EVS　音声符号化　標準化

Technology Reports

VoLTEのさらなる高音質化と音楽の活用を実現する3GPP標準音声符号化方式EVS

VoLTEのさらなる高音質化を実現するEVSの標準化活動において，ドコモは将来のニーズを見据えた音声符号化方式となるように標準化方針の決定に貢献してきた．さらに技術提案を通してFMラジオ並みの高い音声品質と従来の音声符号化方式が苦手としていた音楽の高効率な圧縮の両立に寄与してきた．EVSの登場により，音楽を活用した従来にない豊かなコミュニケーションの実現が期待される．

先進技術研究所　堤（つつみ）　公孝（きみたか）　菊入（きくいり）　圭（けい）

1. まえがき

VoLTE（Voice over LTE）のサービス開始と音声定額料金プランにより，高品質な通話コミュニケーションの価値が注目されている．そのような中，3GPP（3rd Generation Partnership Project）は，音声符号化方式EVS（Enhanced Voice Services）[1]の規格を2014年9月に完成させた．

EVSは，これまでの音声通話サービスを大きく上回る音質を目標に標準化作業が開始された．従来サービスでは，現在のドコモのVoLTEで利用されているAMR-WB（Adaptive Multi-Rate WideBand）*1 [2]に代表される，サンプリング周波数*2 16kHzの広帯域音声に対応した音声符号化方式を用いてAMラジオ並み*3の音質が達成可能であった．一方，EVSはサンプリング周波数32kHzの超広帯域*4音声にも対応しFMラジオ並み*5の音質の達成が可能になることから，EVSのVoLTEへの導入により，音声通話サービスの飛躍的な品質改善が期待できる．また，MPEG USAC（Moving Picture Experts Group Unified Speech and Audio Coding）*6 [3] [4]に代表される，音声に加えて音楽も高音質・高効率に符号化できる非リアルタイムサービス用途の音声・音響統合符号化方式の登場により，リアルタイムサービス用途のEVSにおいても音声のみならず，音楽にも対応することが要求条件に盛り込まれた．さらに，AMR-WBを用いた音声通話サービスが世界的に広がりつつある状況を考慮して，EVSにはAMR-WBと互換性をもつモードも備えることが要求された[5]．

ドコモは2010年よりEVSの標準化に参画し，VoLTEへの導入によるネットワークの変更を最小化するようなEVSの設計条件を設定することで早く・広い普及を目指すべきと主張してきた．これにより，VoLTEにおける無線区間伝送時のデータサイズがAMR-WBと同じになるようにEVSのビットレートが設定され，AMR-WB利用時のVoLTEの無線ネットワーク設計をそのまま利用できるようになった．さらに，将来の音声サービスと音楽サービスとの融合によるVoLTEの進化を見据えて，低ビットレートにおける音楽の音質の重要性を主張した．当初は，音楽への要求条件は音声に比べて比較的低く設定される方向で議論が進められていたが，最終的には音楽の符号化において有利な，EVSよりも長い原理遅延*7をもつ符号化方式と同程

＊1 **AMR-WB**：音声電話サービス等で利用される，AMR-NBよりも高品質な音声コーデック．
＊2 **サンプリング周波数**：マイクなどから入力した音声信号の音圧を1秒間に何回記録するかを表す単位．大きいほど高い音域まで記録できる（単位はHz）．
＊3 **AMラジオ並み**：100Hzから7.5kHzまでの音声帯域を表現できる．
＊4 **超広帯域**：下限周波数が50Hz，上限周波数が14～16kHzの音声帯域．

度の音質，という高い要求条件が設定された．現在の3GおよびVoLTEの音声通話サービスではメロディコール®*8が広く利用されているが，音声に特化された符号化方式であるAMR[6]を利用しているため，必ずしも高い音質を達成できていない．要求条件を達成したEVSにより，このような既存の音楽コンテンツサービスの音質改善に加えて，音声通話サービスにおいて音楽を活用する新しいサービスの登場が期待できる[7]．

3GPPにおける上記の議論を通して，EVSは，①FMラジオ並みの音質を達成する超広帯域音声に対応，②VoLTEへの容易な導入，③音声のみならず音楽も高音質，という特長を持つ音声符号化方式として標準化された．

