域が利用され、更に 15 kHz, 20 kHz の広帯域通話系も開発されている。広帯域系はひずみの小さい高品質伝送を目指しており、主観評価法は放送スタジオ品質の評価法がよりふさわしいと見られている。

表 1・2 ACR カテゴリー

評価カテゴリー	評点
Excellent 非常によい	5
Good よい	4
Fair まあよい(ふつう)	3
Poor 悪い	2
Bad 非常に悪い	1

(1) 試験の種類

(a) 実験室試験とフィールド試験

通話システムの実態を最も良く反映した評価を望むなら、実際の通話システムに試験条件を設定して、呼ごとに利用者に評価を入力してもらうフィールド試験がよい。しかし、この方法は相当数の利用者に協力を依頼し、多数の呼の評価結果が集まるまで長期間を要し、また、評価入力システムを付加するにも相当のコストを要するうえ、通話条件の制御が容易で

ないという問題点がある。そこで主観評価は専ら実験室試験において、試験条件を制御して行うことになる。形態は受聴試験、送話試験、会話試験がある。

(b) 会話試験

遅延時間やエコー、側音に関する評価は、被験者自身が送話してその反応を受けて初めてできるものであるため、会話試験によらざるを得ない。また、通話サービス実態をよりよく反映するのが単なる受聴ではなく会話であることは論を待たない。会話試験は被験者が送話者にもなるため、試験施設の室内音響環境、雑音、騒音などの条件を ITU-T 勧告 P.800 Annex A に定めている。更に会話試験の方法、実験設備、試験設計、被験者の選定、会話のタスク、評価結果の統計処理などが ITU-T 勧告 P.805¹⁰⁾として勧告された。

(c) 送話試験

会話試験の形態の一種で、相手話者が送話しない特殊な場合といえ、側音やエコーの評価に用いられる。会話試験に比べ通話の実態にそぐわない不自然さがあるため、可能なら会話試験を選ぶべきである。

(d) 受聴試験

条件を厳密に制御でき、比較的短時間のうちに多くの条件を評価することができる受聴試験は最も実施しやすい形態である。ITU-T 勧告 P.800 の Annex B に受聴試験としての ACR 試験法が勧告されている。同勧告 Annex D には劣化カテゴリ評価 DCR (Degradation Category Rating), Annex E には比較カテゴリ評価 CCR (Comparison Category Rating), Annex F には標準系との比較による閾値法が勧告されており、評価の目的に応じて適切に使用されるべきである。また、近年の音声符号化方式を含むデジタル系の評価に関しては、ITU-T 勧告 P.800 に準拠しながら、その特殊性を考慮した評価法が ITU-T 勧告 P.830¹¹⁾、そのためのツールが ITU-T 勧告 G.191¹²⁾として勧告化されている。

(2) P. 800 準拠の実験

(a) 音源収録

静粛 (30 dBA 以下) な、残響が過多でない (500 ms 未満、望ましくは 200-300 ms)、程よい容積 (30 m³ 以上 120 m³ 以下) の環境で、例えば ARAEN 送話系を用いて、ITU-T 勧告 P.64¹³⁾ に従って較正後収録する。音声符号化方式などの評価には ITU-T 勧告 P.830 Annex D に勧告される修正中間基準系 (Modified IRS) の利用が適切である。ITU-T 勧告 P.48 の IRS は電話送話系の周波数特性を反映したものであるが、近年の電子化電話機の低域減衰は緩やかであり、急峻な低域遮断フィルタは不要なうえ、デジタル圧縮アルゴリズムに適さないため修正が施された。図 1・2 に送受 Modified IRS と参考の SRAEN 特性を示す。送話系には、IEC581-5¹⁴⁾ に準拠する平坦な周波数特性、低雑音、低ひずみのマイクロホン、アンプを用い、マイクロホンは送話者の 140 - 200 mm に設置して収録することが推奨されている。発声レベルは ITU-T 勧告 P.56¹⁵⁾ 音声レベル計でモニタし、クリッピングひずみを避ける。その他、音声素材の選定、録音保存の実効レベルの正規化法、送話者の年齢男女比など注意すべき条件が記載されている。

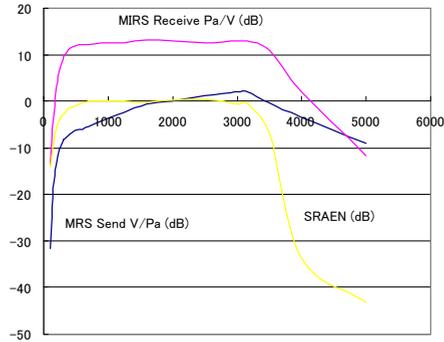


図 1・2 送話 Modified IRS 周波数レスポンス

(b) 実験条件

受話音量は評価に影響を与えるが、オピニオン評価では好ましい音量で行われればよく、室内が静かであれば、片耳レシーバ受聴で、79 dB SPL が好ましいとされる。送話者の声質が評価に影響するのを避けるため、男女各 2 名以上の音声素材を用いるべきである。実験のセッションごとに標準の品質条件（レファレンス）を適切に配置して、異なった環境下の実験と比較できるように配慮する。この目的に供する ITU-T 勧告 P.810¹⁶⁾ MNRU (Modulated Noise Reference Unit) は、G.711¹⁷⁾ で発生するような、音声信号と振幅相関のある雑音を含む音声であり、電話利用者になじみのある音質特徴を持ち、SN 比を任意に制御できるため標準信号として有用である。

(c) 実験設計

ACR 試験はその名のとおり絶対評価であり、被験者が内にもつ尺度により評価する。しかし、心理的に相対評価を下す特性があるため、枠組の影響に注意が必要である。一つのセッションに高品質な条件が大半を占めれば相対的に低い条件が極端に低く評価される危険があり、また、「帯域制限」と「信号対雑音比」というように異なった要因を同時に評価する枠組み設定した場合、被験者の内的尺度が安定しない。1 セッションの時間は 20 分程度として、被験者の疲労を避け、また、条件の提示順序の影響を避けるため、異なったランダム順の複数のセッションを用いる。

(d) 受聴試験手順

1) ACR 試験

オピニオン評価基準は、品質に応じて表 1・2 のように評点を配する。訳語に応じて微妙に意味が異なる可能性があり、国際間の MOS 値の比較では、ずれが含まれることがある。

2) DCR (劣化カテゴリー評価) 試験

ITU-T 勧告 P.800 Annex D に定義される。高品質な標準系に対し被試験系の劣化を表 1・3 に示す劣化カテゴリーで表す。被試験系の劣化が比較的小さく条件間に明確な差が得られないとき使用される。

3) CCR (比較カテゴリー評価) 試験

ITU-T 勧告 P.800 Annex E に定義される、DCR 試験法の変形である。提示されるカテゴリーに標準系よりもよい評価が用意される。これは、被試験系に音声強調、雑音抑圧などの処

理を含む場合に使われることがある。被験者には表 1・2、表 1・3 のカテゴリと評点を示し、標準音の後に被試験音を一対にして提示し、後者が前者に比べてどうか意見を求める。

表 1・3 DCR カテゴリ

評価カテゴリ 訳語	評点
Degradation is inaudible. 劣化が聞き取れない	5
Degradation is audible but not annoying. 劣化が聞き取れるが、煩わしくない	4
Degradation is slightly annoying. 劣化が少し煩わしい	3
Degradation is annoying. 劣化が煩わしい	2
Degradation is very annoying. 劣化が非常に煩わしい	1

