ISO/IEC MPEG　USAC　音声音響統合符号化

Technology Reports

音声と音楽の高効率な圧縮を実現するMPEG標準音声音響統合符号化方式

先進技術研究所　菊入（きくいり）　圭（けい）　仲（なか）　信彦（のぶひこ）

国際標準化機関ISO/IECでMPEG音声音響統合符号化方式（USAC）の標準化が完了フェーズにある．USACは，低ビットレートの音響符号化の課題であった音楽の品質に比べ音声の品質が低い点を改善し，入力信号によらず音声も音楽も高効率に圧縮できることを特徴とする．ドコモは，従来の音響符号化の帯域拡張方式の課題であった拍手や打楽器といった音源の品質改善に貢献するinter-TES技術をUSACに提案し，採用された．USACの登場により，今後のモバイルオーディオサービスのさらなる高音質化が期待される．

1. まえがき

従来，音声と音楽はそれぞれの用途の違いにより異なる符号化方式により情報圧縮されてきた．音声は主に電話用途として，狭帯域（～3.5kHz）の信号を低ビットレートで符号化する音声符号化を用いた．一方，音楽は主に放送・蓄積用途として，広帯域（～20kHz）の信号を高ビットレートで符号化する音響符号化を用いてきた．そのため，モバイル用途に適した低～中ビットレートで音声と音楽が混在したコンテンツを圧縮するには，音声符号化，音響符号化のどちらの符号化技術を用いても，音声と音楽のいずれかの品質が劣化してしまうという課題があった．そこで，国際標準化機関ISO（International Organization for Standardization）[*1]/IEC（International Electrotechnical Commission）[*2]は，音声と音楽の両方の品質を損なうことなく1つの符号化で対応できる技術として，MPEG（Moving Picture Experts Group）[*3]音声音響統合符号化方式（USAC：Unified Speech and Audio Coding）の標準化に着手した．

本稿では，MPEG USACの概要と，ドコモが提案しUSACに採用された帯域拡張方式の品質改善技術である，インターサブバンドサンプル時間包絡整形（inter-TES：inter-sub-band-sample Temporal Envelope Shaping）の概要を解説する．

2. USAC

2.1 USACの要件と想定アプリケーション

USACは，入力信号によらず最良の音声符号化および音響符号化と同等以上の品質を達成することを要求条件に，2007年10月に提案募集がされた．2008年7月に，応募された7件の技術からUSACの基本アルゴリズムとなる参照モデルを1件選定した．その後，参照モデルの性能をさらに改善するために数々のコア実験が行

© 2011 NTT DOCOMO, INC.
本誌掲載記事の無断転載を禁じます．

＊1　ISO：国際標準化機構．情報技術分野の標準化を行う組織であり，電気および電子通信分野を除く全産業分野に関する国際規格を作成する．
＊2　IEC：国際電気標準会議．情報技術分野の標準化を行う組織であり，電気および電気通信の国際規格を作成する．

われ，2011年9月に最終規格草案が発行され，2011年度中に規格化される予定となっている．USACのアプリケーションには，

・モバイル機器向けのマルチメディアダウンロード
・ユーザ作成コンテンツサービス（UGC：User Generated-Contents）
・デジタルラジオ
・モバイルTV
・オーディオブック

が想定された．いずれも，音声・音楽が混在したコンテンツを取扱うアプリケーションであり，USACによるさらなる高音質化が期待される．

2.2 USACの技術的特徴

USACの基本アルゴリズムは，時間周波数変換符号化[＊4]と線形予測符号化[＊5]の組合せによるコア符号化，eSBR（enhanced Spectral Band Replication）[＊6]による帯域拡張，MPEGサラウンド（MPS：MPEG Surround）[＊7]によるマルチチャネル化により構成される．**図1**にUSACデコーダの基本構成を示す．

(1)コア符号化

コア符号化デコーダでは，低周波数帯域の信号を復号する．入力信号の特性に応じて，コア符号化に用いられる符号化は切り替えられる．一般的に，音楽のように定常的な信号では時間周波数変換符号化，音声のように過渡的な信号では線形予測符号化が用いられる．

時間周波数変換符号化には，修正AAC（modified Advanced Audio Coding）が用いられる．修正AACはMPEG-4 AAC[＊8][1]とほぼ同様であるが，スペクトルノイズレス符号化[＊9]にハフマン符号化[＊10]の代わりに算術符号化[＊11]を用いることで，より高効率の符号化を実現している点が異なる．

