

CELPに基づく音声符号化向けのピッチ周波数に依存した適応ポストフィルタ

千葉 大将[†] 鎌本 優^{††} 守谷 健弘^{††} 原田 登^{††}
宮部 滋樹[†] 山田 武志[†] 牧野 昭二[†]

Adaptive Post-Filtering Method Controlled by Pitch Frequency for CELP-Based Speech Coding

Hironobu CHIBA[†], Yutaka KAMAMOTO^{††}, Takehiro MORIYA^{††}, Noboru HARADA^{††},
Shigeki MIYABE[†], Takeshi YAMADA[†], and Shoji MAKINO[†]

あらまし 携帯電話や VoIP (Voice over Internet Protocol) などのように低ビットレートでの通信が要求される条件下では CELP アルゴリズムをもとにした音声符号化方式が使われることが多い。これらの方式では復号信号に対してピッチやフォルマントを強調するポストフィルタ処理を行うことで聴感上の品質を向上させており、8 kbit/s 程度の低ビットレートでの広帯域音声通信をサポートしている ITU-T G.718 ACELP 方式では、復号信号に対して低い周波数帯域のみピッチ強調する Bass post-filter を用いて音声品質を向上させている。本研究では、ピッチ周期情報を用いて、Bass post-filter のピッチ強調を行う周波数帯域とピッチ強調利得を処理フレームごとに適応的に変化させる手法を提案する。提案法の音質を評価するために、広帯域用 PESQ を用いた客観評価と MUSHRA による主観評価を行った結果、クリーン音声の音質を向上させることが確認できた。従来法と比較すると提案法の演算量・記憶容量の増加はごくわずかであり、CELP に基づいた多くの音声符号化方式にも簡単に適用することが可能である。

キーワード 音声符号化, G.718, CELP, ポストフィルタ, ピッチ強調

1. ま え が き

近年、伝送帯域の改善により高精細映像コンテンツも流通することができるようになってきている [1]。一方、音声通信のように低遅延が要求され、優先制御が必要な場合はビットレートを低く抑えることが望まれている。そのような状況下においても音声品質の改善が求められるが、ビットレートや演算量とのトレードオフとなるため、より効率的な音質改善処理が必要となる。その方法の一つとして、復号信号のピッチ周期

やスペクトルを解析しコムフィルタやフォルマント強調処理をデコーダ側で行うポストフィルタによる音質改善処理が研究されてきた [2]~[6]。歴史的に、携帯電話や VoIP (Voice over Internet Protocol) を代表とする低ビットレートでの音声通信において、音声の生成過程をモデル化することで効率的に情報圧縮を行う CELP (Code-Excited Linear Prediction) に基づいた音声符号化アルゴリズムが使われており、その中でも ACELP (Algebraic CELP) と呼ばれる、より情報圧縮の効率性が高い音声符号化アルゴリズムが広く用いられている [7]。CELP に基づく音声符号化アルゴリズムでは、復号信号の聴感的な品質を向上させるために伝送された情報を効率的に利用したデコーダ側でのポストフィルタ処理を行っている [8]~[12]。この具体例として、ピッチやフォルマントの強調処理が挙げられる [13]~[15]。従来のポストフィルタ処理では、符号化方式のサンプリング周波数が 8 kHz と狭いため、ナイキスト周波数周辺までポストフィルタ処理を

[†] 筑波大学大学院システム情報工学研究科, つくば市
Graduate School of Systems and Information Engineering,
University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, 305-8573 Japan

^{††} 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所,
厚木市

NTT Communication Science Laboratories, Nippon
Telegraph and Telephone Corporation, 3-1 Morinosato-
Wakayama, Atsugi-shi, 243-0198 Japan

DOI:10.14923/transinfj.2015JDP7004

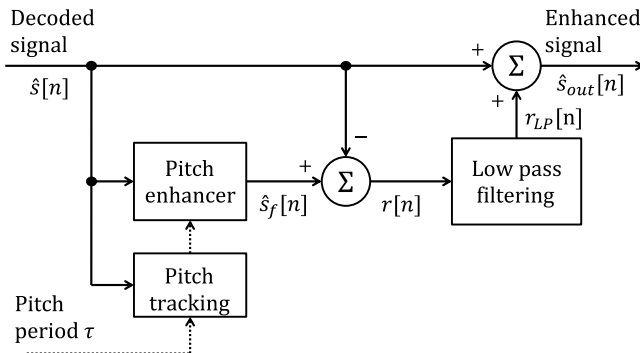


図1 G.718 Bass post-filter のブロック図
Fig. 1 Block diagram of bass post-filter used in G.718.

行っていたが、広帯域音声符号化をサポートしている ITU-T G.718 ACELP 方式では、復号信号に対して低周波数帯域のみピッチ強調する Bass post-filter を採用している。このポストフィルタは、ある固定の周波数帯域でピッチ強調処理を行い、聴感的な音質を向上させている [16], [17]。

本研究では、G.718 の Bass post-filter において、CELP での伝送情報であるピッチ周期を用いて、Bass post-filter のピッチ強調を行う周波数帯域幅とピッチ強調の利得を処理フレームごとに適応的に変化させる手法を提案する。従来の G.718 では、ピッチ強調用の信号を、ある固定のカットオフ周波数の低域通過フィルタに通すことで特定の周波数帯域のみピッチ強調を行うが、ピッチ周期によって音質向上に最適なカットオフ周波数は異なると考えられる。また、強調時の利得は復号信号の周期性に依存して算出されているが、我々はピッチ周期によって音質に適した利得の上限値があることを発見した [18]。よって、本論文では Bass post-filter の低域通過フィルタのカットオフ周波数と利得上限値を処理フレームごとに適応制御する手法について論じる。また、客観評価及び主観評価を行い、どのような音声に対して提案法による音質向上効果が高いのか分析する。

本論文の構成を以下に示す。2. では、G.718 に実装されている、従来の Bass post-filter の処理の概要について述べる。次に 3. では、本研究における提案法の実装を行うために、音声のピッチ周期に該当するピッチ周波数（基本周波数、 F_0 ）と固定のカットオフ周波数及び利得上限値の関係に関する予備実験について述べる。この予備実験より、基本周波数に対して適したカットオフ周波数や利得上限値を算出する回