(3) 結果の評価

評価結果は限られた数の投票から得るため、信頼度幅を算出する必要がある。誤差 ε がある幅の内に入る確率は式(1・1)で表される。95%の信頼度幅では、 $\delta = 0.95$ として、投票数 n があまり大きくないとき、 σ をある条件のオピニオン評点分布の標準偏差とすると、式(1・2)の t_δ を student の分布関数 ϕ を用いた表から検索して

$$P(-t_\delta \frac{\sigma}{\sqrt{n}} < \varepsilon < t_\delta \frac{\sigma}{\sqrt{n}}) = \delta \quad (1 \cdot 1)$$

$$t_\delta = 2 \int_0^\delta \phi[t, (n-1)] dt \quad (1 \cdot 2)$$

信頼度幅 95%の境界値 ε は式(1・3)となる。

$$\varepsilon = t_\delta \sigma / \sqrt{n} \quad (1 \cdot 3)$$

(4) 広帯域高品質系の主観評価

双方向使用時の品質（エコー、遅延の影響など）を別にして、受聴試験に限れば放送スタジオ品質の評価に適した BS シリーズの評価法が利用できる。

(a) ITU-R 勧告 BS. 1116¹⁸⁾

品質劣化が非常に少ない場合の厳密な品質評価法を規定する。レファレンスはいかなる劣化も受けないオリジナル音である。被験者、試験者ともレファレンスと被試験系の区別がつかないダブルブラインド環境で、レファレンスと被試験系を提示し被試験系の劣化の程度を評価する。試験に用いる音源も劣化の分かりやすいものを選び、被験者も評価感度の高い者とし、受音環境も厳しく規定する。

(b) ITU-R 勧告 BS. 1285¹⁹⁾

ITU-R 勧告 BS.1116 は、劣化が極めて小さい系の評価のために大きな労力をかけて精密に実施する試験であるので、劣化の種類によっては不要な労力をかけるおそれがある。無駄を避けるために、事前に実施すべき予備試験を定めている。

(c) ITU-R 勧告 BS. 1534²⁰⁾

MUSHRA (MULTi Stimulus test with Hidden Reference and Anchor) と称され、ITU-R 勧告 BS.1116 ほど精密さを要さない、準高品質の対象に信頼できる結果の得られる試験法を定めている。

(5) ソフトウェアツール ITU-T 勧告 G. 191

ANSI-C ソースコードで提供され、開発中の符号化アルゴリズムに組み込んで ITU-T 勧告 P.800, ITU-T 勧告 P.830 に準拠した試験用音声を作成できる。これにより符号化アルゴリズム周辺の特性の違いによる評価の差異を防ぐことができる。主な機能は以下のとおりである。

- ① 折り返し禁止フィルタつきサンプリングレート変換 フィルタ特性として、ITU-T 勧告 G.712 準拠 PCM 装置周波数特性、平坦特性、ITU-T 勧告 P.48 準拠 IRS 特性、ITU-T 勧告 P.830 Annex D 準拠 Modified IRS 送話及び受話特性など
- ② ランダム及びバースト発生パターンビット誤り、フレーム消失挿入
- ③ ITU-T 勧告 G.711 Appendix 1 の PLC, ITU-T 勧告 P.810 MNRU, ITU-T 勧告 P.56 音量計
- ④ 現行標準符号化方式 ITU-T 勧告 G.711, G.726, G.727, G.722, 及び RPE-LTP (GSM フレート標準)

■参考文献

- 1) 三浦敏敏, “新版 聴覚と音声,” 第4部, コロナ社, 1980.
- 2) 北脇信彦, “音のコミュニケーション工学,” 5章, コロナ社, 1996.
- 3) Inglis, A. H., “Transmission features of new telephone sets,” Bell Syst. Tech. J.17, p.358, 1938.
- 4) “Handbook on Telephony,” CCITT, 1992.
- 5) “Specification for an intermediate reference system,” CCITT Recommendation, p.48, 1988.
- 6) Fletcher, H., “Speech Hearing in Communication,” Von Nostrand Inc., New York, 1953.
- 7) “Effect of transmission impairments,” ITU-T Recommendation, p.11, 1993.
- 8) 難波精一郎, 桑野園子, “音の評価のための心理学的測定法,” コロナ社, 1998.
- 9) “Methods for subjective determination of transmission quality,” ITU-T, p.800, 1996.
- 10) “Subjective evaluation of conversational quality,” ITU-T Recommendation, p.805, 2007.
- 11) “Subjective performance assessment of telephone-band and wideband digital codecs,” ITU-T Recommendation, p.830, 1996.
- 12) “ITU-T Software Tool Library 2005 User's manual,” ITU-T G.191, 2005.
- 13) “Determination of sensitivity/frequency characteristics of local telephone systems,” ITU-T, p.64, 2007.
- 14) “High fidelity audio equipment and systems; Minimum performance requirements - Part 5: Microphones,” IEC Publication, vol.581-5, 1981.
- 15) “Objective measurement of active speech level,” ITU-T Recommendation, p.56, 1993.
- 16) “Modulated noise reference unit (MNRU)” ITU-T, p.810.
- 17) “Pulse Code Modulation (PCM) of Voice Frequencies,” ITU-T, G.711, 1988.
- 18) “Methods for the subjective assessment of small impairments in audio systems including multi-channel sound systems,” ITU-R, BS.1116, 1997.
- 19) “Pre-selection methods for the subjective assessment of small impairments in audio systems,” ITU-R, BS.1285, 2007.
- 20) “Method for the subjective assessment of intermediate quality level of coding systems,” ITU-R, BS.1534.

■3群 - 5編 - 1章

1-2 音声の客観品質評価

(執筆者：北脇彦彦) [2009年3月 受領]

1-2-1 客観品質評価の基礎

人が音声を試聴することによって品質評価する方法を、主観評価法という。主観評価によって求めるユーザ体感品質 (QoE: Quality of Experience) を、計測した物理量から数式またはアルゴリズムによって推定する方法が、客観品質評価法である。QoE と対応がよい客観品質評価法を得るために、聴覚心理的 (Auditory and Psychological) な知見が利用される。本節では、音の大きさの知覚、周波数選択性及び心理学的測定法について述べる¹⁻⁴⁾。

(1) 音の大きさの知覚

重要なことは、耳に入ってくる音の物理的特性とそれが引き起こす感覚との関係を可能な限り定量的に記述することである。人間の耳は、その感度のよさという点でも可聴レベルの広さという点でもすばらしい能力を持っている。音の大きさのことをラウドネス (loudness) という。ラウドネスは感覚的な量であり、周波数や強さによって変化する。音の物理量を感覚量であるラウドネスと結びつける尺度化の研究は、S.S. Stevens (1957) 以来、多くの知見がある。Stevens によると、音の大きさ L は、音の物理的な強さ I のべき関数、 $L = kI^{0.3}$ で表されるという。ここで、 k は被験者などによって決まる定数である。つまり、レベルが 10dB 上がるごとに感覚的な音の大きさは 2 倍になるということである。これらの知見は電話系における音の大きさに関する客観評価に適用されている。

(2) 音の周波数選択性

フーリエ級数によれば、音声のような複合音は正弦波の和からなっている。音の周波数選択性というのは、複合音中の正弦波成分を分解する能力のことをいう。この能力は蝸牛内にある基底膜に備わっており、基底膜上の各位置は限られた範囲の周波数に応答するものと考えられている。周波数選択性はマスキングを利用することによって測定される。マスキングとは、ある音の最小可聴値が他の音の存在によって上昇する現象またはその上昇量のことである。

Fletcher (1940) は、帯域雑音をマスキャー、正弦波をマスキャーに用いた聴覚実験を行い、聴覚抹消系は帯域が連続的に重なり合う帯域通過フィルタ群を含む機能を持っていると結論づけている。このフィルタを聴覚フィルタ (Auditory Filter) という。また雑音の帯域をそれ以上広げても信号のマスキング閾値が上昇しなくなる帯域幅を臨界帯域幅という。

人間の可聴周波数帯域は 24 チャンネルの帯域通過フィルタ群に分割され、基底膜上で 1.5 mm の間隔が 1 臨界帯域幅に対応するといわれている。また、臨界帯域で測った心理的な周波数と物理的な周波数の関係を示す心理尺度として、バーク (bark) がある。

(3) 心理学的測定法

一般に、ユーザが感ずる品質というのは、各人の心の中に大小の量的な性質が備わっているものである。ある閾値 (Threshold) を越えると感覚が生じ、弁別閾 (Differential Limen) を越えるとある状態の感覚が次のレベルへ変化したと認識される。精神物理学的測定法においては、このような心理学的連続体上における弁別過程の散らばりは、正規分布をすることが仮定されている。

図 1・3 に示すように、弁別過程がカテゴリーの境界のすべてに及んでおり、かつその連続体上において正規分布をなしている刺激 S_1 について考えてみよう。 S_1 に対する最頻弁別過程が I_C であるとする、 C の反応が起こるはずである。しかし、 S_1 のときには I_A, I_D とも対応するので、 A または D の反応も起こる。その割合はカテゴリー区間内の弁別の散らばりの面積で表される。正規分布の仮定が正しいとすれば、標準偏差 σ を単位にとることによって、平均値から各カテゴリーの境界までの距離を得ることができる。心理尺度化にあたっては Thurstone の方法がよく用いられる³⁾。