線形予測符号化は，TCX（Transform Coded Excitation）[＊12]とACELP（Algebraic Code Excited Linear Prediction）[＊13]の組合せにより構成される．TCXとACELPは線形予測を用いる点が共通であり，TCXはその予測残差信号[＊14]を周波数領域で符号化するのに対し，ACELPは時間領域で符号化する点が異なる．この構成は，3GPP AMR-WB+[＊15][2]と基本的に同じである．ただし，USACのTCXでは修正AACと同様の修正離散コサイン変換（MDCT：Modified Discrete Cosine Transform）[＊16]が用いられており，またその変換係数も同様に算術符号化により符号化される点がAMR-WB+と異なる．なお，一般的には音声や非常に時間変動の大きい信号にはACELPが用いられ，過渡的ではあるが比較的時間変動の小さい信号にはTCXが用いられる．

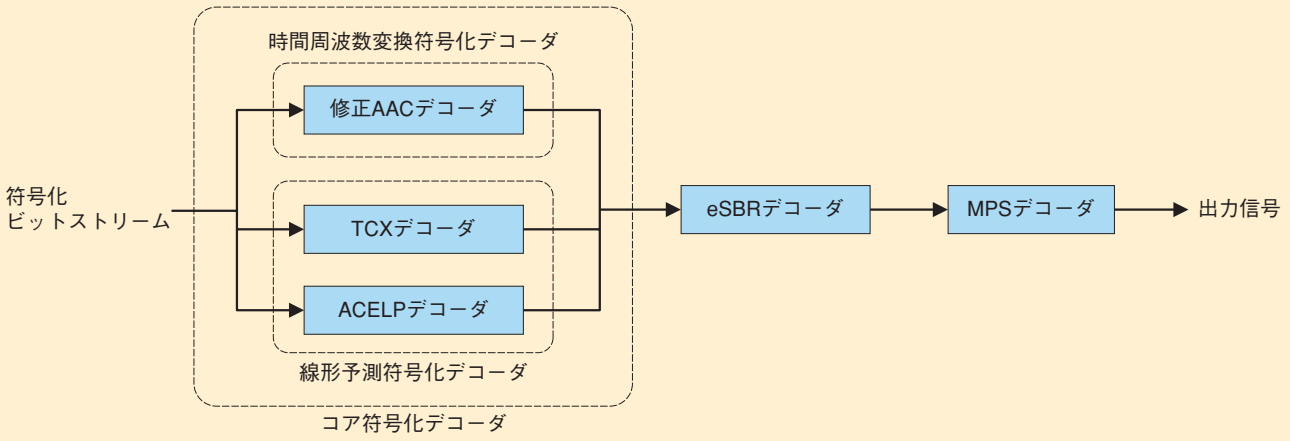

図1　USACデコーダの構成

＊3 **MPEG**：デジタル音声や映像の符号化および伝送方式などの技術標準仕様．ISOとIECの合同専門委員会の作業部会が策定した．MPEG-2では，デジタルテレビ放送やDVDへ適用，MPEG-4では移動端末などの低ビットレート領域まで適用領域を拡大した方式．

＊4 **時間周波数変換符号化**：時間信号を離散コサイン変換（DCT：Discrete Cosine Transform）やMDCT（＊16参照）に代表される直交変換で周波数領域に変換し，周波数領域にて圧縮する符号化方式．

＊5 **線形予測符号化**：線形予測分析を用いて予測可能な成分の冗長性を除去して圧縮する符号化方式．

＊6 **eSBR**：ISO/IECで規格化された帯域拡張方式SBRを発展させた技術．

＊7 **MPEGサラウンド（MPS）**：ISO/IECで規格化されたマルチチャネルオーディオ符号化方式の1つ．マルチチャネルの信号を，実際のチャネル数よりも少ない音響信号とパラメータで表現することで，低ビットレートで高品質の符号化を実現する．