帰直線を導出し、適応的制御の実装方法について記述する。4. では、提案法による音質改善の評価として、ITU-T P.862 PESQ (Perceptual Evaluation of Speech Quality) [19] を用いた客観評価実験結果を示す。この客観評価で提案法の有効性を確認し、ITU-R BS. 1534 MUSHRA (Multi Stimulus test with Hidden Reference and Anchor) [20] による主観評価実験結果も示す。最後に 5. では、本論文をまとめる。

2. G.718 における Bass post-filter の処理概要

G.718 の ACELP 復号後に用いられる Bass post-filter の処理を図 1 に示す。この処理は、復号信号 $\hat{s}[n]$ に対して 10 ms の処理フレームごとに行う。ただし、G.718 では 20 ms ごとに復号化しパケット伝送を行うため、パケット伝送フレーム長 L より先の復号信号 $\hat{s}[n]$ は、

$$\hat{s}[n+L] = \hat{s}[n+L-\tau] \quad (1)$$

としてピッチ周期 τ で外挿し補う。

Bass post-filter のピッチ強調処理に用いるピッチ周期 τ としては、不連続なピッチ周期でのピッチ強調を防ぐために、図 1 で表される Pitch tracking 処理により復号処理で使用されたピッチ周期 $\hat{\tau}$ を補正して使用する。このピッチ周期補正では、復号信号と、復号信号より $\lfloor \frac{\hat{\tau}}{2} \rfloor$ サンプル早い信号の正規化された内積 c_n を

$$c_n = \frac{\sum_{n=0}^{N_{pt}-1} \hat{s}[n] \hat{s}[n - \lfloor \frac{\hat{\tau}}{2} \rfloor]}{v_1 v_2} \quad (2)$$

$$v_1 = \sqrt{\sum_{n=0}^{N_{pt}-1} \hat{s}[n - \lfloor \frac{\hat{\tau}}{2} \rfloor] \hat{s}[n - \lfloor \frac{\hat{\tau}}{2} \rfloor]} \quad (3)$$

$$v_2 = \sqrt{\sum_{n=0}^{N_{pt}-1} \hat{s}[n]\hat{s}[n]} \quad (4)$$

として算出し、 c_n の値によって

$$\tau = \begin{cases} \frac{\hat{\tau}}{2} & c_n > 0.95 \\ \hat{\tau} & otherwise \end{cases} \quad (5)$$

として、 $\frac{\hat{\tau}}{2}$ 周期の周期性が高い処理フレームにおいて、ピッチ強調処理に用いるピッチ周期 τ を修正する。ここで、 N_{pt} は正規化された内積 c_n の算出に用いる復号信号の長さであり、G.718 では $N_{pt} = 280$ が使用されている。また、 $\lfloor \cdot \rfloor$ は床関数を表している。なお、ピッチ周期 τ は真の基本周波数 F_0 の値と異なることもあり得るが、ポストフィルタの性能上の問題は小さい。また、ピッチ周期推定を間違いやすい入力ではピッチ相関が低下して自動的にポストフィルタの影響は小さくなる。

ピッチ強調処理部では、まず復号信号 $\hat{s}[n]$ から前後 τ サンプル離れた信号を用いて生成された長期予測信号 $s_p[n]$ を、

$$s_p[n] = 0.5\hat{s}[n - \tau] + 0.5\hat{s}[n + \tau] \quad (6)$$

という処理により生成する。次に、 $\hat{s}[n]$ 、 $s_p[n]$ より復号信号を全帯域でピッチ強調した信号 $\hat{s}_f[n]$ を生成し、更に $\hat{s}[n]$ と $\hat{s}_f[n]$ よりピッチ強調用信号 $r[n]$ を、

$$r[n] = \hat{s}_f[n] - \hat{s}[n] \quad (7)$$

として生成する。ここで、

$$\hat{s}_f[n] = (1 - \alpha)\hat{s}[n] + \alpha s_p[n] \quad (8)$$

であり、 $\hat{s}_f[n]$ は

$$\alpha = \frac{C_p}{0.5(E_p + 10^{0.1\bar{E}_{pp}})} \quad (9)$$

として処理フレームごとに適応的に算出されるピッチ強調利得 α によって制御される。ここで、 C_p は $\hat{s}[n]$ と $s_p[n]$ の内積であり、 E_p は長期予測信号 $s_p[n]$ のエネルギーである。また、 $\bar{E}_{pp}$ は長期予測信号 $s_p[n]$ の直前の処理フレームまでの平均予測誤差のデシベル領域でのエネルギーである。したがって、無音区間や雑音区間ではピッチ強調の利得である α は小さい値が算出される。なお、 α は式 (9) より算出後、

$$\alpha = \begin{cases} 0.5 & (\alpha > 0.5) \\ 0.0 & (\alpha < 0.0) \end{cases} \quad (10)$$

として上限値・下限値で制限される。式 (9) における $\hat{s}$ と $s_p[n]$ の内積 C_p は、

$$C_p = \sum_{n=0}^{N-1} \hat{s}[n]s_p[n] \quad (11)$$

であり、長期予測信号 $s_p[n]$ のエネルギー E_p は

$$E_p = \sum_{n=0}^{N-1} s_p[n]s_p[n] \quad (12)$$

として求められる。ここで、 N はピッチ強調に用いる信号長である。式 (9) の $\bar{E}_{pp}$ は長期予測信号 $s_p[n]$ の直前の処理フレームまでの平均予測誤差のエネルギーを表している。 $\bar{E}_{pp}$ は直前の処理フレームで長期予測誤差 $e_p[n]$ と過去の $\bar{E}_{pp}$ から適応的に求められる。Bass post-filter における長期予測誤差 $e_p[n]$ は、