本稿では，EVSの概要と主要な品質改善技術に加えて，ドコモの貢献技術について解説し，EVSの性能確認として行った超広帯域音声・音楽に対する音質評価試験の結果について報告する．

2. EVSの技術的特徴

2.1 EVS概要

EVSは，広帯域，超広帯域音声に加えて，従来の電話で利用されてきたサンプリング周波数8kHzの狭帯域音声，CDよりも高いサンプリング周波数48kHzのフルバンド音声にも対応し，また設定可能なビットレート範囲は5.9～128kbpsと非常に広い．

EVSエンコーダ・デコーダの基本構成を**図1**に示す．低ビットレート動作モードでは，音声を効率よく圧縮可能な時間領域符号化と，音楽を効率よく圧縮可能な周波数領域符号化をフレーム*9ごとに切り替えて用いる．一方，入力信号の圧縮に十分な情報量が利用できる高ビットレートの動作モードでは，入力信号によらず周波数領域符号化を用いる．

VoLTEのような移動通信環境下においてはパケットロス*10が避けられないため，これにより生じる復号音の欠落を疑似的に生成した音で補うパケットロス隠蔽技術（PLC：Packet Loss Concealment）が重要である．EVSではパケットロス隠蔽技術も標準化されており[8]，パケットロス前で最後に正常に復号できたフレームの符号化領域に応じて，時間領域パケットロス隠蔽技術と，周波数領域パケットロス隠蔽技術を切り替えて用いる．

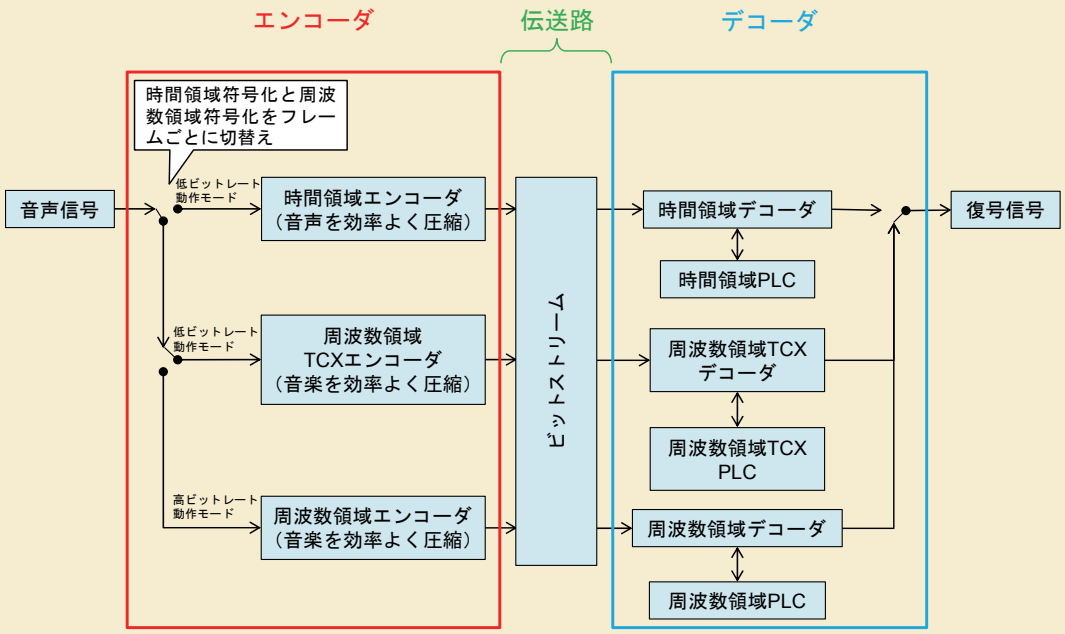


図1　EVSエンコーダ・デコーダの基本構成図

---

*5 **FMラジオ並み**：50Hzから15kHzまでの音声帯域を表現できる．

*6 **MPEG USAC**：MPEG音声音響統合符号化方式．MPEGはデジタル音声や映像の符号化・伝送方式に関する技術標準仕様．ISO/IECの合同作業部会により策定する．

*7 **原理遅延**：原音と比べて復号音がどのくらい遅れて出力されるかの指標．音声符号化方式の仕様で決まり，周波数領域の符号化方式の場合，一般的に長い方が符号化効が向上する．

