

広帯域感を高めた音声帯域拡張技術

藤枝 大 矢頭 隆

近年、NGN (Next Generation Network) やLTE (Long Term Evolution) などに代表されるようなIPネットワークの発展に伴い、従来より高音質な音声通信である広帯域音声サービスが普及し始めている¹⁾。広帯域の音声を伝送するための符号化技術の標準化も精力的に進められており、今後は広帯域音声サービスの普及が加速すると見込まれている。

しかし、広帯域音声サービスへの移行期では、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る。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なければならないため、従来の狭帯域端末と広帯域端末との間では狭帯域の音声での通話しかできない(図1)。

この問題に対処するため、OKIでは狭帯域音声を基に、疑似的に広帯域音声を生成する帯域拡張機能VBE (Voice Band Enhancer) の研究開発を継続して進めてきた。今回、新たな要素技術を開発・導入することによって、より高音質な拡張音声を得ることができた。本稿では、その技術を紹介し、実験による評価結果を報告する。

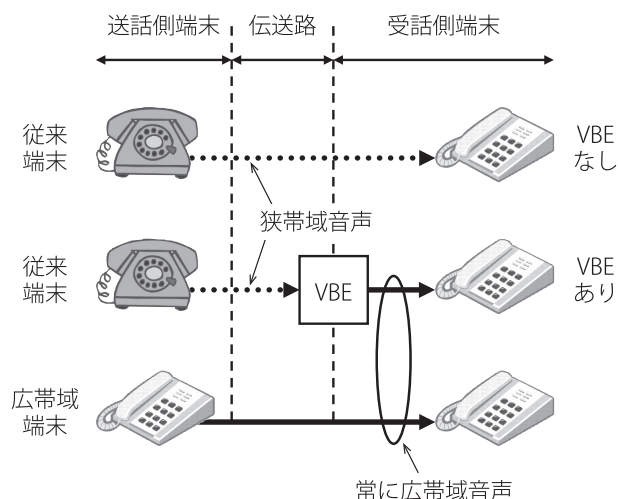


図1 広帯域音声サービスとVBE

帯域拡張機能 (Voice Band Enhancer)

帯域拡張処理は、一般には二つの手法に大別される。一つは、狭帯域音声 (~4kHz) を分析して音源信号と声道情報に分け、それぞれ広帯域 (~8kHz) まで成分・情報を有するように拡張し、再び合成して拡張音声を得る分析合成型の方法である。この方法では、狭帯域音声と広帯域音声の周波数的概形を予め調査 (学習) して対応付けておいて、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ことが多い。もう一つは、アップサンプリングされた狭帯域音声を加工して高域に移動させ、さらに低域の情報に基づいて高域を整形し、狭帯域音声と加算合成して拡張音声を得る加工整形型の方法である。

OKIのVBEは、これらの手法のうち、高域信号のノイズ感が比較的少なく抑えられる特徴を持つ加工整形型の帯域拡張方法をベースとしている。このVBEの処理の流れを図2を用いて説明する。

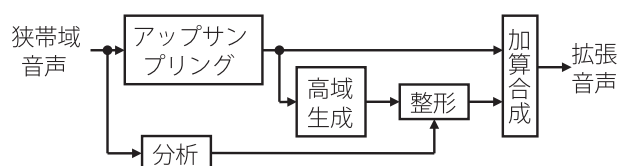


図2 加工・整形型の帯域拡張方法

まず8kHzサンプリングの狭帯域音声が入力された後、処理は実際に帯域を拡張する処理 (上側) と、分析・推定によって高域成分を調整する処理 (下側) とに分かれる。拡張側では、サンプリング周波数を8kHzから16kHzへと変更するアップサンプリング処理が施される。得られた信号をベースバンド信号と呼ぶ。ベースバンド信号は、低域は狭帯域音声と同じ成分を持つが、高域は成分を何も持たない。高域生成処理では、このベースバンド信号に特殊な加工を施すことで高域に成分を有する信号を生成し、HPFで4kHz以上を抽出する。得られた信号を拡張高域信号と呼ぶ。特殊な加工には、例えば低域成分を高域にシフトする処理や低域成分を

4kHzで折り返す処理、全波整流のような非線形な処理が用いられる。なお、このような高域生成処理だけによって得られた拡張高域信号をベースバンド信号に付加するだけでは、高域成分が強すぎたり弱すぎたりといった不自然な拡張音声になってしまうので、振幅やスペクトル包絡の調整が必要となる。これを担うのが分析・推定側の処理であり、狭帯域音声を分析して音響特性を抽出し、これに基づいて高域の情報を推定する。この推定によって得られた情報を先の拡張高域信号に反映させることで、自然性の高い拡張音声を得ることができる。