心理学的知見は複数の要因が複合して存在するときの総合の品質 MOS (Mean Opinion Score) を推定するオピニオンモデルに適用されている。

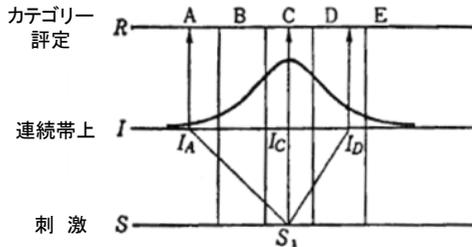


図 1・3 カテゴリーの心理尺度化[3]

(Copyright © 1977 by Yoshihisa Tanaka)

1-2-2 聴覚心理的知見に基づく客観品質評価法

聴覚心理的知見を利用した客観品質評価法の実例として、単独品質要因に対する PSQM (Perceptual Speech Quality Measure) と、複合品質要因に対する OPINE (Overall Performance Index model for Network Evaluation) について述べる。

(1) PSQM 法

PSQM 法は、聴覚内部表現として、臨界帯域幅で分割された聴覚フィルタを用いて、耳内音圧スペクトルをパークスペクトルとして抽出し、ラウドネス補償を施したひずみ量を算出する。算出したひずみ量を、主観品質を表す MOS にマッピングする。

PSQM は符号化ひずみのように時間連続的なひずみを評価対象としていたが、IP (Internet Protocol) 電話におけるパケット損のような時間離散的なひずみにも対処できるように改良され、2001 年に ITU-T 勧告 P.862 PESQ (Perceptual Evaluation of Speech Quality) として標準化された⁵⁾。PESQ の構成を図 1・4 に示す。ひずみ量を聴覚内部表現量として求めるところは PSQM と同じであるが、4 ms 程度の短区間ごとのパワーの変化量を観測して、時間不連続な区間が発見されるとパケット損が起こっている区間とみなして、荷重を調整するアルゴリズムになっている。

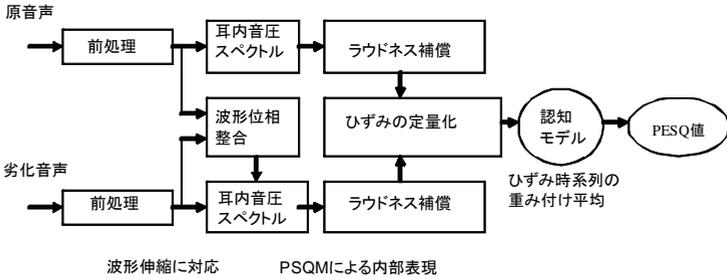


図 1・4 符号化ひずみとパケット損による品質を推定する PESQ の構成[4]

(2) OPINE 法⁶⁾

電話系の総合品質には複数の品質要因が影響を及ぼしている。品質要因が複合して存在するときの総合品質を推定するオピニオンモデル OPINE のコンセプトを図 1・5 に示す。OPINE では、品質評価過程を聴覚における処理過程と心理評価過程の縦属接続として構成している。聴覚系処理過程では、電話系物理パラメータで表される品質要因を、Fletcher モデルに基づく聴覚内部表現量に変換して記述する。

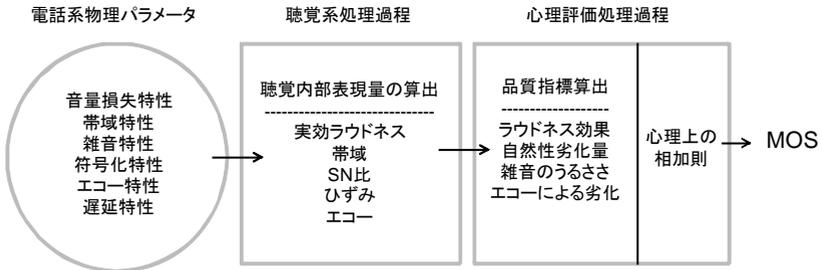


図 1・5 品質要因が複合して存在するときの総合品質を推定するオピニオンモデル OPINE⁶⁾

心理評価処理過程では、評価に影響を及ぼすいくつかの要因に対して、Thurstone の心理モデルに基づく品質指標 PI (Performance Index) を算出する。ここでは、ラウドネス (PI_{EL})、雑音 (PI_N)、帯域制限 (PI_{BL}) の 3 要因の例を図 1・6 に示す。要因ごとの PI は、ある物理量に対応する原点を有する心理尺度値である。総合的な劣化を表す OPI (Overall PI) は、要因ごとの PI の心理尺度上で加算されたものとする。心理尺度上の総合評価の平均値 P は、品質劣化がない状態 P_0 から総合劣化 OPI を減算することによって決定される。これを「非常に良い」から「非常に悪い」のカテゴリーに対応させることによって MOS が算出される。

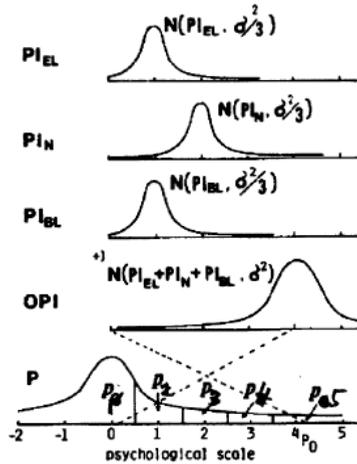


図 1・6 Thurstone の心理モデルに基づく品質指標の算出⁶⁾
(Copyright © 1985 IEICE)

1-2-3 IP 電話の通話音質客観評価法

グラハム・ベルによる電話の発明以来、一貫して続いてきた交換接続による電話網は、ルータやサーバを用いた IP 網による次世代ネットワーク NGN (Next Generation Network) へと、130 年ぶりの大変革を遂げようとしている。本節では、1-2-1 節及び 1-2-2 節で述べた聴覚心理基礎技術をベースにした IP 電話の通話音質客観評価法について述べる。

(1) 客観評価法の分類

客観評価と一口に言っても、様々な品質要因をもつ通信系に、万能な客観評価法は存在しない。表 1・4 に用途別にみた客観評価技術の分類を示す。機器や装置のベンチマーク、QoE 設計やプランニング、QoE 管理やモニタリングへの観点から分類すると、音声信号を入力として受聴品質を推定する音声レイヤ客観評価 (ITU-T 勧告 P.800 シリーズ)、品質パラメータを入力として総合品質を推定するオピニオンモデル (ITU-T 勧告 G.100 シリーズ)、IP パケット情報を入力として受聴品質を推定するパケットレイヤ客観評価 (ITU-T 勧告 P.500 シリーズ) に分けることができる。

表 1・4 別にみた客観評価技術の分類⁴⁾

品質測定用途	ベンチマーク	プランニング	モニタリング
カテゴリ	音声メディアレイヤ	オピニオンモデル	パケットレイヤ
入力情報	音声信号	品質パラメータ	IP パケット
品質推定法	FR法	Model法	NR法、RR法
勧告例	P.862, P.862-2	G.107, G.1070	P.563, P.564

客観品質推定アルゴリズムへの入力信号から客観評価法を分類すると、原信号と劣化信号を比較することで品質推定する FR (Full-Reference) 法, 劣化信号のみを用いる NR

(Non-Reference) 法, 原信号の特徴量に関するサイド情報と劣化信号を用いる RR (Reduced-Reference) 法とに分けられる。

電話帯域 (3.4 kHz) 符号化音声とパケット損を品質推定対象とする勧告 P.862 PESQ や, 広帯域 (7 kHz) 符号化音声とパケット損を対象とする勧告 P.862.2 Wideband-PESQ は FR 法である。In-Service で Non-Intrusive な品質推定を行う勧告 P.563 や RTP/RTCP (Real-time Transport Protocol/RT Control Protocol) パケットのヘッダ情報のみを用いる勧告 P.564 は, NR 法を用いた客観評価法である。