*8 **メロディコール®**：携帯電話の呼出音を，好みの楽曲に変更できるドコモのサービス．ドコモの登録商標．

*9 **フレーム**：エンコーダ・デコーダが動作する周期，およびそれに対応する長さの音声信号．EVSのフレーム長は20msであり，20msに1回符号化・復号化を行う．

以下，時間領域符号化と周波数領域符号化，およびパケットロス隠蔽技術について概要を解説する．

⑴時間領域符号化

時間領域符号化の構成を**図2**に示す．人間の聴覚は，低周波数帯域成分を敏感に聞き分けられるが，高周波数帯域になるほど鈍感になるため，入力信号を低周波数帯域成分と高周波数帯域成分に分けて高周波数帯域を少ない情報量で符号化することにより，音質を維持したまま効率よく情報量を削減できる．

①低周波数帯域成分は，入力信号から算出した線形予測[*11]係数と，線形予測残差信号[*12]を量子化[*13]して伝送するCELP（Code Excited Linear Prediction）により符号化される．線形予測残差信号の符号化には，類似の波形（ピッチ波形[*14]）が時間的に繰り返す音声信号の性質を利用し，直前のピッチ波形からの差分のみを，ピッチ波形の長さとともに符号化する．AMR，AMR-WBをはじめとする音声符号化の多くが，直前ピッチ波形からの差分の符号化に代数符号帳[*15]を用いる．EVSでは，符号語あたりの表現力が向上した高効率代数符号帳が導入され，品質が大幅に向上している．

②高周波数帯域成分は，帯域拡張符号化により符号化する．帯域拡張符号化は，低周波数帯域成分を整形することにより高周波数帯域成分を作り出す技術であり，整形に必要なパラメータだけを低ビットレートで伝送するだけで高品質な高周波数帯域成分が得られる．EVSでは低遅延での帯域拡張を実現するため，時間領域帯域拡張符号化が採用された．

⑵周波数領域符号化

周波数領域符号化の構成を**図3**に示す．周波数領域符号化では，入力信号を修正離散コサイン変換（MDCT：Modified Discrete Cosine Transform）[*16]を用いて周波数領域表現に変換したうえで，MDCTの係数を符号化する．

MDCT係数の符号化は，MDCT係数をサブバンド[*17]に分割してサブバンドごとのスケールファクタ[*18]とそれにより正規化されたMDCT係数をベクトル量子化[*19]する方式と，線形予測残差信号を周波数領域

