

低ビットレート符号化音の客観音質測定法 ～ MPEG-2 AAC AM 相等等符号化音の客観評価 ～

渡辺 馨^{*1} 窪田弘美^{*2} 松島幸雄^{*3} 鹿又健一^{*4} 濱田泰光^{*5}

^{*1}NHK 放送技術研究所 〒157-8510 東京都世田谷区砧 1-10-11

^{*2}(株)ジェニスタ ^{*3}(株)ケンウッド ^{*4}(株)フジテレビジョン ^{*5}東京電子工業(株)

E-mail watanabe.k-de@nhk.or.jp

あらまし 電波産業会(ARIB)・品質評価法調査研究会では、オーディオ信号の低ビットレート符号化方式におけるAM～FM程度の符号化音質を客観的に測定する、客観音質測定法を調査した。調査した客観音質測定法は、ITU-Rで勧告されているPEAQ客観音質測定法と、PEAQ法を改良したジェニスタ社客観音質測定法である。ARIBが1998～99年に行った「MPEG-2 AAC低ビットレート音声主観評価結果」とこれらの客観音質測定法の測定結果との対応を求めた結果、音源カテゴリーを「スピーチ、スピーチ+背景音」に限定すれば0.9強の高い相関をもつことが分かった。また、「音楽」カテゴリーにおいてはジェニスタ社測定法がPEAQ測定法より優れていると考えられる。

キーワード 客観音質測定、MPEG-2 AAC、主観音質、オーディオ符号化

A Study of Objective Measurement Method of Intermediate Audio Quality with Low Bit-rate Audio Coding

Kaoru Watanabe^{*1} Hiromi Kubota^{*2} Yukio Matsushima^{*3} Ken-ichi Shikamata^{*4} Yasumitsu Hamada^{*5}

^{*1}NHK Science and Technical Research Labs. 1-10-11, Kinuta, Setagaya-ku, Tokyo 157-8510, Japan

^{*2}Genista Corp. ^{*3}Kenwood Corp. ^{*4}Fuji Television Company ^{*5}Tokyo Electric Industry Corp.

E-mail watanabe.k-de@nhk.or.jp

Abstract This document presents a validity of objective intermediate audio quality measurement method. Two objective measurement methods, ITU-R PEAQ method and Genista measurement method, have been investigated. The validity is studied between the objective measured values and the subjective quality assessment results of MPEG-2 AAC codec with from 32kbps/2ch to 80kbps/2ch carried by ARIB. As a result, for speech materials both objective measurement methods have high correlation with subjective qualities, and for music materials Genista method has better performance than PEAQ method.

Keyword Objective Audio Quality Measurement, MPEG-2 AAC, Subjective Quality, Low Bit-rate Audio Coding

1. はじめに

人の聴覚的な特性を利用したオーディオ信号の低ビットレート符号化方式が、様々な分野で利用されている^[1]。この符号化音質は、符号化方式、ビットレート、音源などにより大きく変動する。これらの符号化音質の評価には、主観評価法と客観評価法が用いられる。現在のところ、最終的な音質評価や確認は主観評価により行われている。しかし、主観評価は多大な時間や労力を要することから、主観的な音質に対応して客観的に音質が測定できる客観音質測定法が望まれている。

BSデジタル放送等で用いられる、比較的ビットレートが高くFM～CD程度の高音質な符号化音については、ITU-RにおいてPEAQ(Perceived Evaluation of Audio Quality)客観音質測定法^[2]が規格化されている。また、電話音質についてはITU-TでPESQ(Perceived Evaluation of Speech Quality)客観測定法^[3]が規格化されている。一方、AM～FM程度の中間音質については、ITU-Rにおいて主観評価法^{[4][5]}が規格化されている。しかし、これらの中間音質の主観的な音質に対応した客観測定法については未だ規格化されていない。中間音質の客観測定は、デジタル放送の携帯受信やインターネット放送、