＊8 **AAC**：ISO/IECで規格化されたオーディオ符号化方式の1つ．MDCT（＊16参照）により時間周波数変換された信号を符号化する．

＊9 **スペクトルノイズレス符号化**：AACにおいて，量子化されたMDCT（＊16参照）係数をハフマン符号化（＊10参照）により圧縮する．

＊10 **ハフマン符号化**：エントロピー符号化の一種．情報の出現確率に応じて割り当てる符号語が決定する．

(2)帯域拡張

eSBRは，低周波数帯域の信号から高周波数帯域の信号を生成する帯域拡張方式であるMPEG-4 SBR（Spectral Band Replication）[1]を発展させた技術である．eSBRではMPEG-4 SBRに加えて，低周波数帯域のハーモニクス[*17]を規則正しく高周波数帯域に拡張できるHarmonics Transposer，低周波数帯域の周波数スペクトル包絡[*18]を用いて高周波数帯域の周波数スペクトル包絡を符号化するPVC（Predictive Vector Coding），低周波数帯域の時間包絡を用いて高周波数帯域の時間包絡を調整するinter-TES，低周波数帯域の信号と高周波数帯域の信号のサンプリング周波数の比を従来の1:2以外に3:8，1:4に設定できる機能を備えている．これらの技術により，従来のSBRと比較して低ビットレートでの品質が向上している．

(3)マルチチャネル化

MPSは，マルチチャネルの信号をより少ないチャネル数の信号とパラメータにより高効率に符号化する方式である[3]．USACにおいては，MPSによるモノラル信号からステレオ信号への拡張が規定され，パラメータとしてチャネル間位相差（IPD：Inter-channel Phase Difference）[*19]の伝送が可能になり，さらに過渡的な信号用のデコリレータ[*20]（TSD：Transient Steering Decorrelator）が追加された．

3. inter-TES

inter-TESは，音響信号の低周波数帯域と高周波数帯域の時間包絡の相似性を利用して，SBRで生成された高周波数帯域信号の時間包絡を，コア符号化デコーダで復号された低周波数帯域信号の時間包絡を用いて整形する技術である．SBRの構成を**図2**に示す．SBRではフレームをSBR

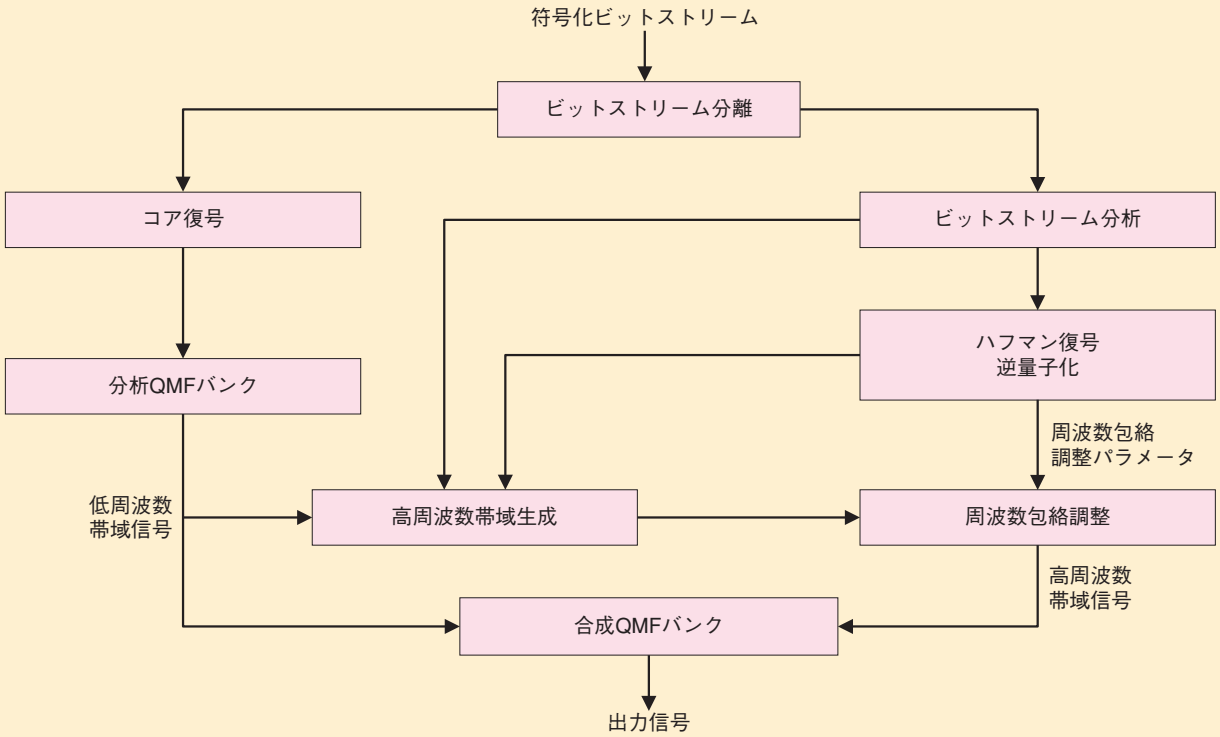

図2　SBRの構成

＊11 **算術符号化**：エントロピー符号化の一種．各情報の出現確率に応じて割り当てる符号語が決定する．一般的に，ハフマン符号化よりも圧縮効率が良いといわれる．