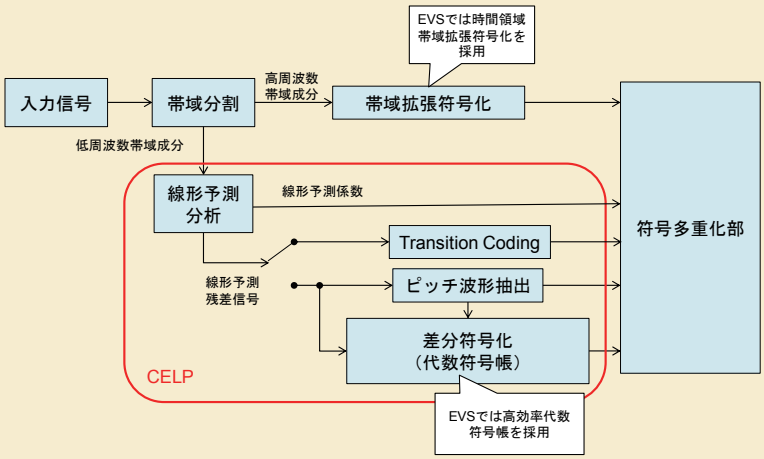


図2　時間領域符号化の構成図

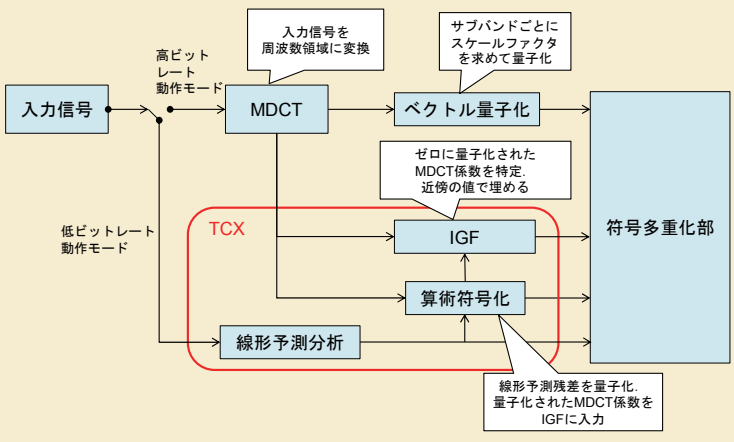


図3　周波数領域符号化の構成図

＊10 **パケットロス**：輻輳などにより，デコードするまでに音声パケットが届かないこと．

＊11 **線形予測**：ある時刻の音声信号を，過去の音声信号の線形和で近似する手法．

＊12 **線形予測残差信号**：入力信号を線形予測した際の予測誤差の信号．

＊13 **量子化**：変換処理にて生成される離散データの値を，飛び飛びの値である粗い区間の代表値に対応づけること．歪みを許しながら大幅に情報量を削減することが可能．

＊14 **ピッチ波形**：音声には類似の波形が繰返し現れる性質がある．その繰返しの1周期分に対応する波形のこと．

＊15 **符号帳**：入力ベクトルを量子化するために，あらかじめ決められた複数の候補ベクトルを登録したもの．

表現に変換して符号化するTCX（Transformed Code Excitation）を用いる方式がある．

TCXは，MPEG USACにおいても採用されていたが，EVSにおいては，誤り耐性能力を高めたうえで圧縮効率を維持できるように改善された算術符号化が導入された．

TCXでは，すべてのMDCT係数を符号化できず，符号化できないMDCT係数は0に量子化されるため，周波数領域で表現した線形予測残差において信号成分がない周波数帯域が生じる．従来，このような周波数帯域に雑音を付加して埋めるNoise Fillingが用いられてきたが，EVSでは，周囲のMDCT係数を用いるIGF（Intelligent Gap Filling）が採用された．

⑶パケットロス隠蔽技術（PLC）

①時間領域パケットロス隠蔽技術

時間領域符号化であるCELPはフレーム間予測＊20を用いるため，パケットロス後に正常に受信できたフレームの復号にもパケットロスの影響が伝搬する性質がある．EVSでは，パケットロス後に正常に受信できたフレームの復号に際して，正常な線形予測係数や線形予測残差信号が得られるよう，直前フレームに依存せずに線形予測係数と線形予測残差信号を符号化するTransition Codingモード＊21が採用され（図2），パケットロスに対する耐性が向上している．

②周波数領域パケットロス隠蔽技術

周波数領域パケットロス隠蔽技術は，基本的に，パケットロス直前のフレームの復号により得たMDCT係数をパケットロスが発生したフレームに複写して復号音を生成する．単純にMDCT係数を複写するとフレーム境界付近で波形の不連続が生じるが，EVSでは，波形の不連続が生じないよう，複写後の波形の位相を周波数領域で調整して波形が滑らかにつなぐ処理を行う．

2.2 EVSの主要な品質改善技術

EVSにはさまざまな改善技術が採用されたが，特に重要と思われる技術について概要を説明する．

⑴時間領域帯域拡張符号化

帯域拡張符号化は，低周波数帯域成分を用いて生成した高周波数帯域成分を，エンコーダから受信した高周波数帯域成分のパワー分布となるよう整形する技術である．

EVSで採用された時間領域帯域拡張符号化は，帯域分割フィルタ＊22から出力した高周波数帯域成分について，線形予測スペクトル＊23を算出して符号化することにより，高周波数帯域成分の大まかな周波数パワー分布を表現する技術である．さらに，復号にあたっては，**図4**に示す通り，符号化された線形予測スペクトルをもつ合成フィルタに，