$$e_p[n] = \hat{s}[n] - \frac{C_p}{E_p} s_p[n] \quad (13)$$

として求められるが、低周波数帯域におけるピッチ調波の谷を抑圧するため、

$$e_{pp}[n] = e_p[n] + 0.9e_{pp}[n - 1] \quad (14)$$

としてディエンファシスされた長期予測誤差 $e_{pp}[n]$ が求められる。したがって、現在の処理フレームの予測誤差エネルギー E_{pp} は、

$$E_{pp} = 10 \log \sum_{n=0}^{N-1} e_{pp}[n]e_{pp}[n] \quad (15)$$

として求められ、現在の処理フレームまでの平均予測誤差エネルギーが、

$$\bar{E}_{pp} = 0.99\bar{E}_{pp} + 0.01E_{pp} \quad (16)$$

として計算される。

以上の処理により算出されたピッチ強調用信号 $r[n]$ は、カットオフ周波数が 500 Hz である対称的な 32 次 FIR フィルタで低域通過フィルタリングされ、

$$r_{LP}[n] = b[0]r[n] + \sum_{k=1}^{16} b[k](r[n - k] + r[n + k]) \quad (17)$$

として低域周波数帯域のみのピッチ強調用信号 $r_{LP}[n]$ を求めるのに用いられる。ここで、 $b[k]$ は低域通過フィルタのインパルス応答である。最終的に、 $r_{LP}[n]$ を

用いて低域周波数帯域のみピッチ強調された復号信号 $\hat{s}_{out}[n]$ を,

$$\hat{s}_{out}[n] = \hat{s}[n] + r_{LP}[n] \quad (18)$$

として生成し、聴感的な音質を向上させている。ここまですべて従来用いられてきた方法である。

3. Bass post-filter のピッチ強調帯域及び利得上限値の可変化

3.1 ピッチ周期に応じたパラメタの適応的制御

時変フィルタとして現在の処理フレームのピッチ周期を用いて復号信号のピッチ構造を強調する Bass post-filter は復号信号の高調波間の雑音を低減することができる。しかし、従来の Bass post-filter が適用される周波数帯域は直接ピッチ周期に依存していない。一般的に、CELP によって圧縮された信号において、ピッチの高い復号信号はピッチの低い復号信号よりも聴感的品質は低い傾向にある。これは、CELP のアルゴリズムにより声の高い音声信号の方が量子化雑音を発生しやすいためと考えられる。そのような量子化雑音を効率的に低減するために、我々は Bass post-filter における幾つかのパラメタをピッチ周期に依存させて制御する方法を考察した。

そこで、音質を向上させる Bass post-filter の特定のパラメタと音声のピッチ周期情報の関係を知るために、PESQ による客観評価 MOS 値 (PESQ score) を用いて、音声ファイルごとに PESQ score が最も向上する特定のパラメタの値を算出する予備実験を行った。そして、各音声ファイルの基本周波数の平均値と PESQ score が最も向上した特定のパラメタの値をペアデータとして、回帰直線を得ることで、Bass post-filter のパラメタとピッチ周期の関係をモデル化することとした。予備実験用データセットとして、二つの文章を読み上げる約 8 秒の男女クリーン音声 100 ファイルと、男女雑音重畳音声 200 ファイルを用いた [21], [22]。音声は広帯域信号であり、複数言語で多数の話者が発話したデータを用いた。雑音としては、SNR が 15 dB または 20 dB のカーノイズとオフィスノイズを用いた。なお、本論文における雑音重畳音声の生成方法は、 -26 dBov に調整した音声と -41 dBov 若しくは -46 dBov に調整したノイズを合成するという標準的な方法を用いた [23]。ピッチ周期情報としては、音声ファイルごとの基本周波数の平均値を STRAIGHT [24], [25] により算出した。

3.2 ピッチ強調帯域の可変化

G.718 における Bass post-filter では、ピッチ強調用信号をカットオフ周波数が 500 Hz の低域通過フィルタに通しているため、ピッチ強調を行う周波数帯域は一定であり、音声の基本周波数によってピッチ強調の影響を受ける調波周波数が異なる。すなわち、基本周波数によって強調される倍音成分が変化することになるため、音質に何らかの影響を与えるのではないかと推測できる。そこで、低域通過フィルタの帯域を処理フレームごとに基本周波数に合わせて可変化することで、ピッチ強調の影響を受ける調波周波数 (言い換えれば強調される倍音成分の個数) が統一されれば、より効果的に音質を向上させることができると期待される。そこで、本研究では音声のピッチ構造を知る手がかりとなるピッチ周期情報を用いて、Bass post-filter における低域通過フィルタの処理フレームごとの可変化を実験的に試みる。

音声ファイルごとの基本周波数の平均値と、PESQ score を最も向上させる Bass post-filter の低域通過フィルタのカットオフ周波数を図 2 に示す。ここで、カットオフ周波数は 50 Hz, 100 Hz, $\dots$, 2000 Hz であり、40 個の 32 次 FIR フィルタを用意し予備実験を行った。また、カットオフ周波数が 250 Hz, 500 Hz, 1000 Hz, そして 2000 Hz の低域通過フィルタの応答は図 3 で表される。この予備実験の結果から、基本周波数が比較的高い女性音声では、G.718 で採用されているカットオフ周波数 $\omega_c = 500$ Hz より高いカットオフ周波数の低域通過フィルタを使用することで、音質が向上する可能性があることが示唆された。また、

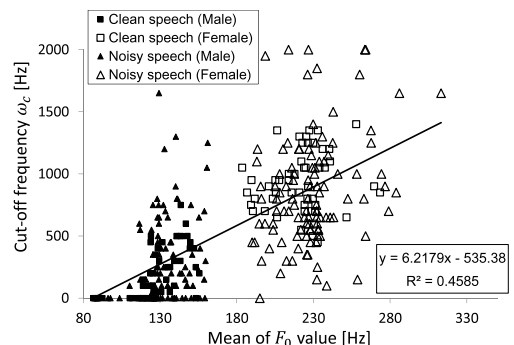


図 2 各音声ファイルにおける平均 F_0 と PESQ score を最も向上させるカットオフ周波数の関係

Fig. 2 Relationship between the cut-off frequency that achieves maximum PESQ score and the mean of F_0 value.