＊12 **TCX**：周波数領域で符号化された励振信号（合成フィルタへの入力信号）を用いて，線形予測により求められる合成フィルタを駆動する符号化方式．

＊13 **ACELP**：過去の励振信号で構成される適応符号帳と複数のパルス列で構成される代数符号帳とからなる励振信号を用いて，線形予測により求められる合成フィルタを駆動する符号化方式．

＊14 **予測残差信号**：参照信号と予測信号との差の信号成分．

＊15 **AMR-WB+**：3GPPで標準化された音声符号化方式AMR-WBを，音楽などの一般的な音響信号にも対応できるように拡張した符号化方式．

＊16 **修正離散コサイン変換（MDCT）**：時系列信号を周波数成分に直交変換する手法の1つ．前後の隣接する変換ブロックを窓かけして重ね合わせる変換で，情報の無駄なく変換ブロックの境界歪みを防ぐことができることから，音響符号化において広く利用されている．

＊17 **ハーモニクス**：基本となる周波数と，その整数倍の周波数で構成される信号成分．

＊18 **周波数スペクトル包絡**：周波数スペクトルの概形．

＊19 **チャネル間位相差（IPD）**：複数チャネルの信号における各チャネル間の信号の位相の差．

＊20 **デコリレータ**：主成分信号との相関が小さくなるような信号を生成する機能．

envelopeと呼ばれる時間セグメントに区切り，さらに分析QMF（Quadrature Mirror Filter）バンク[*21]でSBR envelopeを複数の周波数帯域（サブバンド）に分割し，この周波数帯域ごとにエネルギー包絡を整形する．すなわち，SBRのエネルギー包絡整形の時間分解能はSBR envelope単位である．SBR envelopeが十分に短くない場合には，時間変動が大きい信号が入力されると，時間変動の大きい部分の前後にプリエコー[*22]／ポストエコー[*23]と呼ばれる歪みが生じる．一方で，SBR envelopeを短くすると，時間方向のエネルギー包絡を整形するために多くの情報量が必要となり，符号化に要するビット量は増大する．

この課題を解決するためには，少ないビット量かつ高い時間分解能でエネルギー包絡を制御する必要があり，ドコモが提案したinter-TESにより実現されている．

図3にinter-TESの構成図を示す．まず，コア符号化デコーダで復号された低周波数帯域信号の時間包絡を算出する．符号化ビットストリーム[*24]