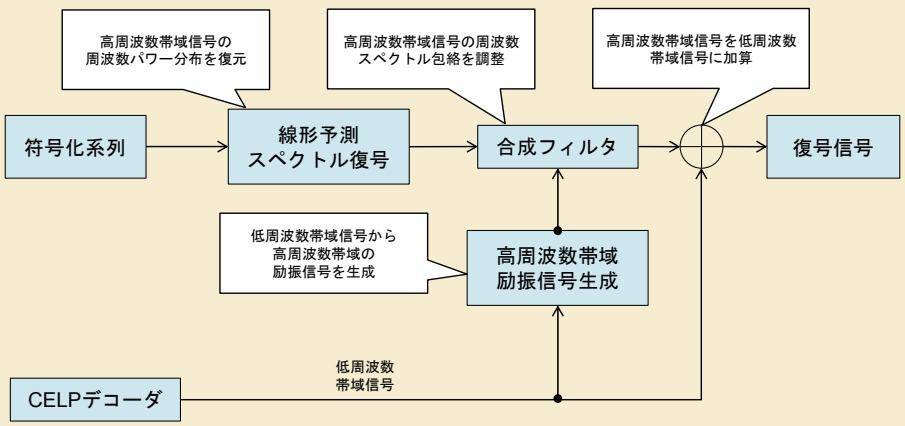


図4　時間領域帯域拡張（復号）

＊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，情報の無駄が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 **サブバンド**：全周波数帯域を複数に分割したうちの1つ．

＊18 **スケールファクタ**：サブバンドのパワーあるいは，振幅を量子化した値．

＊19 **ベクトル量子化**：2以上の長さの数列を，あらかじめ用意した同じ長さの数列のうち類似するもので近似する量子化手法．

＊20 **フレーム間予測**：直前フレームの値からの差分を量子化することで，符号化効率を向上させる手法．

CELPデコーダから出力された低周波数帯域信号を調整した励振信号を入力することで，符号化側で求めた高周波数帯域信号の周波数パワー分布をもつ高周波数帯域信号を生成する．これにより少ない情報量かつ低演算量での高周波数帯域成分の符号化が可能となる．

⑵算術符号化

前述の通り，周波数領域における線形予測残差の量子化において，算術符号化が採用された．

図5に示す通り，隣り合う2つのMDCT係数を2進数表現してビット平面を作り，直前の量子化結果を基準に選んだ符号帳を用いて，高位のビットを量子化する．利用可能な残りのビット数に応じて，下位ビットを量子化する．

十分な数の線形予測残差を符号化できる高ビットレートでは，直前フレームにおける復号で得られた線形予測残差を用いて符号帳を決定する．一方，十分な数の線形予測残差が得られない低ビットレートでは，線形予測スペクトルを用いて符号帳を決定する．

⑶IGF

IGFでは，**図6**のように，符号化できずに欠落した周波数帯域に，近傍の符号化できたMDCT係数を複写するが，複写元のMDCT係数が強いピークをもつ際には，複写先の高域にも強いピークが生じるため音質が低下する．高周波数帯域に生じる不必要な強いピークを抑圧するため，必要に応じて周波数領域での白色化処理[*24]を行う．

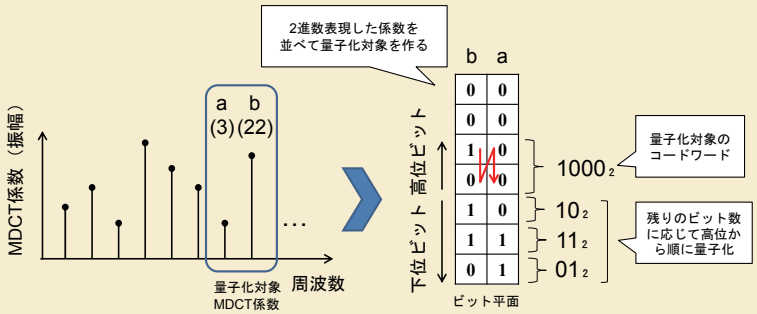


図5　算術符号化

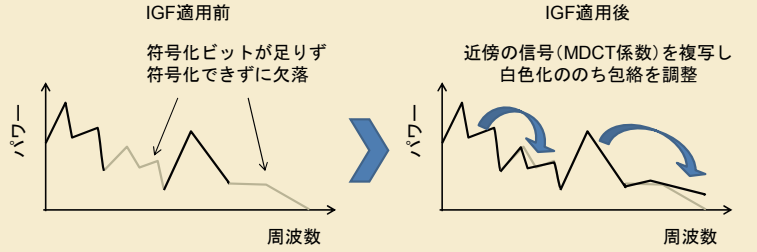


図6　IGF

IGFは，エンコーダ側で周波数スペクトルの概形と，白色化処理のオン・オフの情報を符号化・伝送し，デコーダ側で周波数スペクトルを再現するようMDCT係数を調整する．これにより，IGFは，特に音楽に対して，低ビットレートで高音質な符号化を可能にした．