IMT-2000やインターネットでの音楽配信などの応用を考えれば、重要な課題である。

上記への応用を考慮すれば、スピーチに加えて音楽信号の音質を客観測定できることが望ましい。そこで今回まず、高音質向きのPEAQ客観音質測定法とPEAQ法を改良したジェニスタ社客観音質測定法における中間音質の測定の可能性を、電波産業会 (ARIB) が 1998～99 年に行った「ARIB 低ビット音声主観評価 (MPEG-2 AAC)」^[6] のデータを用いて調べた。また AAC 以外の符号化方式についても、第一歩として「SBR, MP3, PAC 符号化音」の簡易な主観評価を行い、客観音質測定の可能性を検討した。

以下、2章で客観音質測定法の概要、3章で「ARIB 低ビット音声主観評価 (MPEG-2 AAC)」と客観音質測定との対応について述べる。また、4章では AAC 以外の方式の可能性についてふれる。

2. 客観音質測定法

従来、音質の客観的な測定法として SN 比 (Signal to Noise Ratio) や全高調波歪 (THD; Total Harmonic Distortion) が使用されている。しかし聴覚的な特性を利用した低ビットレート・オーディオ符号化装置の音質を測定するためには、SN 比や全高調波歪などは適切でない。そこで ITU-R ではこれらの符号化音の主観音質を客観的に測定できる PEAQ 客観音質測定法を規格化した。この方式は、ITU-R 勧告 BS.1116-1「劣化の少ないオーディオシステムの主観的な評価法」^[7] に則って行った主観評価結果に対応した音質が測定できる^[8]。一方、ジェニスタ社客観音質測定法は、PEAQ 客観音質測定法を基にしてジェニスタ社が独自に幾つかの処理部を改良し、性能の向上を図ったものである。今回、中間音質での PEAQ 客観音質測定法とジェニスタ社客観音質測定法の測定の可能性を調査した。以下、各測定法の概要を述べる。

2.1 PEAQ 客観音質測定法の概要

PEAQ 客観音質測定法の測定原理図を図 1 に示す。この測定法は、原音と被測定音 (例えば AAC オーディオ符号化装置の復号音) の 2 信号を用いて、被測定音の音質を測定する。この測定法には、リアルタイム向きの簡潔なバージョン (Basic Version) と、精度の高い出力が得られるバージョン (Advanced Version) があるが、基本原理は同じである。測定原理は以下の通りである。

- (i) 原音と被測定音を、聴覚系を模擬したアルゴリズム (聴覚モデルと呼ばれる) に入力し、各々の出力を求める。Basic Version の聴覚モデル出力の計算方法を後述する。
- (ii) 原音と被測定音の聴覚モデル出力値から、人が聞いた時の聴覚的な信号劣化の程度 (聴覚歪特性) を表す中間的なモデル出力値を求める。Basic Version は 11 個、Advanced Version は 5 個のモデル出力値をもつ。
- (iii) ニューラルネットワーク構造を持つ認識モデルを用いて、モデル出力値から被測定音の客観音質を計算する。

聴覚モデルは外耳、中耳、内耳の機能を模擬したアルゴリズムで、Basic Version の聴覚モデルは 2048 点の FFT を用いて計算する。Basic Version の聴覚モデルを図 2 に示す。聴覚モデルのアルゴリズムは以下の通りである。

- (i) FFT より 23.4Hz (48kHz サンプルング時) 毎の周波数成分を計算し、これに外耳や中耳の周波数特性を乗算し、重みづけを行う。
- (ii) 内耳の周波数分析機能に相当する 1/4 臨界帯域幅の単位で、周波数成分をグループ化する。
- (iii) 周波数成分に血流などに相当する生理的雑音を加算する。
- (iv) 同時マスクングの周波数軸上の広がりや前向き時間マスクングの時間的な広がりを考慮して、神経興奮パターンを計算する。この結果を聴覚モデルの出力とする。

