



# 音声帯域拡張技術 ～ Voice Band Enhancer ～

青柳 弘美      田代 厚史

音声通信（いわゆる電話）サービスが、今大きく変わろうとしている。そのインフラが、NGN（Next Generation Network）やLTE（Long Term Evolution）などに代表されるようなIPネットワークをバックボーンとしたパケット通信網に統一されようとしている。これにより、今後さまざまな新しい電話サービスが生まれ出されていくことが予想されるとともに、最も基本となる音声の伝送そのものも大きく変わる。これまでの電話サービスのインフラは、システムの制約から音声として伝送できる周波数が300Hz～3.4kHzに制限されていた。一般に人間が聴きとれる周波数は20Hz～20kHzと言われており、その一部しか伝送されていなかったことになる。伝送網が変わることによりこの制限から開放され、より自然に近い音声伝送が可能となる。今後、これまでの高々4kHz程度の狭帯域音声サービスから、4kHzを超える音声までも伝送する広帯域音声サービスの普及が加速されていくだろう。本稿では、この広帯域音声サービス普及における課題と、それに対するOKIの取り組みについて紹介する。

たとえばITU-TではG.729.1<sup>1)</sup>、G.711.1<sup>2)</sup>などが近年制定された。送話側端末、受話側端末双方にこのような広帯域音声符号化機能を組み込み、新しい伝送網を介することにより、広帯域の音声サービスが実現される。

しかし、このような新しい電話サービスは一夜にして既存のものから移り変わるものではない。その移行期では、既存の狭帯域電話サービス網と新しい広帯域電話サービス網が共存する期間が存在し、しかも、初期は既存網の方が大多数を占め、やがて徐々に広帯域網へ移行していくものと考えられる。この移行期間中においては、広帯域サービス網での広帯域端末は十分な効力を発揮できない恐れがある。何故ならば、広帯域音声通信は双方の端末で広帯域音声符号化機能を有している必要があるからである。つまり、従来の狭帯域端末との通話では相手側が広帯域音声符号化機能を有していないため、これまでの狭帯域の音声での通話しかできないのである。広帯域電話サービスの初期においては、通話相手の大多数は従来の狭帯域端末であることが想定され、これでは広帯域端末を利用する価値が半減してしまう（図1）。

この課題に対処するため、OKIでは狭帯域音声信号を基に、その失われた高域の部分を擬似的に生成する帯域拡張機能の実用化を進めている。この機能を利用すると、通信自体は狭帯域音声であっても受話側の端末内部で広帯域の音声信号が得られるため、広帯域端末の能力が十分に発揮されることになる（図2）。以降では、この帯域拡張