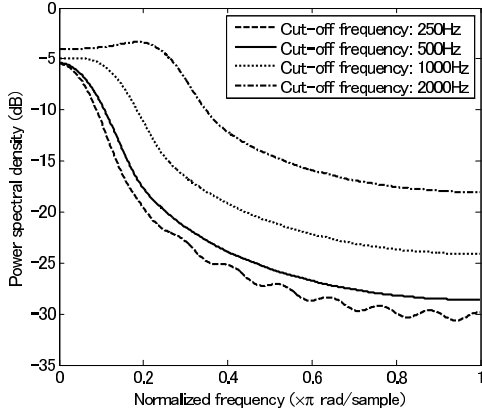


図 3 ピッチ強調帯域の可変化に用いる低域通過フィルタの応答

Fig. 3 Frequency responses of low-pass filters used in adaptive control of parameters of bass post-filter.

カットオフ周波数 ω_c と基本周波数 F_0 において,

$$\omega_c = 6.2179F_0 - 535.38 \quad (19)$$

となる近似直線が得られた. この式 (19) を用いて Bass post-filter の低域通過フィルタを処理フレームごとに可変化すれば, 音質向上につながることが期待される.

3.3 ピッチ強調利得上限値の可変化

式 (9) で算出される利得 α は, 式 (10) のように $\alpha_{max} = 0.5$ として上限を制限されている. しかし, この制限に理論的な妥当性は見出せなかったため予備実験を行ったところ, PESQ score が向上するような利得上限値はピッチ周期に依存性があることが確認できた. そこでピッチ周期に依存させてピッチ強調利得の上限値 α_{max} を可変化させることによる音質改善を試みる.

音声ファイルごとの基本周波数の平均値と, PESQ score を最も向上させる Bass post-filter のピッチ強調利得上限値を図 4 に示す. ここで, 利得上限値は 0.01 刻みで 0.0~1.2 の区間を調査した. なお, 低域通過フィルタのカットオフ周波数は従来どおり 500 Hz に固定した. この予備実験の結果, 基本周波数が比較的低い男性音声では, G.718 で採用されているピッチ強調利得上限値 $\alpha_{max} = 0.5$ より低い値を上限値とすることで音質向上の可能性がある. また, ピッチ強調利得上限値 α_{max} と基本周波数 F_0 において,

$$\alpha_{max} = 0.0036F_0 - 0.1820 \quad (20)$$

となる回帰直線が得られた. この式 (20) を用いて Bass

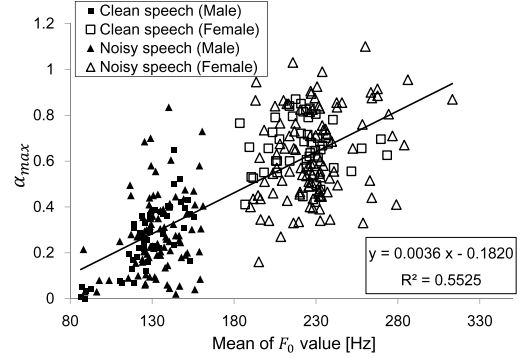


図 4 各音声ファイルにおける平均 F_0 と PESQ score を最も向上させる利得上限値の関係

Fig. 4 Relationship between the upper limit of bass post-filter gain that gave the maximum PESQ score and the mean of F_0 value.

post-filter のピッチ強調利得上限値を処理フレームごとに可変化することで音質向上が期待される.

3.4 ピッチ強調帯域と利得上限値可変化の実装

10 ms の処理フレームごとに, 復号器で使用し補正されたピッチ周期 τ を用いて基本周波数 $\hat{F}_0$ を,

$$\hat{F}_0 = \frac{F_s}{\tau} \quad (21)$$

のように求める. ここで, 標本化周波数は $F_s = 16$ kHz である. この $\hat{F}_0$ を用いて, 式 (19), (20) よりピッチ強調利得上限値 $\hat{\alpha}_{max}$ と低域通過フィルタのカットオフ周波数 $\hat{\omega}_c$ を処理フレームごとに可変化し音質向上を試みた. ただし, $\hat{\alpha}_{max} < 0$ のときは $\hat{\alpha}_{max} = 0$ とした. また, カットオフ周波数の可変化は,

$$\hat{\omega}_c = 50 \left\lfloor \frac{6.2179\hat{F}_0 - 535.38}{50} \right\rfloor \quad (22)$$

として 50 Hz ごとに量子化された 32 次 FIR フィルタによる低域通過フィルタを用いて実装した. ここで, $\lfloor \cdot \rfloor$ は床関数を表し, $\hat{\omega}_c$ の値域は $0 \leq \hat{\omega}_c \leq 2000$ として上限・下限値で制限する. 更に, $\hat{\omega}_c = 0$ のときは Bass post-filter の処理を行わないこととした. 提案法は上記のような実装のため, デコーダ側の ROM にローパスフィルタ係数のテーブルセットを追加するだけで容易に実装することができ, 記憶容量の増加は少ない. また, カットオフ周波数や利得上限値の値は, それぞれ処理フレームごとに式 (22) と式 (20) の回帰式への代入演算によって算出できるため, 計算量の増加もわずかである. なお, 3.2 と 3.3 における回帰直線及び決定係数の導出をクリーン音声と雑音重畳音

声に分けて行うことで、表 1, 2 の予備実験結果を得た。この表より雑音重畳音声では背景雑音の影響により R^2 が低いため、カットオフ周波数と利得上限値の可変化による音質の向上効果は低いと予想される。また、雑音重畳音声ではピッチ相関が低く、 α が小さくなってポストフィルタの効果が小さいということからも音質の向上効果は小さいと予想される。

表 1 カットオフ周波数可変化のためのクリーン音声と雑音重畳音声の回帰直線及び決定係数

Table 1 The regression line and the coefficient of determination in clean speech and noisy speech for the adaptive control of cut-off frequency.

	回帰直線	決定係数 R^2
クリーン音声	$\omega_c = 7.3902F_0 - 727.22$	0.7432
雑音重畳音声	$\omega_c = 5.6698F_0 - 440.26$	0.3549
クリーン及び雑音重畳音声	$\omega_c = 6.2179F_0 - 535.38$	0.4585

表 2 利得上限値可変化のためのクリーン音声と雑音重畳音声の回帰直線及び決定係数

Table 2 The regression line and the coefficient of determination in clean speech and noisy speech for the adaptive control of upper gain limit.