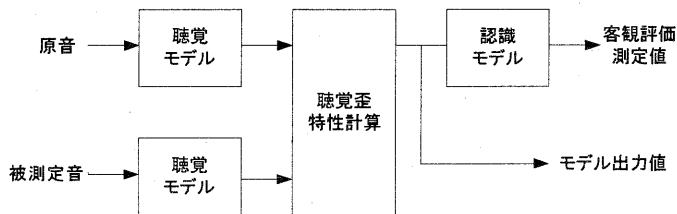


図 1 PEAQ客観音質測定法の原理

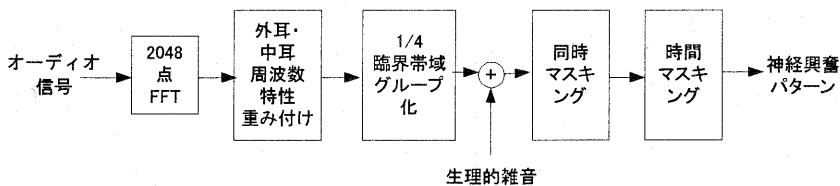


図 2 PEAQ Basic Versionの聴覚モデル

2.2 ジェニスタ社客観音質測定法の概要

ジェニスタ社客観音質測定法は、PEAQ客観音質測定法を基に幾つかの処理部を改良し、性能の向上を図ったものである。改良、検討した箇所は以下の通りである。

(i) 聴覚モデルの最適化

聴覚モデルの各計算部の見直し。一番大きな変更は、同時マスキング効果の数量化を行う周波数軸上の広がり関数の変更である。

(ii) 中間的なモデルの最適化

主観評価データと中間的なモデルの因果関係を調べ、後段のニューラルネットワークを用いた認識モデルの入力値として有効な値の範囲に収まるよう、一部モデルの係数を変更した。

(iii) ニューラルネットワーク構造の認識モデルの検討

(i),(ii)の変更により、ニューラルネットワークの重み係数の変更が必要となる場合がある。別途用意した主観評価データを用いて重み係数の変更を検討したが、変更による有意な差がなかったため、PEAQ測定法の重み係数をそのまま使用している。

3. ARIB低ビットレート音声主観評価結果に対する高音質向き客観音質測定法の適合性

3.1 ARIB低ビットレート音声主観評価

適合性調査に用いた「ARIB低ビットレート音声主観評価(MPEG-2 AAC)」について紹介する。この主観評価は、1998年から99年にARIBにおいて、AM放送が担っているサービス領域を意識したビットレートにおけるMPEG-2 AAC符号化⁹⁾の音質を確認するために実施された。

3.1.1 主観評価法と評価音

評価は2チャンネルステレオで行われた。評価テストに用いられた符号化パラメータを表1に示す。このうち、AACのサンプリング周波数は元の48kHzを一旦半分の24kHzに変換してAAC符号化・復号化を行い、再度48kHzにサンプリング周波数変換した後、主観評価を行っている。AAC符号化ビットレートはステレオ信号で32,48,64,80kbpsである。また、比較音源として、AMステレオを模擬した音(AM模擬)や電話の符号化(ITU-T G.711)も併せて評価している。AM模擬以外の符号化音はソフトウェアにより作成された音である。

評価音源は、極端にクリティカルでない一般的な音源(候補として23音源を準備)を3種のカテゴリー(音楽、スピーチ、スピーチ+背景音)に分け、各カテゴリーから万遍なく比較的クリティカルな計10音源(音楽:3、スピーチ:4、スピーチ+背景音:3)を選定した。選定された評価音源を表2に示す。

主観評価テストは、ITU-R勧告BS.1284¹⁰⁾のR-A R-A法に則って実施された。評定者への提示は、R/A/R/Aの順序で事前に録音したDATテープを用いた。ここで、「R」は基準音、「A」は符号化音(評価音)である。評定者は各試行毎に「R」(基準音)と「A」(評価音)を試聴し、そのうち「A」の「基本オーディオ品質」を評価した。各評定者は、図3に示すITU-Rの5段階品質尺度で、小数点一桁の精度で評点をつけた。統計

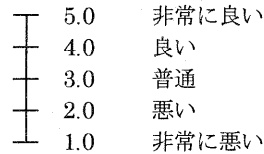


図3 ITU-R 5段階品質尺度

符号化方式	プロファイル	サンプリング周波数 (kHz)	ビットレート (kbps)
AAC	LC	24	32,48,64,80
比較用	AM 模擬	-	-
	電話品質 (ITU-T G.711)	8	64
基準音	-	48	リニア PCM 16bit/sample