## 広帯域音声サービス

広帯域の音声を伝送するためには、広帯域の音声を圧縮・伸張する音声符号化機能が必要となる。各標準化機関では精力的に広帯域音声符号化の標準化を進めており、

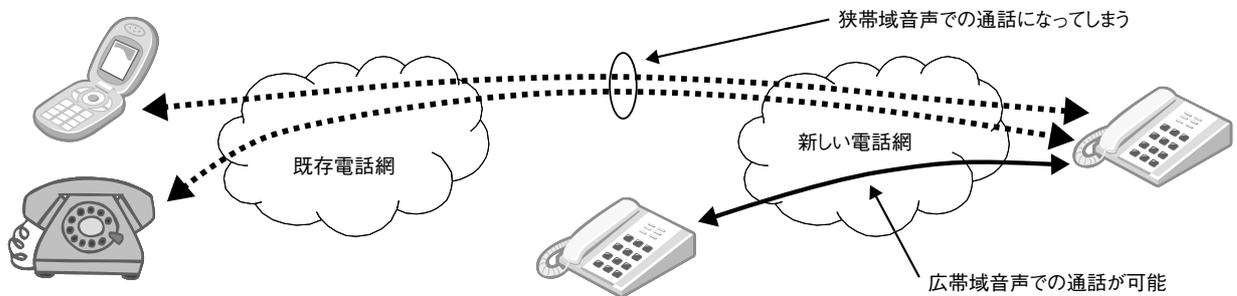


図1 サービス移行期における音声通信

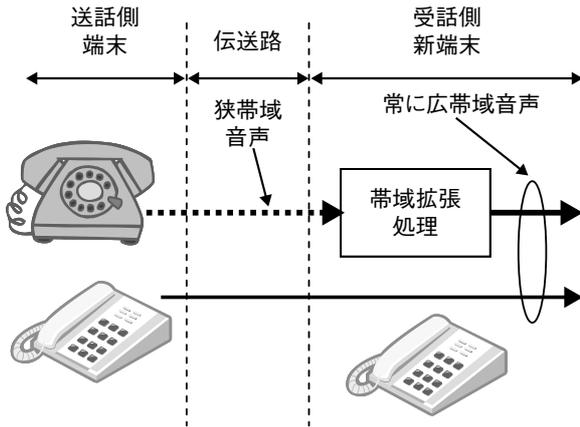


図2 帯域拡張機能

帯域拡張機能の概要と実用化に向けた取り組み、評価結果などについて記す。

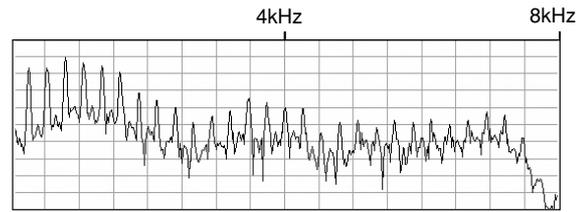
## 帯域拡張機能 (Voice Band Enhancer)

### (1) 拡張手法

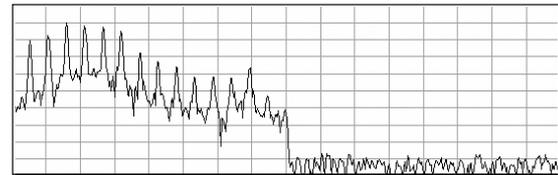
帯域拡張処理は、一般には大きく二つの手法が存在する。一つは、現存する低域（～4kHz）の信号を加工・整形して高域（4kHz～8kHz）に移動させる方法であり、もう一つは、低域信号を分析して残差信号を生成し、これを高域まで考慮して再合成する手法である。特に後者の手法においては、低域信号の周波数的概形と高域信号の周波数的概形を予め調査（学習）して対応付けしておき、これをコードブック（量子化テーブル）として持つ手法を併用する場合がある。これは、低域信号を分析することによりその概形情報を得て、その情報を基に対応する高域の概形情報をコードブックから取り出して合成に利用する、というものである。

これらの手法についてOKIではさまざまな面から調査を実施した。その結果から、前者の加工・整形型は概形の再現性が若干劣るものの、高域信号のノイズ感は比較的少なく抑えられることが分かった。一方、後者の分析・合成型は高域信号のノイズ感が若干増大するという傾向が見られるものの、概形の再現性は比較的良好となることが分かった。また、コードブックを利用する手法は、更に概形の再現性が良好となるが、その設計（学習のさせ方、コードブックの大きさ）には実装上の工夫が必要である。たとえば、入力信号に対する依存性（言語、伝送路特性など）や、メモリ使用量を考慮する必要がある。

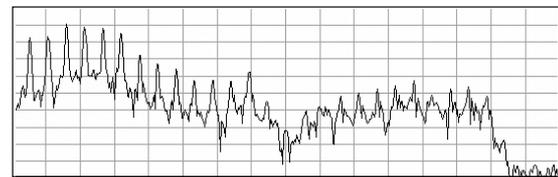
これらのことをふまえ、OKIでは、加工・整形型をベースにして、更に分析を実施して得られた情報を適応



(a) オリジナル広帯域音声



(b) 狭帯域音声



(c) 帯域拡張音声

図3 VBE処理音声の周波数特性

的に加工・整形に利用するという手法を実現し、この帯域拡張機能をVoice Band Enhancer (VBE) と称している。VBE処理音声の周波数特性の一例を図3に示す。

### (2) 入力信号特性

VBEは、その基本拡張性能も重要であるが、実用的なものとするにあたっては、大きく二つの点を考慮した。一つはVBEに対する入力信号の特性であり、もう一つはVBEの出力信号の特性である。

VBEに対する入力信号は、伝送路を介して遠方の端末から送られてくるものである。通信システム（固定網／携帯網／通信会社）が異なれば、使われている音声符号化方式や伝送路の特性も異なり、また、相手端末の使用状況により、背景の雑音の大きさなども異なる。その結果、VBEに入力される信号は実にさまざまな特性を有することになる（図4：次ページ）。帯域拡張の基本処理は、現存する低域（～4kHz）の信号を基に失われた高域（4kHz～8kHz）の周波数部分を生成するため、その拡張特性は、現存部分の信号の特性に少なからず左右されるという性質を本質的に持つ。安定的な拡張性能を発揮させるためには、拡張処理の前段に配するプレ処理が重要な役割を演じ、拡張特性を大きく左右すると言っても過