新方式の特徴

本章では我々の提案するVBEの特徴を説明する(図3)。新方式の大きな特徴は、次の4つである：①有声音と無声音を別々に拡張している、②新しい音響特徴量「ModGI」を考案・使用している、③高域の振幅を推定している、④無声音のスペクトル包絡を推定している。

なお、①はすでに同じような発想から異なる方式が提案されており、②～④の3つの特徴がOKI独自の技術である。①を行う理由は、音声の生成過程が大きく異なるためであり、拡張音の生成と高域情報の推定のどちらにおいても有音/無音を分けることで、帯域拡張処理の安定性と拡張音声の自然性を向上させることができる。②～④の特徴については、以下で詳しく解説する。

(a) 音響特性分析の特徴量「ModGI」(②の特徴)

新しい音響特徴量「ModGI」は、Modified Gradient Indexの意である。Gradient Index (GI) は音声の無音らしさを表すパラメータとして帯域拡張の分野でしばしば使われている²⁾。多くの場合、GIは有音無音判定に使われるが、我々は無声音の拡張高域成分の振幅推定に用いる。後述するように、GIは「値が小刻みに変動しやすい」という欠点を持つため、これによって推定した

振幅値を用いると新たな雑音が発生してしまう。ModGIはこの欠点を改善したものとなっている。

従来の音響特徴量GIは、次式で定義される。

$$GI(n) = \frac{\sum_{m=n+1}^{n+N-1} \Psi(m) |s(m) - s(m-1)|}{\sqrt{1/N \cdot \sum_{m=n}^{n+N-1} s^2(m)}} \quad (1)$$

ここで、 n , m は時間のインデックス、 N は特徴量算出のための区間長、 $s(n)$ は分析対象の信号であり、

$$\Psi(m) = \frac{1}{2} |\psi(m) - \psi(m-1)| \quad (2)$$

$$\psi(m) = \text{sgn}\{s(m) - s(m-1)\} \quad (3)$$

である。式中の $\Psi(m)$ は傾きが変化したかどうかを0または1で表しているの、GIは「傾きの変化の度合い」を表すパラメータであると言える。音声波形において、無声音は有声音に比べて信号値(音圧)の上下が激しいことから、無声音ではGIが大きくなる。しかし、GIは $\Psi(m)$ のような離散的な系列を含んでいることから、値が小刻みに変動しやすい。

次に、新しい特徴量ModGIについて説明する。小刻みな変動を減らすためには、離散的な値を使わないようにすればよい。そこで、次式のようにModGIを定義する。

$$\text{ModGI}(n) = \frac{\sum_{m=n+1}^{n+N-1} |s(m) - 2s(m-1) + s(m-2)|}{\sqrt{\sum_{m=n}^{n+N-1} s^2(m)}} \quad (4)$$

式(4)の分子は傾きの変化量を意味するので、小刻みな変動を抑えながらModGIにGIと似た性質を持たせることができる。

GIとModGIの算出例を図4に示す。図4において、(a)は分析対象の信号、(b)はGI、(c)はModGIの算出結果を表している。図より、ModGIがGIに比べて小刻みな変動が改善されていることがわかる。

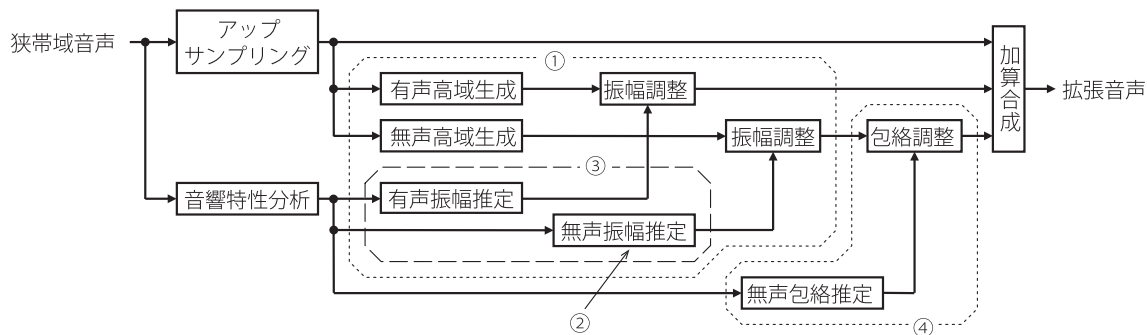


図3 OKI-VBEの処理の流れ

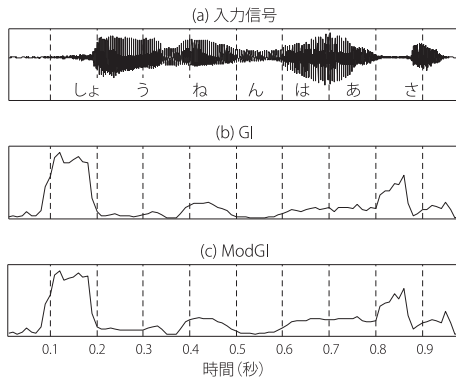


図4 GIとModGIの比較

(b) 拡張高域成分の振幅推定 (③の特徴)