表1 ARIB主観評価テスト符号化パラメータ

ファイル名	音源名	カテゴリー	出典元
lbr5.aif	クラシック #65	音楽	SQAM
lbr7.aif	ABBA #69	音楽	SQAM
lbr13.aif	ウルフルズ-2	音楽	東芝 EMI
lbr9.aif	男声スピーチ (蟹江)	スピーチ	IMPACT2
lbr10.aif	女声スピーチ (高橋)	スピーチ	IMPACT2
lbr21.aif	女声スピーチ+背景音なし	スピーチ	JWAVE
lbr17.aif	男声スピーチ+背景音なし	スピーチ	JWAVE
lbr15.aif	男声スピーチ+ABBA	スピーチ+背景音	JWAVE+SQAM
lbr16.aif	男声スピーチ+クラシック	スピーチ+背景音	JWAVE+SQAM
lbr20.aif	女声スピーチ+クラシック	スピーチ+背景音	JWAVE+SQAM

表2 ARIB評価テスト用音源

的に意味のある主観評価結果を得るために、非専門家、専門家問わず20名以上の評定者を募集したところ、33名の評定者が参加した。そのうち非専門家は10名、専門家は23名であった。

3.1.2 評価結果の分析

評価尺度として5段階品質尺度を用いたため、各評価音は1.0～5.0の評点をつけられる。統計分析はその評点を用いて、まず評定者の信頼性の検定を行い、信頼できると判断された評定者のみを用いて評価結果を導出した。また分散分析の結果、非専門家・専門家間の有意差が存在することがわかったため、評定者全体の集計のほか、非専門家・専門家別の集計を行った。その後、評点の平均値および95%信頼区間を算出した。個別の評価結果は本報告の目的ではないためここでは示さないが、興味のある方は文献[6]を参照してほしい。

3.2 客観音質測定法の適合性

ARIB低ビットレート音声主観評価に使用した評価音の客観音質を、PEAQ客観音質測定法、およびジェニスタ社客観音質測定法を用いて測定した。これらの客観音質測定法は、「BS.1116主観評価法」の主観評価結果に対応した、-4.0～0.0の測定値を出力する。一方「ARIB低ビットレート音声主観評価(MPEG-2 AAC)」は評価尺度として5段階品質尺度を用いており、各評価音の評点は1.0～5.0の値となる。このように主観評価値と客観測定値が異なるが、相関や散布図を用いて適合性を求めることができる。ここで、PEAQ客観音質評価値はBasic Versionを用いて測定した。

ARIB低ビットレート音声主観評価結果に対する高音質向き客観音質測定法の適合性を、以下の観点から検討した。

- (1) 全評価音源に対する適合性
- (2) 音源カテゴリーに対する適合性
- (3-1) ビットレート固定時の、音源に対する適合性
- (3-2) 音源固定時の、ビットレートに対する適合性

ところで、3.1.2節で述べたとおり主観評価値は、評定者全体の集計のほか、非専門家・専門家別に集計している。しかし後述する表3の様に専門家、非専門家別の相関と、評定者全体の相関に大きな差は見られなかった。これは、評定者カテゴリー間の主観評価値の相違に比べて、客観測定値との相違が大きいためと考えられる。ここでは多くの評定者から得た主観評価値を使用することが適切であると考え、評定者全体のデータを用いて検討を行った。

3.2.1 全評価音源に対する適合性

全評価音源に対する客観音質測定値と主観評価値の相関および散布図を、表3および図4に示す。表3のとおり両客観測定法とも評定者全体、専門家、非専門家別の主観評価結果に対する相関は0.7強で、両測定法に大きな差は見られない。この相関値はITU-RがPEAQ客観音質測定法の制定時に行った未知データベースとの相関値²⁾と比較して、相関の高い場合の値0.85には及んでいないが、低い値0.67よりも大きな値である。これより両客観音質測定法とも、ある程度の相関

性があるといえる。ただし、ITU-Rでの結果0.85～0.67は高音質についての評価で主観評価が難しいことに対し、ARIB低ビットレートの結果、約0.74は比較的音質差が大きいことを考慮すると、より大きな相関が期待されるかもしれない。