	回帰直線	決定係数 R^2
クリーン音声	$\alpha_{max} = 0.004F_0 - 0.2363$	0.7218
雑音重畳音声	$\alpha_{max} = 0.0034F_0 - 0.162$	0.4862
クリーン及び雑音重畳音声	$\alpha_{max} = 0.0036F_0 - 0.1820$	0.5525

以上の実装によりカットオフ周波数または利得上限値の可変化を実装することができるが、カットオフ周波数可変化は声の高い音声に対して、また、利得上限値可変化は声の低い音声に対して PESQ score の改善度が大きかったことを我々は確認した。したがって、図 5 に示すように、ある周波数をしきい値として処理フレームごとにカットオフ周波数を可変化するか、利得上限値を可変化するか切り替える手法を提案する。この手法を実装するために、本章のデータセットを用いて、切り替えに適したしきい値 $F_{Threshold}$ を探索した。なお、雑音重畳音声では明確な傾向が見られないため、クリーン音声のみ用いた。図 6 は、横軸のピッチ周波数以上のピッチ周波数の処理フレームではカットオフ周波数を可変化し、それ以外の処理フレームでは利得可変化を行う手法に対して、男女クリーン音声それぞれの平均 PESQ score を算出したものである。ここで、エラーバーは 95 % 信頼区間を表している。この結果から、男女クリーン音声で有意差はないが、系統的な傾向はあることが確認できる。したがって、 $F_{Threshold} = 220$ Hz 付近の値がしきい値として適していることを確認した。そこで、図 7 で表されるように、現在の処理フレームのピッチ周波数が 220 Hz 以上の処理フレームではカットオフ周波数 ω_c を可変化し、それ以外では利得上限値 α_{max} の可変化を行う手法を実装した。

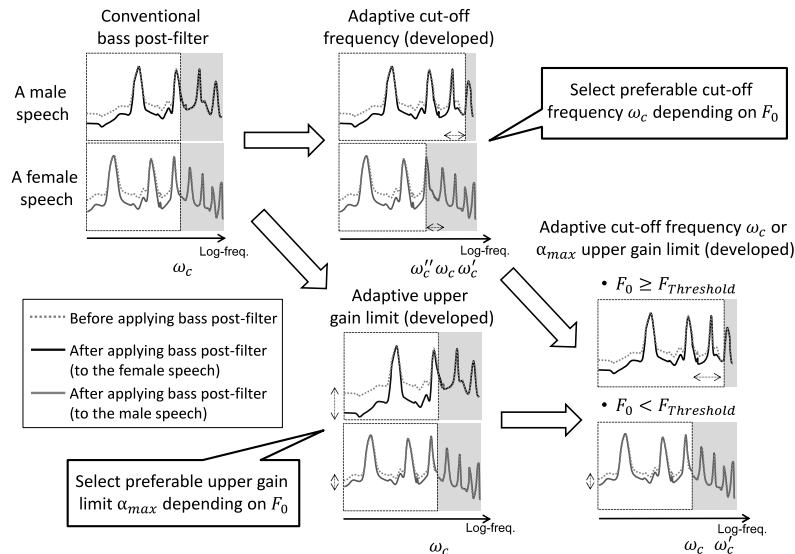


図 5 Bass post-filter における従来及び提案法による周波数スペクトルへの影響
Fig. 5 Effectiveness of conventional and developed bass post-filter.

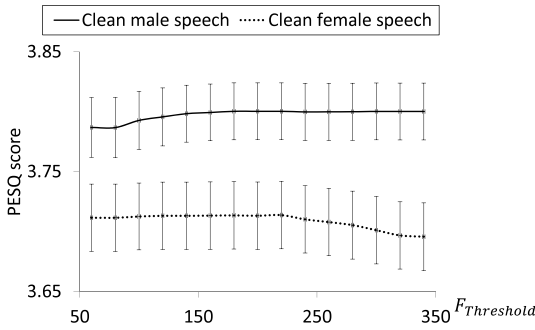


図 6 カットオフ周波数または利得上限値可変化のどちらかを決定するしきい値となるピッチ周波数の探索
Fig. 6 Search of pitch frequency that is the threshold to determine either the cut-off frequency or gain the upper limit allowed adaptive control.

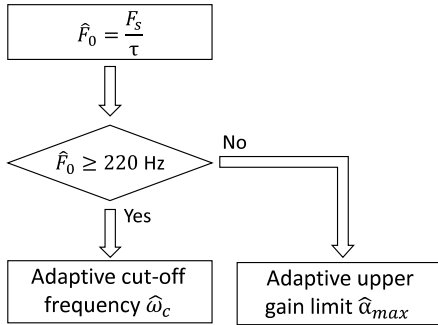


図 7 カットオフ周波数または利得上限値可変化のフローチャート
Fig. 7 A flowchart of the developed method that controlled either the cut-off frequency or gain the upper limit adaptively.

4. 評価実験

4.1 実験条件

評価実験では、3. で使用したデータセットとは異なる評価用データセットを用いた。評価用データセットは二つの文章を読み上げる約 8 秒の男女クリーン音声 231 ファイルと雑音重畳音声 200 ファイルを用いた。音声は広帯域信号であり、複数言語で多数の話者が発話したデータを用いた。雑音としては、SNR が 15 dB または 20 dB のカーノイズとオフィスノイズを用いた。評価実験用符号化方式は G.718 Layer 1 (8 kbit/s) とした。評価は、カットオフ周波数とピッチ強調ゲイン利得上限値が固定されている従来法 (G.718) と、カットオフ周波数を可変化する手法 (Adaptive ω_c)、ピッチ強調利得上限値を可変化する手法 (Adaptive α_{max})、そして 220 Hz より小さい基本周波数ではピッチ強調利得上限値を可変化し、220 Hz 以上の基本周波数ではカットオフ周波数を可変化する手法 (Adaptive ω_c or α_{max}) で行う。