2.3　ドコモ貢献技術

⑴高周波数帯域成分改善技術

EVSにおける高周波数帯域成分の符号化は，時間領域符号化では帯域拡張符号化により低周波数帯域成分を用いて行われ，また周波数領域符号化ではIGFにより他の周波数帯域の成分を用いて行われることが多い．そのため，入力信号の高周波数帯域成分と高周波数帯域成分の符号化の元となる周波数帯域の信号成分とに，時間方向におけるパワー分布に不整合が生じる場合や，高周波数帯域成分の符号化の際に時間方向のパワー分布に歪みが生じてしまう場合がある．そこで，エンコーダにおいて入力信号の高周波数帯域成分の時間方向のパワー分布が平坦か否かを検出して，その結果を送信することで，デコーダにおいて結果に基づき高周波数帯域成分を平坦化するなどの処理を行い，入力信号の時間方向のパワー分布との整合を確保できる．この平坦化処理は，フレームの符号化領域に応じて時間領域，周波数領域で行われる．時間領域符号化の際には，時間方向のパワー分布に

＊21 Transition Codingモード：前フレームの信号を用いた予測をできる限り排除して誤りの伝搬を抑制するように設計されたACELP（Algebraic CELP）の一符号化モード．エンコーダにおいて，入力信号の立ち上がりを含むフレームの次のフレームで選択されるように設計されている．

＊22 帯域分割フィルタ：入力信号を複数の周波数帯域に分割するデジタルフィルタ．

＊23 線形予測スペクトル：線形予測係数で決まるIIRフィルタの周波数スペクトル．

＊24 白色化処理：信号の電力の周波数分布が一様になるようにする処理．

おける急峻な立ち上がりをデコーダで実現するために，エンコーダでの立ち上がりの検出結果を送信してデコーダで立ち上がりを強調する処理を行う．**図7**に急峻な立ち上がりを強調する高周波数帯域成分改善技術の(b)適用前と(c)適用後の信号の周波数スペクトルの時間変化を表すスペクトログラムを示す．図7においては，縦軸が周波数，横軸が時間であり，色が黒から明るくなるにつれてパワーが大きくなることを示している．なお，入力信号である原音のスペクトログラムを(a)に示した．図7より，適用前では信号の急峻な立ち上がりの前に高周波数帯域に歪みが発生しているが，高周波数帯域成分改善技術適用後には歪みが抑制されていることが分かる．

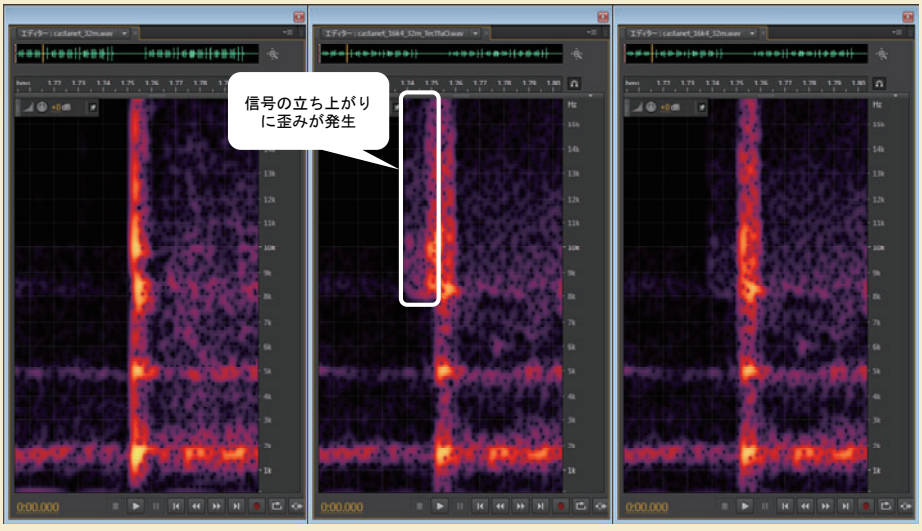
(2)パケットロス耐性改善技術

CELPでは，パケットロス時のピッチ長推定誤りが，回復時の不連続音の原因となる．CELPでは，線形予測係数を良好に算出するため，符号化対象フレームの次フレームの一部を先読み信号としてあらかじめ読み込んで利用する．この先読み部分の信号について算出したピッチ長を伝送することで，余分に遅延を増加させることなく，次フレームのピッチ情報を取得し，パケットロス時のピッチ長推定精度を改善した．