従来は拡張高域成分の振幅（以下、高域振幅と呼ぶ）を調整するために、「有声音では高域を弱め、無声音では強める」などというような拡張高域成分を抑制／強化するというアプローチをとっていた。抑制／強化するための拡張ゲインは、低域成分から適応的に算出される。このようなアプローチは、最終的な高域振幅は振幅調整前の拡張高域成分の振幅に依存するため、最適な高域振幅を与える拡張ゲインを算出するのは容易ではなかった。

そこで、我々は低域成分から高域振幅を推定する方式を考案した。拡張ゲインは、高域振幅推定値を振幅調整前の高域振幅で割ることによって得られる。これにより、拡張高域成分の生成方法に依存しない高域振幅の調整が可能となる。

前述したとおり、高域振幅の振る舞いは有声音と無声音で異なるので、別々の方法で推定する。有声音の高域振幅は、主にベースバンド信号のスペクトルの傾きを高域にも反映させるようにして推定する。一方、無声音の高域振幅は、前述した音響特徴量ModGIに非線形変換を施すことで推定する。さらに、無声音の高域振幅が大きくなるのはベースバンド信号の振幅が小さいときだけなので、この性質を利用することで、従来は調整が困難だった無声音の高域振幅を適切に推定することができるようになる。

(c) 無声音のスペクトル包絡推定 (④の特徴)

帯域拡張技術において、これまで無声音のスペクトル包絡についての言及は皆無であった。そこで、我々は無声音のスペクトル包絡を整形する独自方式を考案し、帯域拡張によって無声音の音韻が変質してしまう問題を解決した。今回は、持続時間が長いために帯域拡張による音韻の変質が知覚されやすい2つの無声音「歯摩擦音」と「硬口蓋摩擦音」のスペクトル包絡を整形する方法を考案した。2つの無声摩擦音のスペクトル包絡の例を、図5に示す。

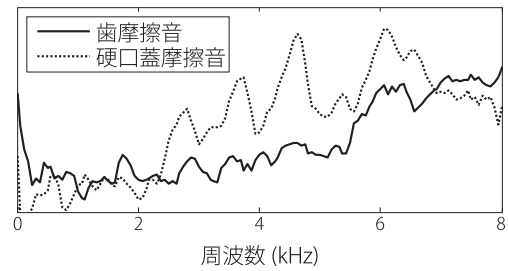


図5 無声摩擦音のスペクトル

図5を見ると、どちらも周波数が高くなるほどパワーが大きくなっているが、硬口蓋摩擦音の方が低い周波数から大きくなっていることがわかる。例えば「～です。」で終わるフレーズを拡張した際、誤って最後の無声音の4kHz～5kHzが強く拡張されると、拡張音は「～でし。」と聴こえてしまう。そこで、4kHz～5kHzのパワーを適切に与えることで、このような無声音の変質の回避に成功した。

評価

以上で説明したVBEの性能について、主観評価を行った結果を紹介する。帯域拡張音声に対し、人が感じる主観的な拡張度合いを評価するために、対比較評価を社内にて実施した。試験条件を表1に示す。

表1 試験条件

音声サンプル	日本語、男性女性各2名、各5センテンス
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）
音声処理種類	ダウンサンプリング、旧VBE、新VBE
被験者	5名（社内）

評価音の作成は計算機上でを行い、用意された音声ファイル（1センテンス／ファイル、雑音重畳無し）に対して、ダウンサンプリング、OKIテクニカルレビュー第215号¹⁾時点での方式（旧VBEと呼ぶ）、本稿で紹介した方式（新VBEと呼ぶ）を施し、その結果および無処理の音声をファイルに書き出す。これら目的音は、受聴時にはヘッドフォンの片方のみ音が再生される。どちらの耳で受聴するかは被験者の好みによって選択できるようにした。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。