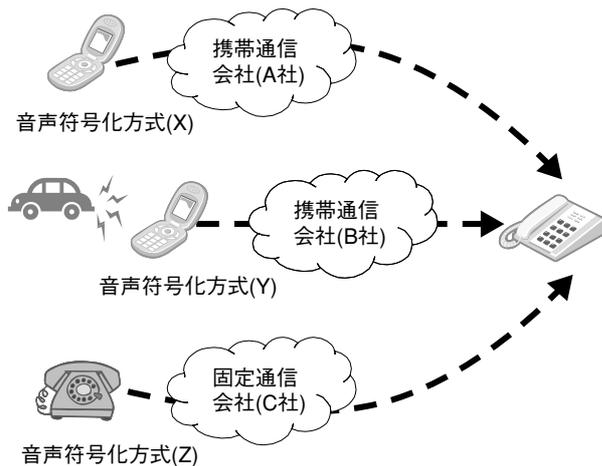


図4 入力信号の多様性

言ではない。OKIでは、さまざまな通信パターンを実測・想定し、プレ処理による補償方法にノウハウを蓄積してきた。

### (3) 出力信号特性

一方、VBEの出力信号は、最終的に電気・音響変換デバイス（いわゆるスピーカ）を経て、人間の耳に到達する。その直前のアナログ回路の特性も含め、スピーカから放出される音響の特性も実にさまざまである（図5）。

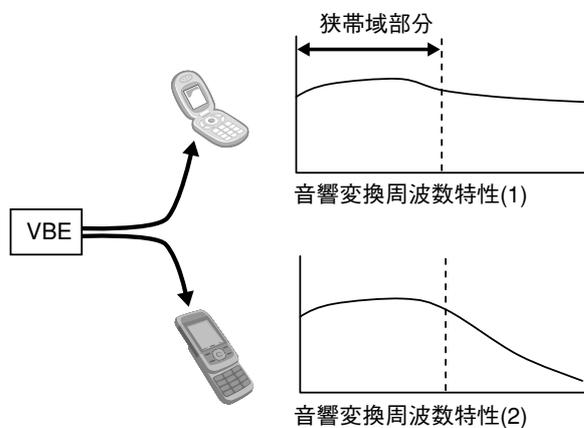


図5 音響変換特性の多様性

安定した音質を人間の耳に到達させるためには、これを補償するポスト処理もまた重要な役割を演じる。現存する電話端末に組み込まれているスピーカ（もしくはレシーバ）は、一般的に狭帯域の信号（～4kHz）を出力することを主眼としているため、広帯域の信号（4kHz～）に対してはその出力特性が保証されない。つまり、比較

的きれいに出力されるものもあれば、大きく減衰してしまうものもある。広帯域用に設計されたスピーカを利用すると比較的安定した特性を得られるが、まだまだ一般的ではなくコストも高い。OKIでは、一般的な狭帯域レシーバでも安定した拡張特性が得られるよう、ポスト処理による補償方法にノウハウを蓄積してきた。

余談ではあるが、音声通信端末を製造する会社では、「音のマイスター」、「音の神様」などと呼ばれる人々が必ずと言って良いほど存在する。彼らは、製品の最終的な音質を調整・決定する人々であり、彼らの音作りを容易にするためのインターフェースを準備・工夫することも実用化の上で重要なポイントである。

## 主観評価

ここでは、VBEの基本となる帯域拡張処理の性能について、その評価結果を紹介する。帯域拡張音声に対し、人が感じる主観的な音声品質を評価するため、ITU-T勧告P.800.1<sup>3)</sup>に従ったオピニオン評価を第三者機関にて実施した。評価音作成系と試験系を図6に、試験条件を表1に示す。

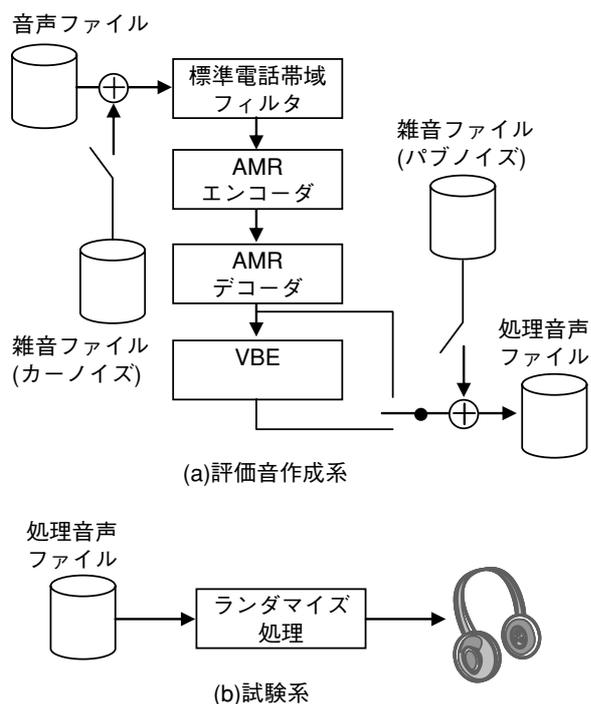


図6 評価音作成系と試験系

評価音声作成系は計算機上で動作し、用意された音声ファイル（2センテンス/ファイル、雑音重畳無し）に対して、標準電話帯域フィルタ、AMR<sup>4)</sup>エンコーダ、AMR