オピニオンモデルに関しては, OPINE をベースとした E-model が, 1998 年に勧告 G.107 として標準化された⁷⁾

(2) IP 電話のオピニオンモデル

E-model による総合品質推定アルゴリズムを図 1・7 に示す。E-model では, 端末要因・ネットワーク要因・環境要因に起因する品質評価値を心理尺度上で加減算することによって, 総合の品質を表す指標 (R 値) を得る。

R 値は 0~100 の数値で表され, S/N 比をベースとする基本的な系の心理評価値 (R_0) から, 「音量感 (I_s)」「遅延・エコー感 (I_d)」「ひずみ感 ($I_{e,eff}$)」といった品質要因ごとに心理尺度上で表現した評価値を, 加減算することによって計算される。A はモバイルなどの利便性が満足度に与える効果を補完するものである。

E-model には, 品質要因に関する 20 の入力パラメータがあり, 出力評価値である R 値はこれらのパラメータの関数として表現される。我が国の標準化機関である TTC (Telecommunication Technology Committee) においては, E-model の入力パラメータのうち環境要因や端末要因などは標準的な特性を想定した既定値を用いることとしており, 評価すべきパラメータを「音声符号化」「パケット損」「遅延」「エコー」に絞り込んでいる。すなわち, 得られた R 値は, 標準的な端末を用いて標準的な音響環境で通話をしたときの通話品質を表したものである。

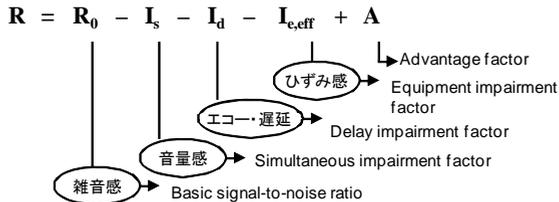


図 1・7 IP 電話の総合品質を推定する E-model アルゴリズム⁴⁾

また, R 値は MOS との間に一定の関係があり, 勧告 G.107 には両者のマッピング関係式が示されている。これらの結果から, 総務省では電話品質を表 1・5 のようにクラス分けしている。

我が国の IP 電話サービスに対する通話品質基準は, E-model に基づいて算出される評価指標である R 値によって規定することが, 総務省令によって決定された。これに対応して, TTC では具体的な通話品質評価法を標準化した⁸⁾。TTC 標準では, 総合の会話 MOS と対応する R 値, 及び R 値を補完する受聴 MOS と対応する PESQ を採用している。

表 1・5 品質クラス分類の例⁷⁾

品質 品質尺度	クラスA	クラスB	クラスC
	固定電話並	携帯電話並	IP電話並
総合品質(R値)	>80	>70	>50
伝送遅延	<100 ms	<150 ms	<400 ms

■参考文献

- 1) B.C.J. ムーア, “大串健吾 監訳, 聴覚心理学概論,” 誠信書房, 1994.
- 2) 三浦種敏 監修, “聴覚と音声,” 電子情報通信学会, 1980.
- 3) G.P. Gullford, “Psychometric Methods,” N.Y.: McGraw-Hill, 1st ed. 1936.
- 4) 田中良久, “心理学的測定法,” 東京大学出版会, 1977.
- 5) “Perceptual evaluation of speech quality (PESQ),” ITU-T Recommendation p.862, Feb. 2001.
- 6) 小坂直敏, 箕一彦, “基本的支配要因を対象とした通話品質客観評価モデル,” 電子通信学会論文誌, vol. J68-A, no. 1, Jan. 1986.
- 7) “The E-model, a computational model for use in transmission planning,” ITU-T Recommendation G.107, July 2002.
- 8) “IP電話の通話品質評価法,” TTC 標準 JJ-201.01, April 2003.

■3群 - 5編 - 1章

1-3 映像の主観品質評価

(執筆者：川田亮一) [2011年1月 受領]

映像の品質評価には主観品質評価と客観品質評価がある。前者では人間が映像を観察することによりその品質を数値化する。後者は、これを装置により行う。人間の評価をなるべく精度良く再現できるように、研究開発が進められている。主観評価法については、その結果を互いに比較可能とするように、標準方式が定められている。本節ではその代表的なものについて解説する。

1-3-1 評価対象

(1) 映像の QoE の決定要因

映像の品質評価を行う際には、何を評価対象としているのかに注意する必要がある。例として家庭でテレビを見る場合、その映像の品質は、大別して次の3段階を通して決まる。

- ・ 撮像系. ここではテレビカメラの品質が映像品質を決定する^{1,2)}。
- ・ 圧縮・伝送系. ここでは、伝送帯域制限のための圧縮や、伝送時のビットエラーが、映像品質に影響を及ぼす。テレビの国際中継の場合には、フレームレートの変換処理^{3,4)}が入ることもあり、この影響も無視できない。
- ・ 表示系. ディスプレイの品質は主観評価値に大きな影響を及ぼす。特に、最近の家庭用薄型テレビは、表示品質向上のため高度な映像処理・表示制御を組み込んでいる^{5,6)}。画面サイズはもとより、この映像処理の主観品質への影響も極めて大きい。

表示系が映像の主観品質へ及ぼす影響は非常に大きいですが、その評価は実際のディスプレイを使用して比較評価するしかない。以下では、その前段階、特に圧縮・伝送などの処理系の品質評価について説明する。

(2) 映像のデジタル圧縮伝送ノイズ

地上波デジタル放送のように、映像をデジタル圧縮して伝送する場合に重畳しうるノイズは、符号化ノイズと伝送路エラーノイズに分かれる。

(a) 符号化ノイズ

符号化ノイズは、伝送のあるなしにかかわらず、映像のデジタル圧縮符号化 (MPEG-2 や ITU-T 勧告 H.264 など) に伴って必ず発生するノイズである。デジタル圧縮符号化では、一般に、映像を構成する1枚1枚の画像の相関を利用する。このため圧縮しやすい絵柄とにくい絵柄が存在する。つまり、同じ圧縮ビットレートでも、品質はコンテンツに大きく依存する (PSNR が 10dB 程度異なってもおかしくはない)。

符号化ノイズは、その見え方から、ブロックノイズやモスキートノイズなどに分類される。

(b) 伝送路エラーノイズ

伝送路エラーノイズは、例えば放送で電波状況が悪い場合に、誤り訂正によっても回復不可能な量のビットエラー (符号誤り) があるときに発生する。映像がストライプ状に崩れたりフリーズしたりするため、非常に目立つ劣化となる。

同様のノイズが、IP 伝送におけるパケット損失やパケット遅延ジッタ、更には圧縮ストリームを復号するデコーダの障害によっても発生する。このため、放送の監視用途としては、IP など下位レイヤだけでなく実際にベースバンド映像まで戻した信号を見ることが必須である。

(3) 評価用テスト映像

上記では処理系の劣化要因としてデジタル圧縮符号化ノイズについて説明した。しかしその他の処理、例えばフレームレート変換処理では、これとは別の劣化が発生するので、評価ポイントが異なる。このほか、映像の高画質拡大処理など、処理によってどのような部分に劣化が発生するかが異なるため、各処理に適したテスト映像を選択することが重要である^{7,8)}。

(4) 主観評価法の分類

主観評価法を大別すると、2 重刺激法と 1 重刺激法に分けることができる。

- ・ 2 重刺激法：評価したい映像と比較対象映像を比較しながら評価値を出す。DSCQS (Double stimulus continuous quality-scale ; 2 重刺激連続品質評価尺度) 法や DSIS (Double stimulus impairment scale ; 2 重刺激劣化尺度) 法などがある⁹⁾。
- ・ 1 重刺激法：評価したい映像のみを提示して評価値を出す (絶対評価)。SSCQE (Single stimulus continuous quality evaluation ; 単一刺激連続品質評価) 法などがある⁹⁾。

更に、1 重刺激法の枠組みを利用しながら 2 重刺激法と同様の結果が得られる新しい方式として、SAMVIQ (Subjective assessment methodology for video quality) 法がある¹⁰⁾。DSCQS や DSIS、SSCQE がテレビジョン評価を主目的として作成されたのに対して、SAMVIQ は PC や携帯端末などのマルチメディアを意識して作成された。