また，EVSでは線形予測係数をフレーム間予測により符号化しているため，特に音声開始部分でのパケットロス後の復帰フレームで線形予測フィルタが不安定となり，復号音に大きなリップルが生じる．そこで，線形予測フィルタが過剰なゲインを持たないようデコーダ側で線形予測フィルタを補正する．エンコーダ側でパケットロス時の線形予測フィルタをシミュレートして，フィルタが不安定になるフレームを検出し，検出結果に応じてデコーダ側で線形予測フィルタを補正することで，**図8**(a)に示したとおりリップルによる音質低下を防ぐことができる．提案手法の有効性をPESQ（Perceptual Evaluation of Speech Quality）*25 [9]により評価した．エラーパターンには，音声の開始区間がロスするものを作成して用いた．提案手法のありの場合と，なしの場合の評価スコアの差を図8(b)に示す．エラーバーは95％信頼度区間を示す．評価結果から，有意な音質改善が確認できる．

(3)演算量削減技術

低ビットレート―高ビットレート間で動作モードを切り替えるレートスイッチングの際，内部サンプリングレートが変化するため線形予測係



(a)原音　(b)適用前　(c)適用後

図7　高周波数帯域成分改善技術の効果

*25 **PESQ**：参照信号と被試験信号の差から音声品質を推定する客観的評価方法．

数を求め直す必要があるが，通常求めるのと同じ方法を用いると多くの演算が必要となる．そこで，線形予測スペクトルに対する周波数領域におけるリサンプリング*26により線形予測係数を求めることで，演算量を従来の3分の1程度に低減した．

3. EVSの性能

3.1 EVS選定試験

EVS選定試験において，各試験あたり32名の被験者が参加する合計24の主観評価試験[10]が3カ所の評価機関において行われた．超広帯域信号に対しては原音からの劣化を5段階で評価するDCR（Degradation Category Rating）試験*27を実施した．試験では，パケットロスなしの雑音なし音声，雑音付音声，音楽，およびパケットロス条件下の雑音なし音声を入力とした[11]．

3.2 評価結果

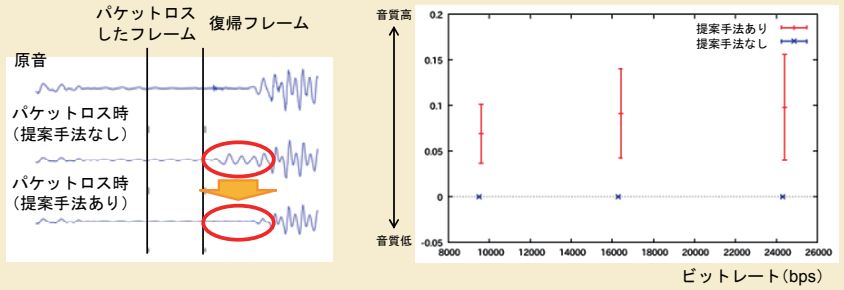
選定試験結果の抜粋を**図9**に示す．エラーバー*28は95％信頼度区間を示す．雑音なし音声での試験では，EVSはリファレンスコーデックG.722.1 Annex C*29 [12]の半分のビットレートで同等以上の音質を達成している．

実利用環境に近い雑音付音声に対する試験では，AMR-WBで利用可能な最高ビットレートである23.85kbpsよりもEVSの13.2kbpsの方が高音質を達成しておりAMR-WBからEVSへ移行した際の音質向上が有意に確認できる．同様に，パケットロス条件や音楽に対する性能向上も確認できる．AMR-WB＋*30は，80msの長い遅延を持つコーデックであるが，EVSは同等の品質を32msの低遅延で実現していることがわかる．