4.2 PESQ による客観評価

クリーン音声と雑音重畳音声の客観評価結果を図 8 に示す。図中の棒グラフとエラーバーは、各データセットにおける PESQ score の平均値と 95% 信頼区間を表している。女性クリーン音声では Adaptive ω_c が音質向上に有効であり、男性クリーン音声では Adaptive α_{max} が音質向上に有効である可能性が確認できた。また、ピッチ周期により処理ごとにカットオフ周波数がピッチ強調利得上限値のどちらかを可変化する Adaptive ω_c or α_{max} は、男性音声、女性音声どちらに対しても音質向上の可能性がある。なお、

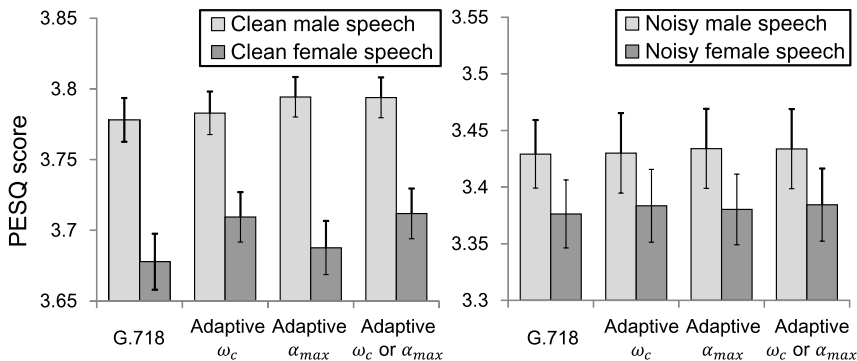


図 8 広帯域 PESQ による提案法の客観評価結果
Fig. 8 Objective evaluation of speech quality with wideband PESQ.

雑音重畳音声では PESQ score にほぼ差はない。これは 3.2 と 3.3 における回帰直線の傾向とも一致する。

以上の PESQ score の平均値による客観評価において、男女クリーン音声で統計的な音質向上が確認された、処理フレームごとにカットオフ周波数若しくは利得上限値を可変化する提案法 (Adaptive ω_c or α_{max}) と従来法の PESQ score の差を観察した。図 9, 10 の一つのプロットは、それぞれ男性クリーン音声、女性クリーン音声のデータセットの一つのファイルにおける提案法と従来法との PESQ score の差を表しており、この数値が高いほど提案法による音質向上の効果が高いことを意味している。ここで、横軸が各音声ファイルの基本周波数の平均値を示している。この結果から、提案法は音声ファイルの中でも基本周波数が比較的高い、または低いほど音質が向上することを確認した。特に基本周波数の平均値の高い女性音声においては、PESQ score が従来法より約 0.15 以上向上する音声ファイルがあることが確認できた。また、従来法と比較して PESQ score が低下している音声ファイルが少なく、低下したときの PESQ score の差も比較的小さい。したがって、この提案法は基本周波数の高い音声に対して従来の Bass post-filter よりも品質改善効果が高い処理であると考えられる。また、基本周波数が低い音声に対しては、従来の Bass post-filter によるピッチ強調で音質が劣化する可能性があり、提案法で利得の上限を小さくすることでピッチ強調処理を緩和して品質改善を実現している。

4.3 MUSHRA による主観評価

従来法 (G.718) と提案法 (Adaptive ω_c or α_{max}) において、前節の客観評価で PESQ score の差が大きかった六つの日本人男女クリーン音声について、8 人の被験者による MUSHRA での主観評価結果を図 11 に示す。なお、雑音重畳音声は前節の PESQ による客観評価において従来法と提案法に有意差がなかったため、試験を実施しない。図中のマーカーは各音声ファイルにおける提案法と従来法の MUSHRA スコアの差の平均値であり、エラーバーは 95% 信頼区間を表している。すなわち、正の値が大きいほど提案法の音質が向上していることを示している。まず、六つの男女クリーン音声全体の評価値 (Total) では、提案法であるカットオフ周波数若しくは利得上限値の可変化は従来の固定の場合よりも統計的な有意差をもって音質が向上する音源 (Female 1) があることを確認した。また、MUSURA スコアの差の平均値が 0 より小さい

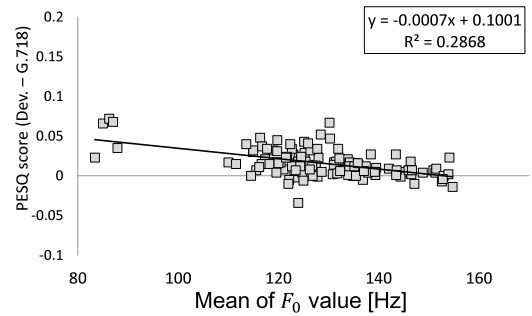


図 9 男性クリーン音声における提案法と従来法の PESQ score の差

Fig. 9 The difference of the PESQ scores between the developed method and the conventional method as for the files of male clean speech.

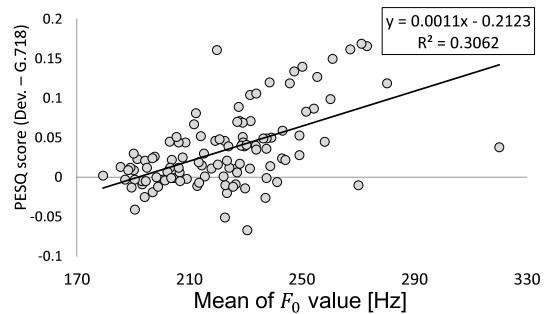


図 10 女性クリーン音声における提案法と従来法の PESQ score の差

Fig. 10 The difference of the PESQ scores between the developed method and the conventional method as for the files of female clean speech.

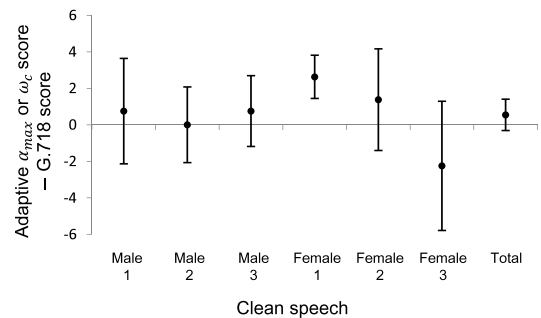


図 11 MUSHRA 法を用いた提案法の主観評価結果

Fig. 11 Subjective evaluation of speech quality with MUSHRA.