以下、これらの方式について説明する。

1-3-2 DSCQS 法

DSCQS 法は、常に 2 種類の映像を比較しながら評価を進め、0-100 の連続値で点数をつけていく (図 1-8) ため、2 重刺激連続品質評価尺度と呼ぶ。

この方式は、ITU-R 勧告 BT.500⁹⁾の第 5 節 (Annex 1) において、以下を含む項目が細かく規定されている^{*}。

- ・ 主観評価者の人数 (非専門家の評価者を 15 人以上)
- ・ ディスプレイの大きさ、明るさ、コントラスト
- ・ ディスプレイから評価者までの距離
- ・ 映像の呈示順番、時間 (2 種類の映像 [各 10 秒間] を交互に 2 回繰り返して呈示)

評価実験に先立って、評価者への事前の説明や、評価作業に慣れるための練習セッションが行われる。この準備作業は、評価データを信頼性の高いものにするために、重要である。

^{*} BT.500 は、テレビジョンの主観画質評価法の勧告として、初版が 1974 年に作成された。以来、改定を重ね、現在は第 12 版となっている。従来より広く使用されている勧告である。

実験時間は、評価者の集中力が持続するよう、30分以内と勧告されている。

DSCQS による品質劣化尺度は、伝送品質の要求条件の勧告で指標として用いられるなど、重要である。例えば ITU-R 勧告 BT.800¹⁾では、テレビの番組素材伝送用のデジタル圧縮符号化の要件として、圧縮映像と原映像の品質差を DSCQS で 12%以下と勧告している。

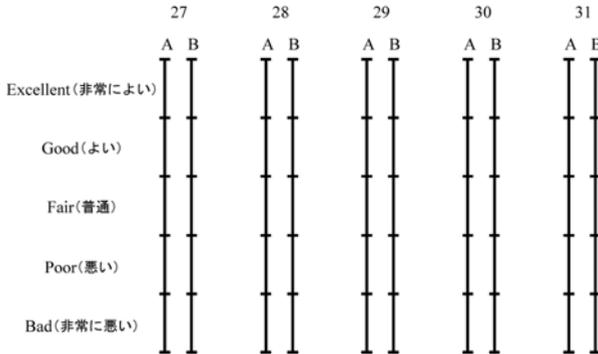


図 1・8 評価シートの一部

1-3-3 DSIS 法

この方式は ITU-R 勧告 BT.500⁹⁾の第 4 節 (Annex 1) に規定されている。

上記の DSCQS 法では、順番に呈示する 2 種類の映像を共に評価するが、DSIS 法では、先に呈示する映像が原画、次に呈示する映像が評価対象映像と決まっている。評価は、原画との差異の大きさを、次のとおり、5 段階で表す。

- 5 : imperceptible (差異がわからない)
- 4 : perceptible, but not annoying (差異が分かるが気にならない)
- 3 : slightly annoying (気になるが、妨害にならない)
- 2 : annoying (妨害になる)
- 1 : very annoying (非常に妨害になる)

1-3-4 SSCQE 法

この方式は ITU-R 勧告 BT.500⁹⁾の第 6.3 節 (Annex 1) に規定されている。

DSCQS 法と DSIS 法では、上記のとおり、2 種類の映像 (各 10 秒間) を比較しながら評価していく。一方、SSCQE 法では、5 分間以上の長さを持つ映像について一つずつ、絶対評価を行う。その際には、時間の経過につれて変動する評価値を、専用のスライダーを動かしながら、連続的に記録していく (図 1・9)。評価値は、連続評価尺度 (非常に悪い・・・非常に良い) である。

SSCQE 法では、実際の視聴環境になるべく近いかたちで品質評価をすることになる (例

えば家庭でのテレビ視聴では、2重刺激のように原画と見比べるわけではない）。また、映像のデジタル圧縮の符号化劣化量が、時間を追うごとに（絵柄によって）変動することに対応している。更に、本手法によれば、音声も含めた総合評価を実施することができる。

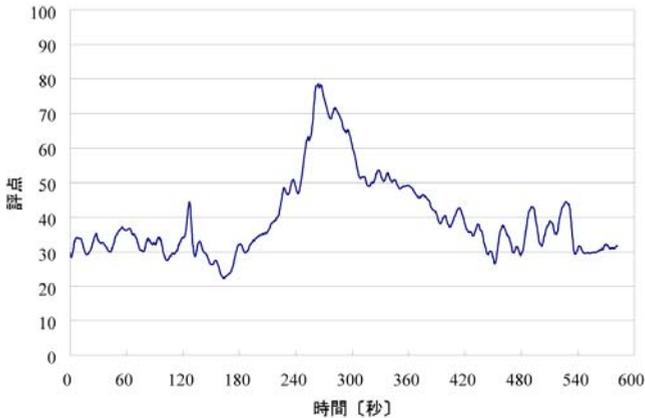


図 1・9 SSCQE 法による評価結果の例

1-3-5 SAMVIQ 法

SAMVIQ 法は、ITU-R 勧告 BT.1788 に規定されている 10)。BT.500 が主にテレビジョン評価を前提としていたのに対し、BT.1788 では、PC や携帯端末上でのマルチメディア視聴環境を想定して、映像品質評価法を勧告している。

SAMVIQ 法では、図 1・10 に示すように、PC 上で映像ファイルを再生して主観評価実験・データ集計が可能となる。DSCQS 法に比べて、評価時間や準備作業負荷が小さくて済み、かつ、同等の評価精度が得られるとされる¹²⁾。

SAMVIQ 法の特徴としては、次のようなものがあげられる。

- ・ 評価したい処理映像に加え、原映像も評価に加える。被験者には、原画と分かっているもの (explicit reference) に加え、分からないもの (hidden reference) も処理映像に交えて呈示する。これによって、DSCQS 法と同等の評価が可能となる。
- ・ 被験者は、何度でも映像を自分で再生することができる。各映像とも、1 回目の再生時には、途中で止めることはできないが、2 回目以降は、再生・中止は自由なタイミングで可能である。0 から 100 までの評点を各映像につける。
- ・ explicit reference 以外は、どの処理映像がどの再生ボタンに割り当てられているかは、被験者には分からない (テスト映像ごとにランダムになっている)。例えば図 1・10 では、REF とあるのが explicit reference、A から G に処理映像と hidden reference が順不同で割り当てられている。

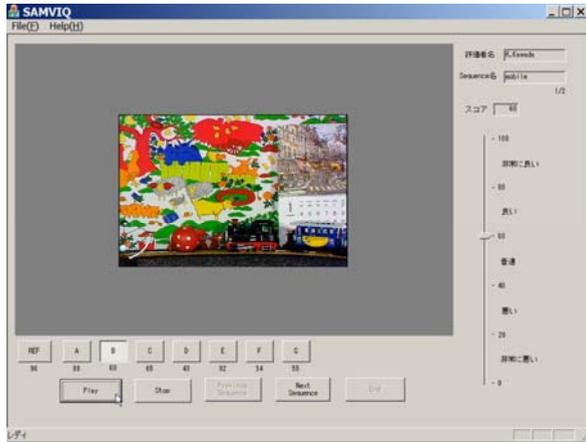


図 1・10 SAMVIQ 法のツール画面例

■参考文献

- 1) 阿部, “撮像デバイスの最近の動向,” 映情メ誌, vol.59, no.3, pp.348-349, 2005.
- 2) 高橋, “CMOS イメージセンサの低ノイズ化動向,” 映情メ誌, vol.62, no.3, pp.303-306, 2008.
- 3) 川田, 浜田, 松本, “動き補正テレビ方式変換の改善,” 映情メ誌, vol.51, no.9, pp.1577-1586, 1997.
- 4) 野尻, “動き補正型方式変換画像の変換劣化に対する客観評価量の検討,” テレビ誌, vol.50, no.8, pp.1080-1090, 1996.
- 5) 栗田, “薄型テレビにおける高画質化の技術動向,” 映情メ誌, vol.61, no.9, pp.1267-1271, 2007.
- 6) 吉田, 藤根, 山本, 古川, 上野, 菊池, 小橋川, 山田, 竹田, 杉野, “大型 LCD-TV 用倍速フレームレート変換技術とその画質改善効果の検証,” 映情メ誌, vol.62, no.5, pp.778-787, 2008.
- 7) “Test pictures and sequences for subjective assessments of digital codecs conveying signals produced according to Recommendation ITU-R BT.601,” ITU-R Recommendation BT.802-1, 1994.
- 8) “映像情報メディア学会テストチャート,” <http://www.ite.or.jp/shuppan/testchart/index.html>.
- 9) “Methodology for the subjective assessment of the quality of television pictures,” ITU-R Recommendation BT.500-12, 2009.
- 10) “Methodology for the subjective assessment of video quality in multimedia applications,” ITU-R Recommendation BT.1788, 2007.
- 11) “User requirements for the transmission through contribution and primary distribution networks of digital television signals defined according to the 4:2:2 standard of Recommendation ITU-R BT.601 (Part A),” ITU-R Recommendation BT.800-2, 1995.
- 12) (社) 電波産業会, “SAMVIQ 法に関する実験報告書,” 2006.