4. あとがき

本稿では，EVSについて，その概要と主要な技術を解説し，音質を評価する選定試験の結果を報告した．

EVSは，音声通話時の音質をFMラジオ並みに高めるうえに，VoLTEへの導入が容易であり，音声に加えて音楽も低ビットレートで高音質に符号化できる音声符号化方式である．これらの特長を活用することで，音声通話サービスやメロディコールのような音楽コンテンツを用いた既存サービスのさらなる高音質化が可能になるだけでなく，新しいモバイル音声コミュニケーションの誕生が期



(a)適用時の音声波形　(b)提案手法適用前後の評価スコアの差

図8　ドコモ提案パケットロス隠蔽技術の効果

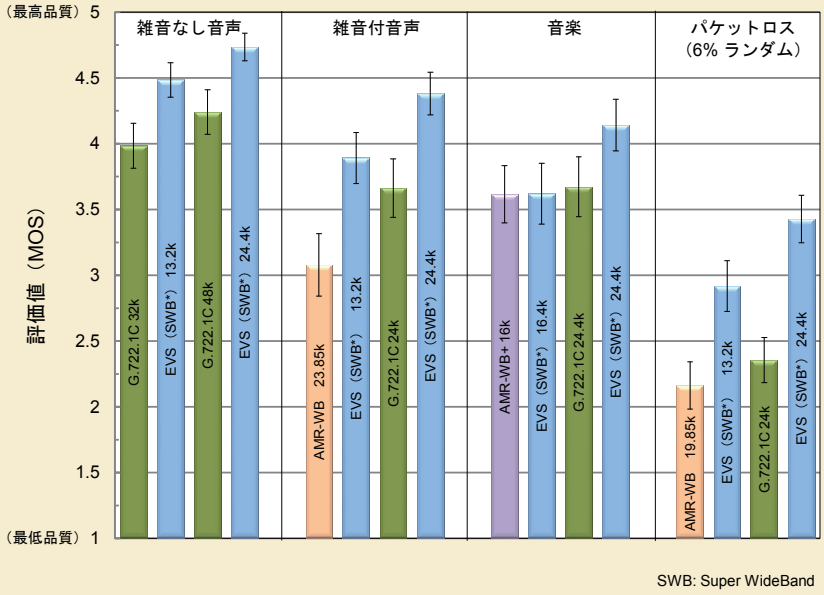


図9　主観評価試験結果

*26 **リサンプリング**：デジタル信号をアナログ信号に戻したうえで，別のサンプリング周波数を用いて再度サンプリングすること．

*27 **DCR試験**：評価対象信号が品質評価の基準となる参照信号に対してどの程度劣化しているかを尺度とする主観評価試験手法．参照信号および評価対象信号を受聴する．ITU-T P.800に規定されている．

*28 **エラーバー**：誤差範囲を示す棒線．

*29 **G.722.1 Annex C**：ITU-T標準の超広帯域対応音声コーデックであり，Polycom社の電話会議装置で用いられている．

*30 **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．

待できる．

文　献

[1] 3GPP TS26.441 V12.0.0: “Speech codec speech processing functions; Enhanced Voice Service (EVS) speech codec; General description,” 2014.

[2] 3GPP TS26.171 V12.0.0: “Speech codec speech processing functions; Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; General description,” 2014.

[3] ISO/IEC 23003-3: “Information technology – MPEG audio technologies – Part 3: Unified speech and audio coding,” 2012.

[4] 菊入，ほか：“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，”本誌，Vol.19, No.3, pp.18-23, Oct. 2011.

[5] 3GPP TR22.813 V10.0.0: “Study of Use Cases and Requirements for Enhanced Voice Codecs for the Evolved Packet System (EPS),” 2010.

[6] 3GPP TS26.071 V12.0.0: “Mandatory speech CODEC speech processing functions; AMR speech Codec; General description,” 2014.

[7] 3GPP TS26.447 V12.0.0: “Codec for Enhanced Voice Services (EVS); Error Concealment of Lost Packets,” 2014.

[8] 金子，ほか：“VoLTEに対応したメロディコールの高音質化，”本誌，Vol.22, No.4, pp.29-33, Jan. 2014.

[9] ITU-T P.862.2: “Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs,” 2007.

[10] ITU-T P.800: “Methods for subjective determination of transmission quality,” 1996.

[11] 3GPP AHEVS-311: “EVS Permanent Document EVS-8b: Test plans for selection phase including lab task specification,” 2014.

[12] ITU-T G.722.1: “Low complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss,” 2005.