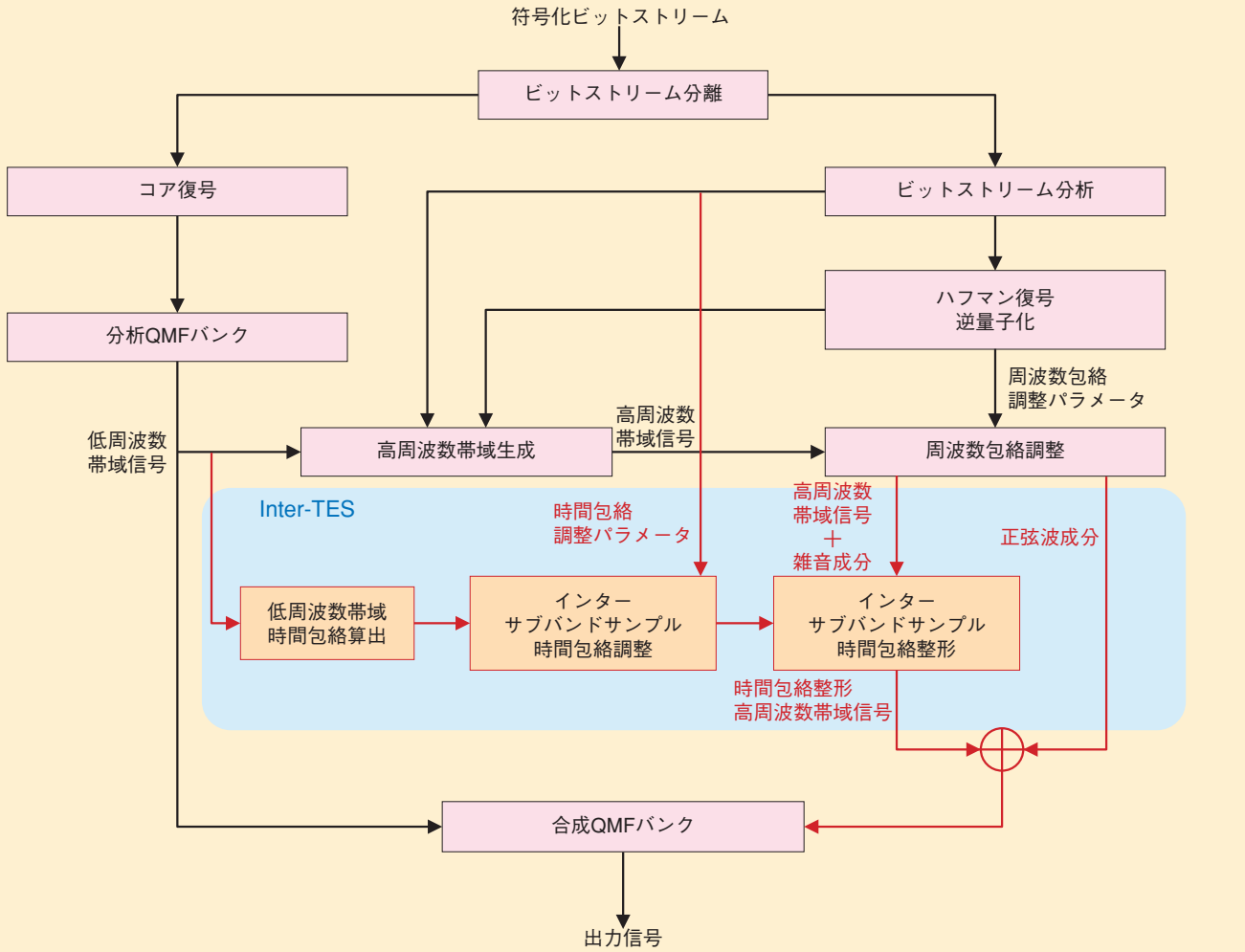

図3　inter-TESの構成

＊21 **QMFバンク**：入力信号を複数の周波数成分に分割するフィルタバンクの一種．

＊22 **プリエコー**：信号の振幅が急激に増加する直前において量子化誤差がエコーのような歪みとして知覚される現象．

＊23 **ポストエコー**：信号の振幅が急激に減少した直後において量子化誤差がエコーのような歪みとして知覚される現象．

＊24 **ビットストリーム**：符号化されたビットの列．

に含まれる時間包絡調整パラメータを用いて，低周波数帯域信号の時間包絡を調整し，高周波数帯域信号の時間包絡を整形するためのゲインを算出する．算出されたゲインを高周波数帯域信号に適用することにより，時間包絡を整形した高周波数帯域信号を得る．

図4にinter-TESにより時間包絡を整形される前後の高周波数帯域信号の時間波形を示す．横軸は時間，縦軸は信号の大きさをあらわす．入力信号の高周波数帯域信号(a)はインパルス信号のように急峻に立ち上がった後に急峻に立ち下がっているのに対し，SBRで生成される高周波数帯域信号(b)は立下がりの後ろに歪みが生じている．一方，inter-TESにより時間包絡整形された信号では，立下がりの後ろに生じていた歪みが抑制されており，より入力信号の時間包絡に近くなっていることがわかる．

主観評価試験により，inter-TESの性能を評価した．主観評価試験は8名の被験者によるMUSHRA（Multi Stimuli with Hidden Reference and Anchors）法[*25][4]で行い，ビットレートはモノラル16kbit/s，24kbit/s，評価音源は2種類の拍手音源である．**図5**にinter-TESによる時間包絡整形した信号の主観評価試験結果を示す．ここでは，inter-TES有りの場合とinter-TES無しの場合での試験スコアの差をプロットした．エラーバーは95％信頼度区間を示す．この結果より，16kbit/sおよび24kbit/sにおいて，inter-TESによる主観品質の改善を確認できた．

