

# 発展するVoIPの客観的音声品質評価 MOS値に近づく最新アルゴリズム「PESQ」

興和 開発本部 先端研究所

取締役所長 大林康二 研究課長 柏木一浩 研究課員 菊川泰士

電話の音声品質を評価する方法として、多くの被験者が実際に耳で聞いた評価データを統計するMOSに代わり、一定のアルゴリズムで客観的に品質評価を行う仕組みが登場している。VoIP市場が拡大する中で、こうした効率的な品質測定方法も重要度を増すと思われる。そこで、VoIPアナライザー製品を足がかりに音声品質評価ビジネスにも乗り出している興和の大林康二氏、柏木一浩氏、菊川泰士氏に、標準化されている品質評価法について解説していただく。

音声品質を決める要素としては、遅延、エコー、クラリティー（明瞭度）がある。ゆっくりした遅延は、大きくても明瞭度には影響しない。エコーは、主として話し手側に影響する。明瞭度には、符号化方法やネットワーク環境が影響する。この解説では、明瞭度に絞って、提唱されているさまざまな品質評価法をまとめる。

## 1. 主観的評価と客観的評価

人間が音響や音声の品質のよさを心理的判断に基づいて行う評価を「主観的品質評価(Subjective Assessment)」という。主観的評価は、音声(Voice & Speech)と音響(オーディオ)の場合で判断基準に差がある。

音声の場合は、「人の声はこうあるべき」と経験上記憶にあるので、参照の音声なしに絶対品質尺度(Absolute Category Rating: ACR)で評価でき

る。しかし、個々人の能力に依存して品質評価が変わる可能性がある。この影響を少なくするためには、あらかじめ設定した標準となる系(Reference System)を用い、標準系を評価者に聞かせておいて、標準系と比較して相対値として品質を評価する比較品質尺度(Comparison Category Rating: CCR)や、制御された被測定系を用いて絶対品質評価が行われる。

音響の場合は、本来こうあるべきだという記憶は通常ないので、原音と比較して劣化の度合いを評価する。また、品質低下も、音声に比べわずかな変化でも音響価値が妨害されるので、妨害品質尺度(Degradation Category Rating: DCR)が用いられる。

主観的評価では、音声や音響を多くの聞き手に聞かせて品質に主観的な評点をさせ、これを統計的に処理して平均の評点を得るオピニオン評価法が採

られる。評点をMOS(Mean Opinion Score:平均オピニオンスコア)と呼ぶ。

しかし、標準系や制御された被測定系を用意し、多くの被験者に品質評価をさせるには、多大の経費、労力、時間を費やすことになる。この負担を軽減するために、主観的評価の要因となる物理的な特徴量を選び、これらを客観的に測定することにより、人間が評価したものと同等の値を推定する客観的品質評価法(Objective Assessment)が実際のフィールドでは用いられる。客観的品質評価のアルゴリズムは、できるだけMOSに近い評価が得られるようにさまざまな工夫がなされている。

表1 音声品質を左右する要因

課題	品質要因
端末機器の品質要因	符号化ひずみ 入力依存性 話者依存性 メディア依存性
ネットワークの品質要因	遅延 ジッター エコー 伝送誤り パケットの損失 時間クリッピング
総合品質	環境騒音 ラウドネス オピニオン

白雑音  
あらゆる周波数の成分をほぼ同量ずつ含んだ音を指す。ホワイトノイズともいう。高い周波数の音が弱くなるような音はピンクノイズという

表2 音声品質評価法に関する標準規格

信号	データレート	主観的評価	客観的評価
広帯域音響	約64-128kbps	ITU-R BS.1116	ITU-R BS.1387 (PEAQ)
狭帯域音声	約8-64kbps	ITU-T P.800	ITU-T P.861 (PSQM) ITU-T P.862 (PESQ)

音声や音響の品質を決める要因には、大きく分けて個々の端末機器の品質要因とネットワークの品質要因がある。また、これらが加算的あるいは互いに影響し合ってネットワークを介した音声や音響の総合品質を決める。通信系の品質要因を表1に示す。

音声やオーディオを評価する場合、現在ITUから勧告されているどの標準に準拠した方法を用いればよいかを表2に示す。

音声の主観的評価にはITU-TのP.800勧告(MOS)客観的評価にはITU-TのP.861勧告(PSQM)とP.862勧告(PESQ)がある。音響の主観的評価にはITU-RのBS.1116勧告があり、客観的評価にはITU-RのBS.1387勧告(PEAQ)がある。

## 2. MOS

ネットワークについてMOSの試験を行ういくつかの方法がITU-T P.800に記載されている。一部を記すと、MOSを求める以下のような方法がある。

- ・絶対品質尺度(ACR): MOS  
被測定系の品質を絶対品質で評価
- ・妨害品質尺度(DCR): D-MOS  
被測定系の品質を標準系からの品質の劣化量で評価
- ・比較品質尺度(CCR): C-MOS

被測定系の品質を標準信号との比較で評価

ACRに基づくMOSを表3に示す。ACRは劣化を受けていない参照音声との比較は行っていない。通常の電話では、相手の生の音声を聴きそれを参照信号として比較を行うことはできないわけだから、絶対品質尺度法は電話による通信方法になかったものである。

しかし、複数の異なった聞き手の結果を集計して、きちんとした比較を行うためには、すべての聞き手のオピニオンを1つの絶対的な尺度に調整する必要がある。この尺度として、音声に白雑音を付加して、その付加の強度を変えてさまざまな音声品質の標準系とする方法がある(MNRU: Modulated Noise Reference Unit、変調雑音発生標準装置)このテストから比較品質尺度法のスコアC-MOSが得られる。テストの前に、MOSの5段階の各グレードに対応するMNRU信号を聞かせて音声品質の基準を確かめさせるわけである。このテストの条件に従って、数十人の評価者に対して、一連の同一の断片的音声サンプルを聞かせる。評価者は、それぞれの音声サンプルに5段階の音声品質グレードをつける。各評価者の結果を統計処理して、MOSが計算される。きちんと手順を

表3 絶対品質尺度に基づくMOS値

スコア	品質
5	非常によい(excellent: E)
4	よい(good: G)
3	まあよい(fair: F)
2	悪い(poor: P)
1	非常に悪い(bad: B)

踏んで行えば、違う場所で行っても、このテストの再現性は極めてよい。

ただし、MOSを求めるためには、標準系や環境を整備するための設備、専門の試験員や多くの被験者が必要で、評価者側にも大変な努力が求められる、多大な経費や労力、時間を費やすこととなる。実際に使用されている電話網の品質を日常的に調べるためには、客観的品質評価法のほうが有用である。

## 3. 客観的品質評価の標準化

アナログである音声を効率よく伝送するためには、デジタルに変換して符号化するコーデックを用いる。コーデックの客観的品質評価には、符号化ひずみを客観的に測定し、主観的評価であるMOSの値と相関のよい評点(スコア)を出力するひずみ尺度が必要である。そこで、人間が音を知覚し、判断を入れて認知する仕方に極力類似させて客観的なモデルを作る必要がある。

PSQM(Perceptual Speech Quality Measurement: 知覚的音声品質評価法)は、人間の音声知覚の仕方を考慮した客観的で、自動的で、再現性のある音声品質評価法の要求に応えるた