■3群 - 5編 - 1章

1-4 映像の客観品質評価

(執筆者：林 孝典) [2011年1月 受領]

1-4-1 客観品質評価の必要性

1-3節で述べたように、映像通信サービスの QoE (Quality of Experience) 評価は、人間が知覚する映像メディアの品質を視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、主観品質評価試験の実施には多大な労力と時間、専用の評価設備などが必要となる。このため、映像通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、映像メディア信号やサービスにかかわる物理的な特徴量から QoE を推定する客観品質評価技術の確立が必要となる。特に、サービス提供中に品質を把握する“インサービス品質管理”を実現するためには、QoE をリアルタイムかつ長時間把握することが必要となり、通信中に測定可能な物理量から QoE を推定する客観品質評価技術は核となる。近年、この分野の技術進展は目覚ましく、国際的には ITU (International Telecommunication Union) が客観評価技術の議論をリードしている。

1-4-2 ピーク信号対雑音比

映像品質の基本的な客観評価尺度の一つとして、ピーク信号対雑音比 $PSNR$ (Peak Signal to Noise Ratio) があり、式(1.4)、式(1.5)で表される。

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \quad (1.4)$$

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N [V_{in}(i, j) - V_{out}(i, j)]^2 \quad (1.5)$$

ここで、 $V_{in}(i, j)$ は評価の基準となる映像(基準映像)の画素、 $V_{out}(i, j)$ は評価対象の映像(評価映像)の画素、 $M \times N$ は画像サイズ、 MSE (Mean Square Error) は基準映像と評価映像の画素ごとの差分自乗平均を表す。映像符号化方式などの検討において、映像品質を表す指標として広く用いられている。また、動画の場合は、フレームごとに計算した $PSNR$ の平均値で映像品質を表現する。

しかしながら、デジタル映像の圧縮・伝送技術の高度化に伴い、画質劣化[†]が多種多様となり、また、コンテンツの空間的・時間的な特徴によって画質劣化の見え方が大きく異なるため、QoEを表す尺度として $PSNR$ を用いることは不適切となった。

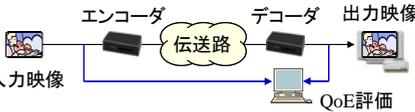
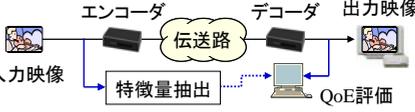
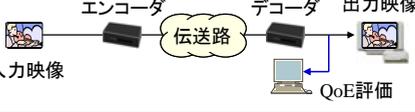
各種画質劣化の特徴量をとらえるパラメータとしては、ANSI (American National Standards Institute) 規格 T1.801.03²⁾が提案されており、市販の映像品質測定機器においてこれらのパラメータを出力するものがあるが、 $PSNR$ と同様、QoE を推測するための参考尺度として利用されるべきである。

[†] デジタル映像の圧縮・伝送により発生する画質劣化には、ブロックひずみ、ぼけ、エッジビジネス、モスキートノイズ、ジャーネスなど様々である。各種画質劣化の定義については、ITU-T 勧告 P.930¹⁾を参照されたい。

1-4-3 客観品質評価法の分類

PSNRに代わり、デジタル映像に対するQoEを客観的に評価する手法を構築するため、ITU-Tでは2000年5月、映像メディア信号を用いた客観品質評価の三つの枠組みをITU-T勧告J.143³⁾としてまとめた。表1・6に示すように、客観評価法を、①基準映像と評価映像を直接比較して評価するFull Reference法(FR法)、②基準映像から抽出した特徴量と評価映像を使用して評価するReduced Reference法(RR法)、③評価映像のみを使用して評価するNo Reference法(NR法)の三つに分類している。

表1・6 映像の客観品質評価の枠組み

	評価方法	利用シーン
二重刺激	Full Reference法(FR法) 	<ul style="list-style-type: none"> ・ 実験室ベンチマーク ・ 配信素材品質確認 ・ 符号化装置評価 ・ 放送波ループバック評価
	Reduced Reference法(RR法) 	<ul style="list-style-type: none"> ・ 遠隔地間のベンチマーク ・ ネットワーク品質評価 ・ 品質常時監視
単一刺激	No Reference法(NR法) 	<ul style="list-style-type: none"> ・ 品質常時監視 ・ 受信品質評価

最初に標準化された映像品質客観評価技術は、SDTV (Standard Definition TeleVision) 映像のMPEG-2符号化ひずみを評価可能なFR法であり、ITU-T勧告J.144⁴⁾として制定された。本勧告はTV放送サービスの映像品質評価を想定していたが、近年、通信回線を介した映像配信サービスの普及により、当該サービスで用いられる多様な符号化方式・ビットレートやIPネットワーク特有の劣化であるパケット損失による劣化映像を評価可能な客観品質評価法の構築が強く要望されてきた。そこで、VQEGではPCや携帯端末(画像解像度:QCIF, CIF, VGA)を想定した映像配信サービスを対象とした客観品質評価法の検討を進め、2008年8月にITU-T勧告J.246⁵⁾(RR法)、J247⁶⁾(FR法)として国際標準化された。その他の品質評価技術の最新標準化動向については、1-5節を参照されたい。また、代表的な検討例を文献7)、8)に挙げる。

1-4-4 ITU-T 勧告 J.144 で規定されている客観品質評価法の一例

本節では、ITU-T勧告J.144 Annex Aで規定されている客観品質評価方式の一つである英BT (British Telecom)方式を示す。図1・11に示すように、本方式ではまず「検出部」において基準信号と評価映像信号の時空間的なズレを整合(マッチング)するとともに、符号化品

質劣化量を表現する六つのパラメータを算出する。具体的には、PSNRのほか、映像の空間周波数分布やエッジの差分など、人間の視覚周波数応答やマスクング効果などを加味した以下の品質推定パラメータを算出する：

- ① fXPerCent：基準映像と評価映像の輝度信号を水平方向に-4～+4画素シフトさせて9×9画素ブロック単位で輝度信号の差分（誤差）を求めた結果から、各シフト量において誤差が最小となったブロック数の最大値が総ブロック数に占める割合を算出。
- ② MPSNR：基準映像と評価映像の輝度信号を9×9画素ブロックごとにマッチングし、その差分データからPSNRを算出。
- ③ PySNR(3,3)：映像信号を空間周波数領域で分析するパラメータとして、マッチング後の基準映像と評価映像の輝度信号のそれぞれを3階層ピラミッド変換してPSNRを算出。
- ④ TextureDeg：映像の複雑さを表現するパラメータとして、評価映像の輝度信号において水平方向の画素差分値の符合が変化した割合を算出。
- ⑤ EDif：エッジ部分の劣化をとらえるパラメータとして、基準映像と評価映像の輝度信号を用いて作成したエッジ抽出画像から差分値を算出。
- ⑥ SegVPSNR：色合いの変化をとらえるパラメータとして、②と同様の方法でマッチングした色差信号（V成分）からPSNRを算出。

次に「統合部」では、上記六つの品質推定パラメータの時間平均値を導出し、その結果を積み付け加算して映像品質客観評価値を計算する。アルゴリズムの詳細はITU-T勧告J.144を参照されたい。