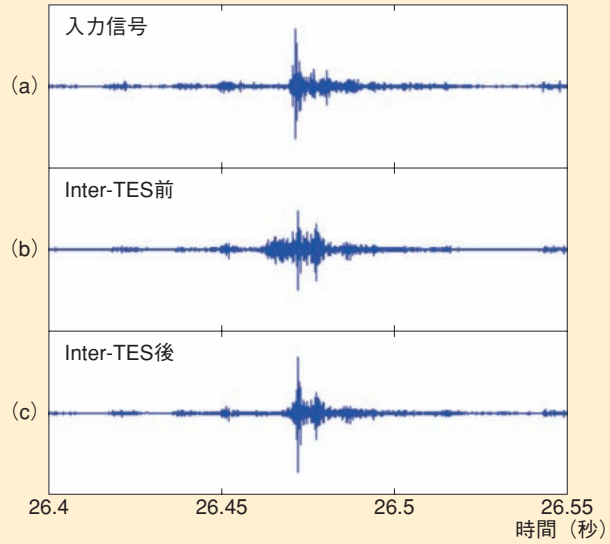

図4　高周波数帯域信号の時間波形

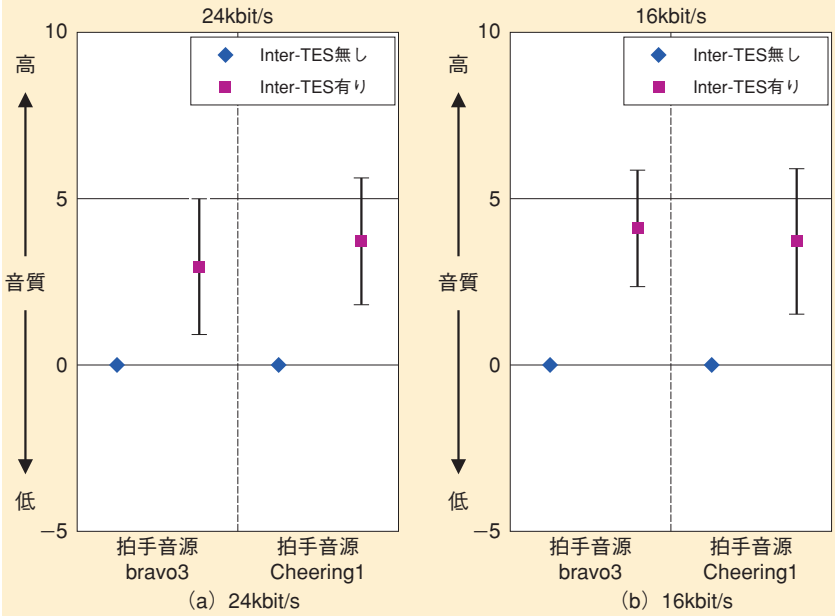

図5　inter-TESの主観評価試験結果

＊25 **MUSHRA**：音声符号化方式，音響符号化方式の性能を評価するための主観評価試験方法の1つ．複数の評価対象音と原音と比較して0～100点の範囲で同時に採点する．評価対象音は，性能評価する符号化方式の復号信号に加え，原音，原音を帯域制限した信号を含む．

4. あとがき

MPEGにおけるUSACの標準化は，2011年6月に性能確認試験実施され，2011年7月に試験を実施した各試験サイトの試験結果が報告され，9月に報告書が提出された．各試験サイトの試験結果より，USACの要件にあるとおり，音声・音楽信号の両方に対して最新の音声符号化・音響符号化以上の性能が確認された．USACはHE（High-Efficiency）-AAC＊26 v.2[1]の後継符号化方式となることが期待される．HE-AAC v.2は着うた®＊27などの音楽配信，BeeTV™＊28などの動画配信，ワンセグなどのモバイル向け放送に利用されている．USACも同様のサービスへの応用が可能であり，モバイルオーディオサービスのさらなる高音質化に貢献できると確信する．

文　献

[1] ISO/IEC 14496-3:2009：“Information technology - Coding of audio-visual objects – Part 3: Audio,” Aug. 2009.

[2] 3GPP TS.26.290 V. 6.3.0：“Audio codec processing functions; Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions,” Jun. 2005.

[3] ISO/IEC 23003-1:2007：“Information technology – MPEG audio technologies – Part 1: MPEG Surround,” Jan. 2007.

[4] ITU-R Recommendation BS.1534-1：“Method for the subjective assessment of intermediate quality level of coding systems,” Jan. 2003.

＊26 **HE-AAC**：MPEG-4 AACに対し約半分のビットレートで同等の音質を実現する音声圧縮・符号化方式．MPEG-4 AACの拡張仕様．

＊27 **着うた®**：株式会社ソニー・ミュージックエンタテインメントの登録商標．

＊28 **BeeTV™**：エイベックス・エンタテインメント株式会社の商標．