試験は、無処理の参照音、参照音にダウンサンプリング処理を施した劣化音、劣化音を旧VBEで拡張した旧VBE拡張音、劣化音を新VBEで拡張した新VBE拡張音の4つを自由に聴き比べられる状態で行う。受聴後、旧VBEと新VBEのうち拡張感（「帯域が拡張された」「高域成分が感じられる」という印象）の高い方を選択してもらい、それぞれが選択された比率を算出して評価値とする。ただし、拡張感が同等で優劣がつけられなかった場合には、両者の選択回数を+0.5とした。このようにして得られた評価結果を図6に示す。また、音声ファイルの一つである男性がセンテンス「ゲームをして遊ぶ」を発話したときの参照音、新旧VBE拡張音のスペクトログラムを図7に示す。

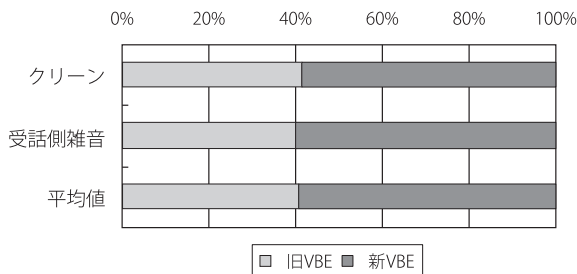


図6 拡張感の主観評価の結果

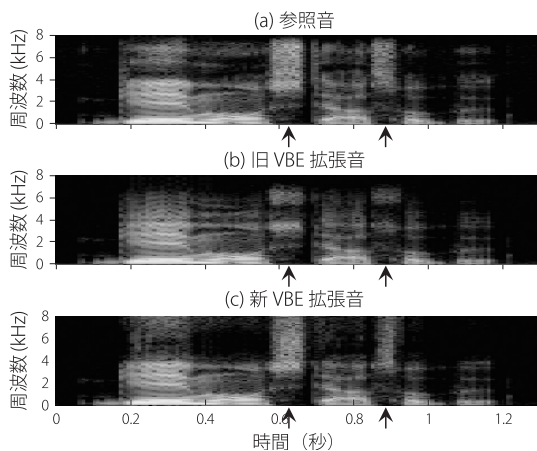


図7 拡張音のスペクトログラム

図6から、雑音の有無に関わらず、旧VBEに比べて新VBEの方が20%評価値が高く、拡張感が向上したことがわかる。次に、図7の無声摩擦音のスペクトル包絡（矢印の時刻）に注目する。0.6秒と0.9秒あたりに、それぞれ「して遊ぶ」の"shi"と"so"がある。この2か所に注目

すると、旧VBEに比べて、新VBEはスペクトルの再現性が高く、無声音スペクトル包絡推定がうまく動作していることがわかる。

まとめ

本稿では、以下の3つの特徴を有する新しいVBEを提案した。特徴の1つ目として、音響特徴量としてGIの特性を改善したModGIを用いることで、雑音感を軽減させた。2つ目として、従来のように拡張ゲインを直接算出するのではなく、高域成分の振幅を推定することで、自然性と拡張感の高い適切な拡張ゲインを得ることができた。3つ目として、無声音のスペクトル包絡を推定して整形することで、拡張によって無声摩擦音の音韻が変質してしまう問題を解決した。これらの新しい技術を取り入れることによって、OKIの旧方式よりも拡張感が向上したことが示された。今回は主に無声音部分の改善効果を得たが、今後は有声音部分の再現性向上も図っていきたい。◆◆

参考文献

- 1) 青柳弘美、田代厚史：音声帯域拡張技術～Voice Band Enhancer～、OKIテクニカルレビュー、第215号、Vol.76、No.2、pp.54-57、2009年10月
- 2) B. Iser, W. Minker, G. Schmidt: Bandwidth Extension of Speech Signals, 2008 ed., pp.36-38, 2008, Springer

筆者紹介

藤枝 大：Masaru Fujieda. 研究開発センタ メディア処理技術研究開発部
 矢頭 隆：Takashi Yazu. 研究開発センタ メディア処理技術研究開発部

TIPS 【用語解説】

歯摩擦音

舌先を歯茎に近づけて発声する摩擦音。日本語（ヘボン式ローマ字）では"s"で書かれる子音で、英語ではseeやsitの語頭の発音に相当する。

硬口蓋摩擦音

前舌を硬口蓋（口腔の上前部）に近づけて発声する摩擦音。日本語では"sh"で書かれる子音で、英語では例えばsheやshipの語頭の発音に相当する。

対比較評価

実験実施者が複数の評価対象の中から2つを一組として被験者に提示して、被験者は提示された2つの評価対象のうち好ましい方を選択する、という手続きを繰り返す評価方法である。