音源 (Female 3) があるが、有意差はない。

5. む す び

本論文では、ITU-T G.718 Bass post-filter におい

て、ピッチ強調帯域とピッチ強調利得をピッチ周期に依存させて処理フレームごとに可変にするポストフィルタ処理について論じた。Bass post-filter の低域通過フィルタのカットオフ周波数と利得上限値が音声のピッチ周期に依存しているか予備実験により調査した結果、どちらもほぼ線形的な依存性がある傾向を確認し、その回帰直線を利用して処理フレームごとにカットオフ周波数、利得上限値の適応的制御を実装した。この提案法とカットオフ周波数と利得上限値が固定の従来法を比較したところ PESQ による客観評価において、提案法はカットオフ周波数の可変性は女性クリーン音声に対して、また、利得上限値の可変性は男性クリーン音声に対して音質向上させることを確認した。なお、3., 4. の実験結果より、雑音が重畳した音声に対しては従来法と提案法に差は見られなかった。これらの結果は、3.4 の表 1, 2 より、背景雑音の影響により回帰直線の決定係数 (R^2) が低いことから予想できた。また、背景雑音のためにピッチ相関が低下して、ポストフィルタの利得が小さくなる傾向が確認でき、ポストフィルタの効果そのものが小さいことにも原因があると考えられる。したがって、提案法は雑音重畳音声に対しても音質を劣化させることはない、総合的に有効な音質向上手法である。また、カットオフ周波数と利得上限値の可変性を特定の周波数で切り替える手法を MUSHRA 法によって主観評価したところ、6 データ中 1 データについて統計的に有意な品質改善効果が確認できた。この結果から、入力音声の調波周波数に対して、統一的にピッチ強調を行うことが聴感的音質の向上に有効である可能性がある。この手法は G.718 だけでなく、広帯域の音声符号化方式に対して適用可能であり、基本周波数の比較的高い、または低い音声において聴感的音質向上が期待される。

文 献

- [1] <http://www.hikaritv.net/4k/>
- [2] M.M. Sondhi, C.E. Schmidt, and L.R. Rabiner, "Improving the quality of a noisy speech signal," Bell Syst. Tech. J., vol.60, pp.1847–1859, 1981.
- [3] D. Malah and R. Cox, "A generalized comb filtering technique for speech enhancement," Proc. ICASSP, pp.160–163, 1982.
- [4] 長瀬裕実, "コームフィルタリング雑音抑圧法の基本性能と PCM 音声の品質改善への応用," 信学論 (A), vol.J65-A, no.6, pp.572–578, June 1982.
- [5] J.S. Lim, A.V. Oppenheim, and L. Braidia, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-26, pp.59–71, 1995.
- [6] J.-H. Chen and A. Gersho, "Real-time vector APC speech coding at 4800 bps with adaptive post filtering," Proc. ICASSP, pp.2185–2188, 1987.
- [7] 守谷健弘, 音声符号化, コロナ社, 1998.
- [8] P. Kroon and B.S. Atal, "Quantization procedures for the excitation in CELP coders," Proc. ICASSP, pp.1649–1652, 1987.
- [9] D.E. Veeneman and B. Mazor, "Enhancement of block-coded speech," Proc. ICASSP, pp.193–196, 1987.
- [10] W.B. Kleijn and K.K. Paliwal, "Speech coding and synthesis," Elsevier Science, pp.105–108 & 244–245, 1998.
- [11] A.M. Kondoz, Digital Speech: Coding for Low Bit Rate Communication Systems, Second Edition, pp.255–258, John Wiley & Sons., 2004.
- [12] J.-H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," IEEE Trans. Speech Audio Process., vol.3, no.1, pp.59–71, Jan. 1995.
- [13] ITU-T Recommendation G.729, March 1996.
- [14] 3GPP TS 26.171, Adaptive Multi-Rate Wideband (AMR-WB), April 2001.
- [15] 藤原 洋, マルチメディア情報圧縮, 共立出版, 2000.
- [16] ITU-T Recommendation. G.718, June 2008.
- [17] M. Jelinek and R. Salami, "Wideband Speech Coding Advances in VMR-WB Standard," IEEE Trans. ASLP, pp.1167–1179, 2007.
- [18] H. Chiba, Y. Kamamoto, T. Moriya, N. Harada, S. Miyabe, T. Yamada, and S. Makino, "Adaptive post-filtering controlled by pitch frequency for CELP-based speech coder," Proc. Asilomar Conf. on Signals, Systems, and Computers, Nov. 2014.
- [19] ITU-T Recommendation. P.862.2, July 2007.
- [20] ITU-R Recommendation. BS.1534-1, March 2003.
- [21] <http://www.ntt-at.co.jp/product/wideband/>
- [22] <http://www.ntt-at.co.jp/product/demwa02/>
- [23] 3GPP, "EVS Permanent Document EVS-7a: Processing functions for qualification phase," S4-130155, 2013.
- [24] H. Kawahara, H. Katayose, A.de Cheveigne, and R.D. Patterson, "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity," Proc. EUROSPEECH, vol.6, pp.2781–2784, Sept. 1999.
- [25] H. Kawahara, "Vocoder のもう一つの可能性を探る: 音声分析変換合成システム STRAIGHT の背景と展開," J. Acoust. Soc. Jpn., vol.63, no.8, pp.442–449, 2007.

付 録

3.4 の表 1, 2 に示した、クリーン及び雑音重畳音声ファイルにおける平均 F_0 と PESQ score を最も向

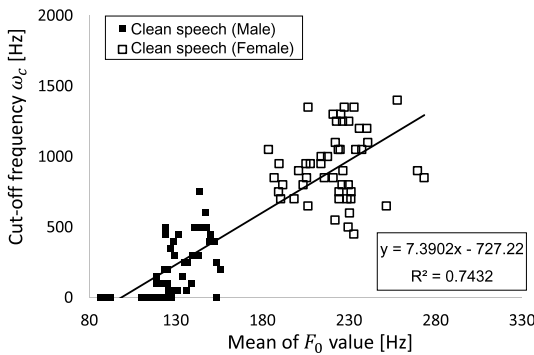


図 A・1 クリーン音声ファイルにおける平均 F_0 と PESQ score を最も向上させるカットオフ周波数の関係
Fig. A・1 Relationship between the cut-off frequency that achieves maximum PESQ score and the mean of F_0 value in clean speech.