表1 試験条件

音声サンプル	英語、男女各5センテンスペア
音声条件	① クリーン音声（雑音重畳なし） ② 受話側雑音重畳（パブノイズ、S/N=20dB程度） ③ 送話側雑音重畳（カーノイズ、S/N=15dB）
音声処理種類	AMR、AMR+VBE
被験者	10名（英語ネイティブな一般人）

デコーダ、VBE処理（ON/OFF）を順に施し、その結果をファイルに書き出す。送話側雑音重畳は、標準電話帯域フィルタ処理の前段で、用意された雑音ファイルを加算することにより実施する。送話側に重畳する雑音は、カーノイズ（走行中の車の中で録音した雑音）を使用し、S/Nは15dBとした。受話側雑音重畳は、VBE処理の後段で、用意された雑音ファイルを加算することにより実施する。受話側に重畳する雑音は、パブノイズ（飲食店の中で録音した雑音）を使用した。受話側雑音は直接耳に聴こえるものであるため、20kHz程度までの周波数成分を含み、かつステレオフォーマット（バイノーラル）とした。この場合、目的音は片耳から、雑音は両耳から聴こえることとなるため、S/Nの規定は難しく、片耳（目的音+雑音）でのS/Nが20dBとなるように設定した。

試験系は、録音された処理音声をランダムに再生してヘッドフォンで受聴（両耳）する。受話側雑音重畳の条件以外は片方のみ音が再生され、どちらの耳で受聴するかは被験者の好みにより選択される。被験者はファイルごとに品質レベルを5段階（評価が高い方から5点、4点、3点、2点、1点）で評価し、全ての被験者の評点の平均値（MOS値）を算出することにより最終的な評価結果を得る。結果を図7に示す。

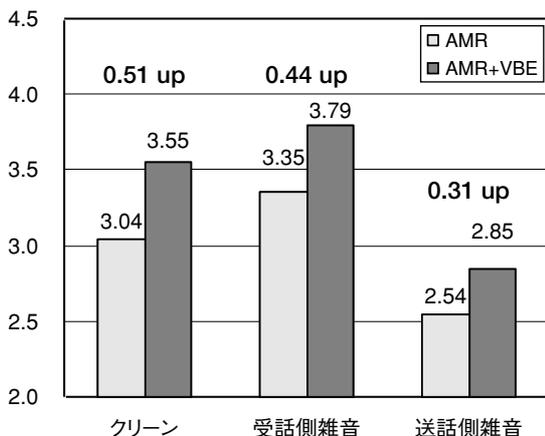


図7 評価結果

図7から、クリーンな音声に対しては0.5ポイントのMOS値上昇が見られ、大きな効果があることが分かる。また、受話側、送話側の背景に雑音がある場合においても、VBEによる効果が見られることが分かる。これらのことから、本VBEを使用することにより、従来の狭帯域音声よりも音質が向上した拡張音声を得られると言える。

## まとめ

本稿では、広帯域音声サービス普及における課題と、それを解決するためのOKIの取り組みについて紹介した。今後も継続した方式改良により、その実用性を高めていく。現時点では、オリジナルの広帯域音声と比べると、その音質は今一歩劣っている。しかし、解決すべき課題は既に明確になっており、課題解決に向けた検討を鋭意進めている。近い将来、オリジナルと遜色の無いレベルの拡張音声を得られることが期待される。広帯域音声サービス普及を加速するために、OKIのVBEがその一助となれば幸いである。◆◆

## 参考文献

- 1) ITU-T Recommendation G.729.1 (2006) : G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729
- 2) ITU-T Recommendation G.711.1 (2008) : Wideband embedded extension for G.711 pulse code modulation
- 3) ITU-T Recommendation P.800.1 (2006) : Mean Opinion Score (MOS) terminology
- 4) 3GPP TS 26.071 V6.0.0 (2004) : Mandatory speech CODEC speech processing functions; AMR speech CODEC; General description (Release 6)

## ● 筆者紹介

青柳弘美 : Hiromi Aoyagi. 通信システム事業部 eおと事業推進部 技術統括チーム  
田代厚史 : Atsushi Tashiro. 通信システム事業部 eおと事業推進部 技術統括チーム