3.2.2 音源カテゴリーに対する適合性

ARIB低ビットレート音声主観評価は、3.1節のとおり、評価用音源を3種の音源カテゴリー（音楽、スピーチ、スピーチ+背景音）に分け、各カテゴリーから万遍なく、計10音源を選定し評価している。この音源カテゴリー毎の客観音質測定法の適合性を検討した。

音源カテゴリー毎の散布状況を観測したところ、「スピーチ、スピーチ+背景音」「音楽」の2カテゴリーに分類できることがわかった。そこで、「スピーチ、スピーチ+背景音」「音楽」のカテゴリーにわけて分析を行った。

相関	評定者全体	専門家	非専門家
PEAQ 測定法	0.751	0.745	0.744
ジェニスタ社測定法	0.741	0.739	0.725

表3 客観音質測定値と主観評価値の相関

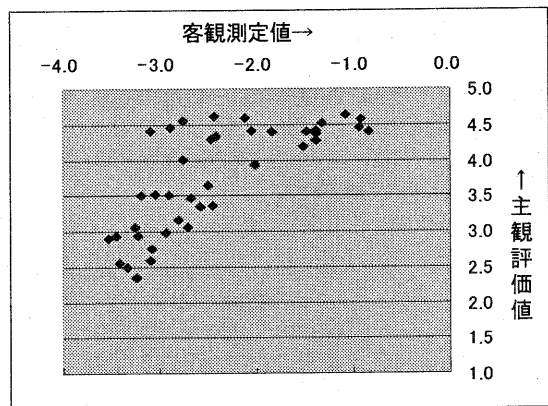


図4(a) PEAQ客観音質測定値と主観評価値

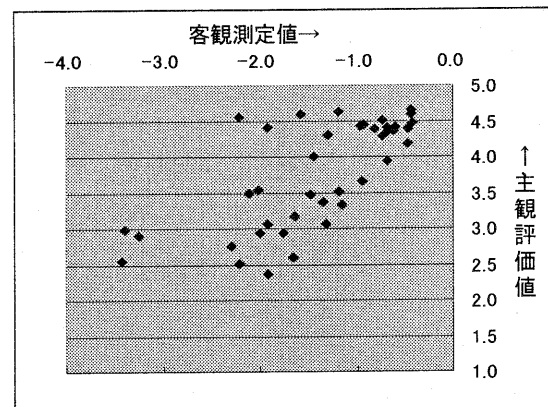


図4(b) ジェニスタ社客観音質測定値と主観評価値

音源カテゴリ毎のPEAQ客観音質測定値、およびジェニスタ社客観音質測定値と主観評価値の相関および散布図を、表4および図5に示す。図5において「スピーチ、スピーチ+背景音」カテゴリは◆で、「音楽」カテゴリは■で示している。両客観測定法とも、全カテゴリの相関よりも、「スピーチ、スピーチ+背景音」「音楽」カテゴリ毎の相関が大きくなっている。特に、「スピーチ、スピーチ+背景音」の相関は両客観測定法とも0.9強で、主観評価値とよく相関のとれた音質が測定されているといえる。一方、「音楽」カテゴリについては、ジェニスタ社客観音質測定法の相関がPEAQ客観音質測定法よりよい値を示している。また、図5の散布図からジェニスタ社客観音質測定値が-4.0~-0.5の範囲に分布しているのに対し、PEAQ客観音質測定値の分布は-4.0~-2.0の範囲となっている。

カテゴリ毎の相関	全カテゴリ	スピーチ、スピーチ+背景音	音楽
PEAQ 測定法	0.751	0.916	0.772
ジェニスタ社測定法	0.741	0.911	0.849