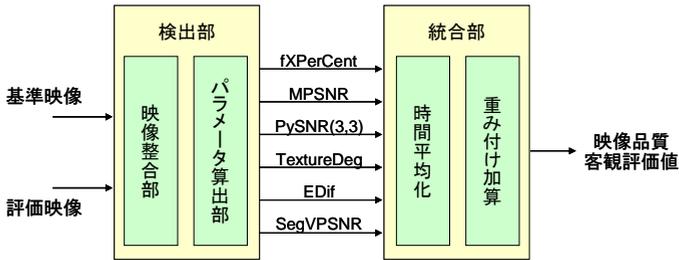


図1-11 ITU-T勧告J.144で規定されている客観品質評価モデルの例

1-4-5 マルチメディア品質の客観評価

従来のQoE評価技術は、主に音声・映像の個別メディアに対する品質を対象とした検討が中心であるが、最終的には、各メディア間の相互作用や同期の影響などを考慮して、総合的なQoEを客観評価できる仕組みを構築することが必要である。ITU-Tでは2003年5月、マルチメディア品質の客観評価法のフレームワークを勧告J.148⁹⁾としてまとめている(図1-12)。音声・映像の個別メディアのQoEを入力とし、マルチメディア品質(総合品質)のほか、映像品質の影響を受けた音声品質あるいはその逆の品質などを出力する概念が示されている。また、どのようなコンテンツを想定して評価するのかにより、マルチメディア品質の統合方法が変わる可能性があるため、タスクも入力情報として与えられている。

ITU-T勧告P.911¹⁰⁾では、映像配信サービスを対象として、音声メディアの品質(平均オピ

ニオン評点)を MOS_A 、映像メディアの品質を MOS_V とした場合、これを統合したマルチメディア品質 MOS_{AV} を式(1・6)で導出している。

$$MOS_{AV} = \alpha + \beta MOS_A MOS_V \quad (1 \cdot 6)$$

ここで、 α 、 β は定数を表す。ほかの検討例¹¹⁾では、各メディア品質がマルチメディア品質に及ぼす影響度の違いを考慮できるように、式(1・7)を適用する場合が多く見られる。

$$MOS_{AV} = \alpha + \beta MOS_A + \gamma MOS_V + \delta MOS_A MOS_V \quad (1 \cdot 7)$$

ここで、 α 、 β 、 γ 、 δ は定数である。テレビ電話サービスの総合品質評価法を規定する ITU-T 勧告 G.1070¹²⁾の中でも、同様の考え方が適用されている。

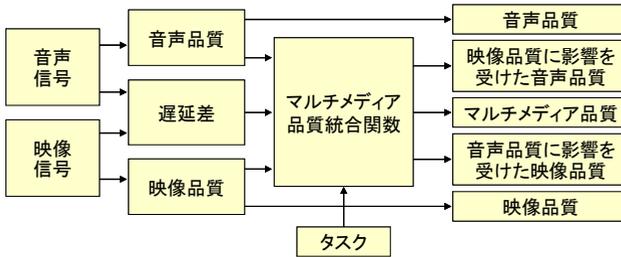


図 1・12 マルチメディア品質客観評価のフレームワーク

■参考文献

- 1) “Principles of a reference impairment system for video,” ITU-T Recommendation P.930, Oct. 1998.
- 2) “Digital transport of one-way video signals parameters for objective performance assessment,” ANSI T1.801.03, Feb. 1996.
- 3) “User requirements for objective perceptual video quality measurements in digital cable television,” ITU-T Recommendation J.143, May 2000.
- 4) “Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference,” ITU-T Recommendation J.144, March 2004.
- 5) “Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference,” ITU-T Recommendation J.246, Aug. 2008.
- 6) “Objective perceptual multimedia video quality measurement in the presence of a full reference,” ITU-T Recommendation J.247, Aug. 2008.
- 7) M. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” IEEE Transactions on Broadcasting, vol.50, no.3, pp.312-322, Sept. 2004.
- 8) P.L. Callet, C. Viard-Gaudin, and D. Barba, “A convolutional neural network approach for objective video quality assessment,” IEEE Transactions on Neural Networks, vol.17, no.5, pp.1316-1327, Sept. 2006.
- 9) “Requirements for an objective perceptual multimedia quality model,” ITU-T Recommendation J.148, May 2003.
- 10) “Subjective audiovisual quality assessment methods for multimedia applications,” ITU-T Recommendation P.911, Dec. 1998.
- 11) D.S. Hands, “A basic multimedia quality model,” IEEE Transactions on Multimedia, vol.6, no.6, pp.806-816, Dec. 2004.
- 12) “Opinion model for video-telephony applications,” ITU-T Recommendation G.1070, April, 2007.

■3群 - 5編 - 1章

1-5 標準化の動向

(執筆著：高橋 玲) [2011年1月 受領]

QoE に関する国際標準化はITU (International Telecommunication Union) を中心に行われている。ITU-T (ITU - Telecommunication Sector) は通信にかかわる研究委員会の組織であり、SG12 (Study Group 12) が性能と QoS (Quality of Service) に関するリード SG として位置づけられている。SG12 は、メディア品質評価法に関する P シリーズ勧告 (P.900 シリーズを除く) とネットワーク伝送計画に関する G.100/G.1000 シリーズ勧告を所掌している。SG9 はケーブルテレビジョンに関するリード SG であり、その活動にはケーブルテレビジョン品質に関する検討課題を含む (P.900 シリーズ及び J シリーズ勧告を所掌)。一方、放送に関する品質は ITU-R (ITU - Radiocommunication Sector) の SG6 が検討しており、BT シリーズ及び BS シリーズの品質関連勧告が所掌に含まれる。また、映像品質評価法に関しては ITU-T と ITU-R に横断的な映像品質専門家会合 (VQEG: Video Quality Experts Group) があり、各 SG の技術諮問機関として機能している。本節では、これら機関における音声及び映像の主観・客観品質評価法の標準化動向について説明する。

1-5-1 主観品質評価法

ITU において標準化されている主観品質評価法を表 1・7 に示す¹⁾⁷⁾。ここでは評価対象とするメディアを、音声、オーディオ、映像、マルチメディアの四つに分類している。また、評価基準を用いない「絶対評価」と、品質の比較対照となる基準を提示する「相対評価」それぞれについて、評価カテゴリーによる離散的な評価 (カテゴリー尺度) と連続尺度を分けて整理している。これらの評価法は国際的にも定着した評価法であり、単に評価尺度を規定するだけでなく、実験機器の特性、被験者の選定、評価環境 (防音室など) の特性、被験者へのインストラクション、評価手順など、評価の普遍性と再現性を確保する観点から様々なガイドラインが定められている。現在でも、新たなサービス・方式の評価への対応を目的に、新たな評価法の標準化が検討されており、評価法の選択に際しては、評価の目的や評価対象とする品質要因 (符号化ひずみ、パケット損失、遅延時間など) を考慮する必要がある。また、標準化された評価法を用いた具体的な実験計画については、VQEG や音声品質専門家会合 (SQEG: Speech Quality Experts Group, SG12 内の課題 7 に該当) が実施している国際的な主観品質評価試験計画が参考になる⁸⁾。

表 1・7 主観品質評価法の技術分類と国際標準

	絶対評価		相対評価	
	カテゴリー尺度	連続尺度	カテゴリー尺度	連続尺度
音声	P.800(ACR法)	BT500(SSCQE法)	P.800(DCR法)	
オーディオ	BS.1284(ACR法)		P.800(CCR法)	
			BS1284(DCR法)	
		BS1284(CCR法)		
映像	P.910(ACR法)	BT500(SS法)	BS1116(DBTS-HR法)	BT500(DSCQS法)
	BT500(SS法)		BT500(DCR法)	BT500(SCS法)
			P.910(DCR法)	BT500(SDSCFE法)
マルチメディア	P.911(ACR法) P.920(ACR法)※	P.911(SSCQE法)	P.911(DCR法)	

※P.911は片方向(受視聴)評価, P.920は双方向(会話)評価

1-5-2 客観品質評価法

客観品質評価法は、物理的な特微量分析に基づいて主観品質を推定する方法である。客観品質評価は、多大な労力・時間と専用の評価設備を必要とする主観品質評価を行うことなく、主観品質を評価することができ、品質評価の効率化の観点から非常に重要である。客観品質評価技術はその適用領域とそれを考慮した技術アプローチから表 1・8 のように分類できる。以下、各アプローチについて標準化の動向を紹介する。

表 1・8 客観品質評価法の分類

	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	パラメトリックプランニングモデル	ビットストリームレイヤモデル
入力情報	メディア信号	パケットヘッダ情報	品質設計・管理パラメータ	ペイロード情報(復号化前)
主な評価目的	コーデック最適化 サービス実力把握	インサービス品質管理	品質設計 インサービス品質管理	インサービス品質管理