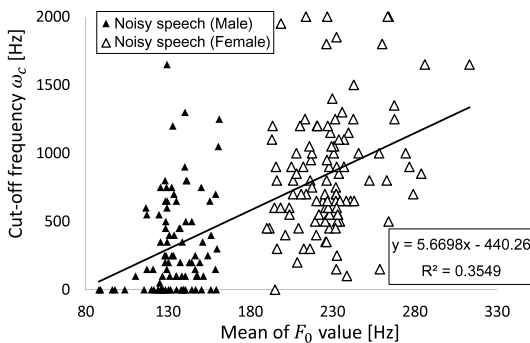


図 A・2 雑音重畳音声ファイルにおける平均 F_0 と PESQ score を最も向上させるカットオフ周波数の関係
Fig. A・2 Relationship between the cut-off frequency that achieves maximum PESQ score and the mean of F_0 value in noisy speech.

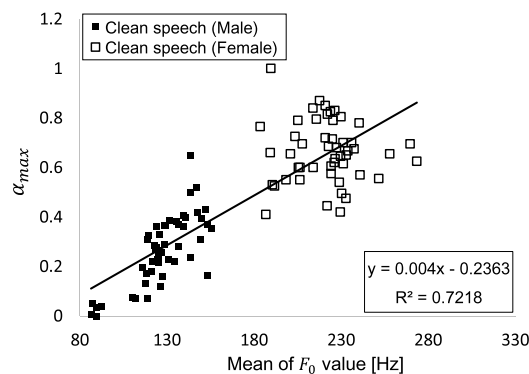


図 A・3 クリーン音声ファイルにおける平均 F_0 と PESQ score を最も向上させる利得上限値の関係
Fig. A・3 Relationship between the upper limit of bass post-filter gain that gave the maximum PESQ score and the mean of F_0 value in clean speech.

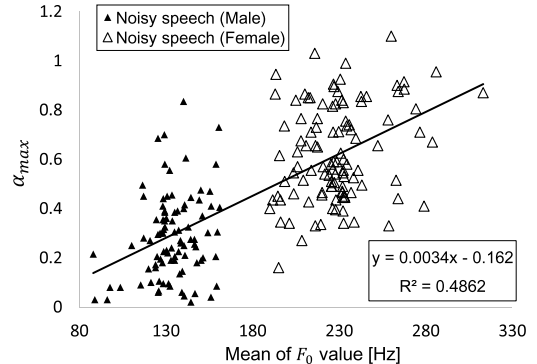


図 A・4 雑音重畳音声ファイルにおける平均 F_0 と PESQ score を最も向上させる利得上限値の関係
Fig. A・4 Relationship between the upper limit of bass post-filter gain that gave the maximum PESQ score and the mean of F_0 value in noisy speech.

上させるカットオフ周波数及び利得上限値を図 A・1, A・2, A・3, A・4 に示す。

(平成 27 年 1 月 15 日受付, 4 月 28 日再受付,
7 月 2 日早期公開)



千葉 大将

響学会会員。

2013 筑波大学情報学群情報科学類卒.
2015 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻・博士前期課程了. 修士 (工学). 同年バイオニア株式会社に入社. 在学中, 音声符号化, アレー信号処理に関する研究に従事. 日本音



鎌本 優 (正員)

2003 慶應義塾大学理工学部物理情報工学科卒. 2005 東京大学大学院情報理工学系研究科システム情報学専攻修士課程了. 同年日本電信電話株式会社入社. 現在, NTT コミュニケーション科学基礎研究所研究主任. 信号処理, 情報理論に関する研究及び音声音響符号化の標準化に従事. 博士 (情報理工学). 日本音響学会, 情報処理学会, IEEE 各会員.



守谷 健弘 (正員)

1978 東京大学工学部計数工学科卒。1980 東京大学大学院工学系研究科計数工学専門課程修士課程了。同年日本電信電話公社入社。現在 NTT コミュニケーション科学基礎研究所特別研究室長。音声音響信号の符号化の研究開発、標準化に従事。工学博士。本会より学術奨励賞、論文賞、業績賞、功績賞受賞。NTT Fellow, IEEE Fellow.



原田 登 (正員)

1995 九州工業大学情報工学部機械システム工学科卒。1997 九州工業大学大学院情報工学研究科情報システム専攻修士課程了。同年日本電信電話株式会社入社。現在、NTT コミュニケーション科学基礎研究所主幹研究員。音声・楽音の高効率符号化、音声情報処理、VoIP の研究に従事。日本音響学会、Audio Engineering Society, IEEE 各会員。



宮部 滋樹 (正員)

2003 神戸大学工学部電気電子工学科卒。2007 奈良先端科学技術大学院大学博士後期課程了。2008 米ジョージア工科大学客員研究員。2009 東京大学特任研究員。2010 東京大学助教。2011 筑波大学助教。現在に至る。音響信号処理の研究に従事。博士(工学)。日本音響学会、IEEE、各会員。



山田 武志 (正員)

1994 大阪市立大学工学部情報工学科卒。1999 奈良先端科学技術大学院大学情報科学研究科情報処理学専攻博士後期課程了。同年筑波大学講師。2005 筑波大学准教授(助教授)。現在に至る。音声認識、音環境理解、多チャネル信号処理、メディア品質評価、e ラーニングの研究に従事。博士(工学)。IEEE、情報処理学会、日本音響学会、日本言語テスト学会、各会員。



牧野 昭二 (正員：フェロー)

1979 東北大学工学部卒。1981 東北大学大学院修士課程了。同年日本電信電話公社入社。2009 筑波大学生命領域学際研究センター教授。現在に至る。電気音響変換器、音響エコーキャンセラ、ブラインド音源分離などの音響信号処理の研究に従事。工学博士。IEEE Distinguished Lecturer. IEEE Fellow. 本会 Fellow.