表4 カテゴリ毎の客観音質測定値と主観評価値の相関

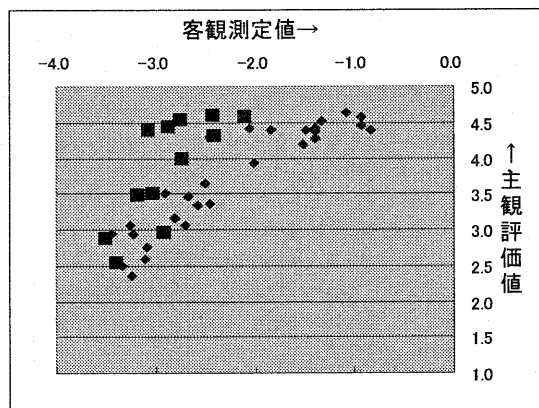


図5(a) PEAQ客観音質測定値と主観評価値

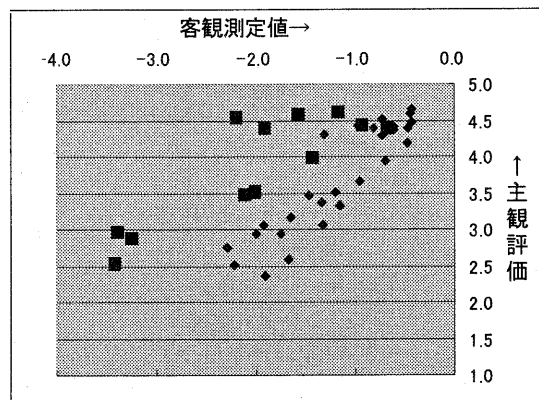


図5(b) ジェニスタ社客観音質測定値と主観評価値

以上より、音源カテゴリを「スピーチ、スピーチ+背景音」に限定すれば、両客観音質測定法とも実用に応用できる可能性が高い。また、「音楽」カテゴリにおいてはジェニスタ社客観音質測定法がPEAQ客観音質測定法より優れていると考えられる。

3.2.3 ビットレート毎、音源毎の適合性

ビットレートを固定とし、全10音源の主観評価値とPEAQ客観音質測定値およびジェニスタ社客観音質測定値との相関を求めた。相関値を表5に示す。音源毎の相関は大きくばらついており、現状同一ビットレートにおける音源毎の音質差を測定することは困難であるといえる。

一方、音源の種類を固定し、10種類の音源から1つを選択し、これをビットレート80,64,48,32kbpsと可変した時の主観評価値とPEAQ客観音質測定値、ジェニスタ社客観音質測定値との相関を算出した。この相関値を表6に示す。表6の結果から、一部を除きほぼ全ての音源で相関が高い。とりわけジェニスタ社客観音質測定法の場合、全ての音源で相関が0.9よりも大きな値となった。またPEAQ客観音質測定法の場合、音源カテゴリが「音楽」の場合を除いて相関が0.9よりも大きな値となった。これより客観音質測定法において音源を固定すれば、ビットレートを変化させた事による音質の変化が測定可能であると考えられる。ただし、ビットレートは4種類のみであり、今後詳細な検討が必要である。

ビットレート一定時の相関	80kbps	64kbps	48kbps	32kbps
PEAQ 測定法	-0.009	-0.128	0.126	0.024
ジェニスタ社測定法	-0.467	-0.308	0.489	0.209

表5 ビットレート固定、音源毎の相関

スピーチ	Lbr 9	Lbr10	Lbr17
PEAQ 測定法	0.982	0.995	0.975
ジェニスタ社測定法	0.994	0.998	0.999
スピーチ	Lbr 21		
PEAQ 測定法	0.968		
ジェニスタ社測定法	0.968		
スピーチ+背景音	Lbr 15	Lbr 16	Lbr 20
PEAQ 測定法	0.927	0.988	0.928
ジェニスタ社測定法	0.969	0.997	0.977
音楽	Lbr 5	Lbr 7	Lbr 13
PEAQ 測定法	0.712	0.839	0.846
ジェニスタ社測定法	0.945	0.909	0.962

表6 音源固定、ビットレート毎の相関

4. その他のオーディオ符号化方式の適合性

統計的に意味のある主観評価が行われ、その元となる評価音及び原音が容易に入手可能なものは、3章に示した「ARIB 低ビット音声主観評価(MPEG-2 AAC)」のみであった。そこで、AAC以外の符号化方式における客観音質測定法の適合性を調査する第一歩として、表7に示す符号化音について簡単な適合性実験を行い、可能性を検討した。