(1) パラメトリックプランニングモデル

パラメトリックプランニングモデルは品質設計パラメータを入力としてメディア品質を推定する技術であり、その研究は電話サービスの品質設計を、主観品質評価特性を反映しつつ効率的に行うことを目的に始まった。「品質設計」という性格上、机上で評価可能な品質パラメータに基づいて、それにより実現される QoE を評価することが必要である。当時の品質設計では伝送路における信号の減衰や遅延、回線エコーといった品質要因が支配的であった。伝送路の IP 化により「信号の減衰」はなくなったが、遅延、エコーといった従来の要因に加えて符号化ひずみやパケット損失ひずみを加味することが重要となった。

このような背景のもと、ITU-T は 1998 年に E-model を ITU-T 勧告 G.107⁹⁾として標準化した。E-model には、端末要因・環境要因・ネットワーク要因等に関する 21 の入力パラメータがあり、出力評価値である R 値はこれらのパラメータの関数として表現される¹⁰⁾。具体的には、ネットワーク設計者は、端末やネットワークの設計パラメータを変化させたときの総合通話品質を E-model により指標化し、これに基づいてネットワークを最適化する。

日本国内では、郵政省令（事業用電気通信設備規則）の改正に伴い、IP 電話サービスに対する通話品質基準として、E-model に基づいて算出される評価指標である R 値が用いられた。これに対応し、TTC（情報通信技術委員会）は E-model をベースとした具体的な通話品質評価法を標準化した¹¹⁾。

一方、TV 電話サービスを対象としたパラメトリックモデルの標準化も検討され、ITU-T は 2007 年 3 月にこの技術を ITU-T 勧告 G.1070（Opinion Model for Video-telephony applications）として標準化した¹²⁾。更に ITU-T では、G.1070 を発展させ、映像配信サービスを対象としたパラメトリックモデル（GOMVAS: Opinion Model for Video and Audio Streaming Applications）の標準化も目指している（2010 年標準化予定）。

(2) メディアレイヤモデル

メディアレイヤ客観品質評価技術は、原信号と劣化信号を比較することで主観品質を推定する Full-reference 型モデル、劣化信号のみを用いる No-reference 型モデル、原信号の特微量に関するサイド情報と劣化信号を用いる Reduced-reference 型モデルに分類される。

音声メディアに関しては、ITU-T SG12 において標準化が進められており、Full-reference 型モデルとして ITU-T 勧告 P.862 “PESQ (Perceptual Evaluation of Speech Quality)”¹³⁾、及び広帯域 (7 kHz) 音声を対象とした拡張方式である ITU-T 勧告 P.862.2 “Wideband PESQ”¹⁴⁾ が標準化されており広く普及している。また、No-reference 型モデルとして ITU-T 勧告 P.563¹⁵⁾ を標準化している。これらの勧告は電気インタフェースを介して音声信号を取得することを前提にしているが、現在、音響インタフェースを介することで端末の電気音響特性も加味した評価を可能にする新たな標準の策定に取り組んでいる (POLQA)。

オーディオメディアに関しては、ITU-R WP 6Q において標準化が行われており、これまでに Full-reference 型モデルである ITU-R 勧告 BS.1387 “PEAQ (Objective Measurements of Perceived Audio Quality)” を標準化している¹⁶⁾。本勧告は主に符号化による品質劣化の評価を対象としており、IP 伝送において問題となるパケット損失ひずみなどの劣化の評価には適用できないという問題が残る。

映像メディアに関しては、VQEG が中心になって技術的な検討を行っている。現在勧告化が完了しているのはテレビジョン映像を対象とした Full-reference 型モデルである ITU-T 勧告 J.144¹⁷⁾及び ITU-R 勧告 BT.1683¹⁸⁾、QCIF～VGA 映像を対象とした Reduced-reference 及び Full-reference 型モデルである ITU-T 勧告 J.247, J.246 である。現在、VQEG では、SDTV 映像を対象とした No-/Reduced-reference 型モデル、更に HDTV 映像を対象としたモデルの検討も行っている (2010 年標準化見込み)。

(3) パラメトリックパケットレイヤモデル

サービス提供中の品質を、サービスを妨害することなく監視・管理することは、サービス品質を維持するうえで極めて重要である。このようなインサービス品質管理においては、処理負荷の観点から、音声・映像メディア (つまり復号化された) 信号を用いることは困難である。そこで、パケットのヘッダ情報のみから品質を推定する技術が望まれる。ITU-T SG12 では、IP 電話サービスを対象としてパケットレイヤ客観品質評価技術の標準化に取り組み、ITU-T 勧告 P.564¹⁹⁾を勧告化した。

ITU-T 勧告 P.564 は特定の方式を記述した勧告ではなく、「客観品質評価技術が満たすべき性能要求条件」を規定した「フレームワーク勧告」である。つまり、この要求条件を満足する客観品質評価技術はすべて「P.564 準拠」として扱われる。

ITU-T SG12 では、今後、IPTV を含めた映像配信サービスを対象として同様の技術を標準化することとしており、この勧告は暫定的に P.NAMS (Non-intrusive parametric model for the assessment of performance of multimedia streaming) と呼ばれている。

(4) ビットストリームレイヤモデル

パケットレイヤモデルでは、ペイロードの内容までは分析に用いないため、コンテンツの違いが主観品質に与える影響を評価することは困難である。例えば、同じビットレートで符号化された映像であっても、信号が複雑であればあるほど (空間的/時間的情報量が多いほど) 符号化による品質は低下する。しかし、ペイロードの信号を完全に復号化して特徴量を分析する場合 (つまりメディアレイヤモデル)、処理負荷が大きくなり、例えば端末における品質情報収集という観点では問題が生ずる。そこでペイロードに含まれる (復号化前の) 符

号化ビット列を分析することによりコンテンツの違いが品質に与える影響も考慮した品質推定を試みるのがビットストリームレイヤモデルである。ITU-T SG12 は現在、標準化コードネーム P.NBAMS (Non-intrusive bit-stream model for the assessment of performance of multimedia streaming) として映像配信サービスを想定した方式検討を開始している。

■参考文献

- 1) “Methods for subjective determination of transmission quality,” ITU-T Recommendation P.800, Aug.1996.
- 2) “General methods for the subjective assessment of sound quality,” ITU-R Recommendation BS.1284, Dec. 2003.
- 3) “Subjective video quality assessment methods for multimedia applications,” ITU-T Recommendation P.910, Sep. 1999.
- 4) “Subjective audiovisual quality assessment methods for multimedia applications,” ITU-T Recommendation P.911, Dec. 1998.
- 5) “Interactive test methods for audiovisual communications,” ITU-T Recommendation P.920, May 2000.
- 6) “Methodology for the subjective assessment of the quality of television pictures,” ITU-R Recommendation BT.500-11, June 2002.
- 7) “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” ITU-R Recommendation BS.1116, Oct. 1997.
- 8) e.g., <http://www.its.bldrdoc.gov/vqeg/projects/multimedia/>
- 9) “The E-model, a computational model for use in transmission planning,” ITU-T Recommendation G.107, July 2002.
- 10) S. Möller, “Assessment and Prediction of Speech Quality in Telecommunications,” Kluwer Academic Publishers.
- 11) “IP 電話の通話品質評価法,” TTC 標準 JJ-201.01, 2003 年 4 月.
- 12) “Opinion model for video-telephony applications,” ITU-T Recommendation G.1070, Apr. 2007.
- 13) “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” ITU-T Recommendation P.862, Feb. 2001.
- 14) “Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs,” ITU-T Recommendation P.862.2, Nov. 2005.
- 15) “Single ended method for objective speech quality assessment in narrow-band telephony applications,” ITU-T Recommendation P.563, May 2004.
- 16) “Method for objective measurements of perceived audio quality,” ITU-R Recommendation BS.1387-1, Nov. 2001.
- 17) “Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference,” ITU-T Recommendation J.144, Mar. 2004.
- 18) “Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference,” ITU-R Recommendation BT.1683, June 2004.
- 19) “Conformance testing for narrowband voice over IP transmission quality assessment models,” ITU-T Recommendation P.564, July 2006.