低ビットレートオーディオ符号化のデモ用として配布されているSBR (mp3pro)^[10]、PAC 符号化^[11]の音源素材を収集した。また、フジテレビ提供の6音源(音楽、人の声、楽器、ノイズ等)を用いて、mp3^[12]ソフトウェア(FhG IIS .mp3 Producer ver2.1)により、符号化音を作成した。また参考のため同じ音源をAACハードウェアにより符号化した。これらの符号化音と原音のCD-ROMを作成し、各自に所有する適当と判断した環境、設備を用いて試聴し、主観的な評価を行った。また3章と同様に、両客観音質測定法を用いて音質を測定した。

試聴した主観評価結果と客観測定値を比較した結果、両者には何らかの相関がありそうとの合意に達しているが、詳細な検討は今後の課題である。

5. まとめ

AM～FM程度の低ビットレート・オーディオ符号化の主観的な音質を客観的に測定する方法を検討するため、高音質向きの客観音質測定法の適用可能性を調査をした。調査した客観音質測定法は、PEAQ客観音質測定法とジェニスタ社客観音質測定法である。ARIBが1998～99年に行った「ARIB 低ビット音声主観評価(MPEG-2 AAC)」のデータを用いて、これら2つの客観音質測定法の適合性を調査した。

その結果、以下のことが分かった。

- (1) 両客観音質測定法とも、音源カテゴリーを「スピーチ、スピーチ+背景音」に限定すれば0.9強の高い相関をもつ。また「音楽」カテゴリーにおいてはジェニスタ社測定法がPEAQ測定法より、よい値を示した。
- (2) 両客観音質測定法とも、音源の種類を固定し、ビットレートを可変した時の相関値は、ほぼ全ての音源で相関が高い。とりわけジェニスタ社測定法は全ての音源で相関が0.9よりも大きい。一方ビットレートを固定した時、音源毎の相関には大きなばらつきが見られた。

これより、すくなくともAAC符号化において、「スピーチ、スピーチ+背景音」に限定すれば、両客観音質測定法とも実用に応用できると考えられる。また、音源カテゴリーが事前に指定できる場合などは、「音楽」カテゴリーも含めてジェニスタ社客観音質測定法の使用も想定できるかもしれない。

今後は、音源カテゴリーによらない客観音質測定法やAAC以外のオーディオ符号化方式との適合性について、引き続き調査を行う予定である。

謝辞

この調査は電波産業会(ARIB)・品質評価法調査研究会において行ったものである。調査研究会メンバーをはじめ関係者一同に感謝の意を表する。

文献

- [1] 小特集「音声圧縮技術」、映情学誌、Vol.55, No.12, pp.1557-1587, December 2001.
- [2] ITU-R Rec. BS.1387-1, "Method for objective measurements of perceived audio quality", Nov. 2001.
- [3] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs", Feb. 2001.
- [4] ITU-R Rec. BS.1534, "Method for the subjective assessment of intermediate quality level of coding systems", June. 2001.
- [5] ITU-R Rec. BS.1284, "Methods for the subjective assessment of sound quality - General requirements", Oct. 1997.
- [6] 電波産業会「MPEG-2 AAC 低ビットレート符号化音質評価実験報告書」
- [7] ITU-R Rec. BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems", Oct. 1997.
- [8] W. C. Treurniet, G. A.Soulodre, "Evaluation of the ITU-R Objective Audio Quality Measurement Method" J.Audio Eng.Soc., Vol.48, No.3, pp.164-173, March 2000.
- [9] ISO/IEC 13818-7, "Generic Coding of Moving Pictures and Associated Audio Information - Part 7:Advanced Audio Coding", Apr. 1997.
- [10] T. Ziegler et al., "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm", 112th AES Convention, May 2002.
- [11] J. D. Johnston et al., "AT&T Perceptual Audio Coding (PAC)" AES Collected Papers on Digital Audio Bit-Rate Reduction, pp.73-82, 1996.
- [12] ISO/IEC 11172-3, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to 1.5Mbps - Part 3: Audio", 1993.

符号化方式	サンプリング 周波数 (kHz)	ビットレート (kbps)・ステレオ	音源素材
AAC ハードウェア	48	64	フジテレビ提供素材
mp3 ソフトウェア	24	32, 48, 64	フジテレビ提供素材
SBR (mp3pro)	不明	48, 64	デモ CD 収録素材
PAC	48	96	デモ CD 収録素材

表7 試聴したオーディオ符号化方式と符号化パラメータ