

ACELP 用ポストフィルタのピッチ強調帯域及び利得の適応化*

☆千葉大将 (筑波大), 鎌本優 (NTT・CS研), 守谷健弘 (NTT・CS研), 原田登 (NTT・CS研), 宮部滋樹 (筑波大), 山田武志 (筑波大), 牧野昭二 (筑波大)

1 はじめに

音声符号化では、復号信号に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的な品質を向上させており [1], 8 kbit/s 程度の広帯域音声通信での運用をサポートしている ITU-T G.718 ACELP 方式 [2] では低周波数帯域のみピッチ強調を行う Bass post-filter が採用されている。本研究では、ピッチ周期情報を用いて Bass post-filter のピッチ強調利得を可変化する事で音質向上を試みる。また、提案手法の音質を評価するために、広帯域用 PESQ [3] による客観評価と MUSHRA [4] による主観評価を行う。

2 G.718 における Bass post-filter

G.718 ACELP 復号後に用いられる Bass post-filter の処理を Fig.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、20 ms のパケット伝送フレーム長 L より先の $\hat{s}[n]$ は、

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

として補正されたピッチ周期 τ で外挿し補う。

ピッチ強調処理は、まず復号信号 $\hat{s}[n]$ から前後 τ サンプル離れた信号より長期予測信号 $s_p[n]$ を、

$$s_p[n] = 0.5\hat{s}[n-\tau] + 0.5\hat{s}[n+\tau] \quad (2)$$

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、さらに $\hat{s}[n]$ 、 $\hat{s}_f[n]$ よりピッチ強調用信号 $r[n]$ を、

$$r[n] = \hat{s}_f[n] - \hat{s}[n] \quad (3)$$

として生成する。ここで、

$$\hat{s}_f[n] = (1-\alpha)\hat{s}[n] + \alpha s_p[n] \quad (4)$$

であり、 $\hat{s}_f[n]$ は

$$\alpha = \frac{C_p}{0.5(E_p + 10^{0.1\bar{E}_{pp}})} \quad (5)$$

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。 α は式 (5) より算出後、

$$\alpha = \begin{cases} 0.5 & (\alpha > 0.5) \\ 0.0 & (\alpha < 0.0) \end{cases} \quad (6)$$

として上限値・下限値で制限される。式 (3) のピッチ強調信号 $r[n]$ はカットオフ周波数が 500 Hz の 32 次 FIR フィルタで低域通過フィルタリングされ、

$$r_{LP}[n] = \sum_{k=0}^{32} b[k]r[n-k] \quad (7)$$

として低周波数帯域のみのピッチ強調用信号 $r_{LP}[n]$ の生成に用いられる。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 $r_{LP}[n]$ から低周波数帯域のみピッチ強調された復号信号 $\hat{s}_{out}[n]$ を、

$$\hat{s}_{out}[n] = \hat{s}[n] + r_{LP}[n] \quad (8)$$

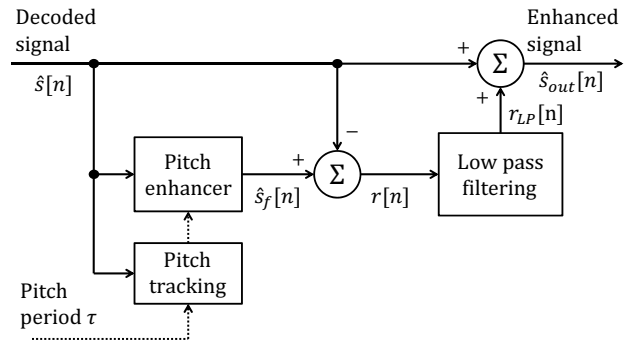


Fig. 1 Block diagram of bass post-filter.

として生成し、聴感的な音質を向上させている。

3 ピッチ周期に応じた利得可変性とピッチ強調帯域の検討

3.1 ピッチ強調利得上限値の可変性

G.718 の Bass post-filter では、ピッチ強調用信号の利得を制御するピッチ強調利得 α の上限値は $\alpha_{max} = 0.5$ で制限される。しかし、音声の基本周波数に応じて音質の向上に適した上限値 α_{max} は異なる可能性があるため、基本周波数に依存させてピッチ強調利得の上限値 α_{max} を可変化する事を試みる。

3.2 ピッチ強調利得上限値の可変性の予備実験

音質を向上させる Bass post-filter のピッチ強調利得上限値と音声のピッチ周期情報の関係を PESQ による客観評価 MOS 値 (PESQ-MOS) を用いて調べた。調査用データセットとして、2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。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。ピッチ強調用信号における利得上限値としては 0.01 刻みで 0.0~1.2 の区間を調査した。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 [5] により算出した。

音声ファイルごとの基本周波数の平均値と、PESQ-MOS を最も向上させる Bass post-filter のピッチ強調利得上限値を Fig.2 に示す。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、

$$\alpha_{max} = 0.0036F_0 - 0.1820 \quad (9)$$

となる回帰直線が得られた。この式 (9) を用いて Bass 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することで音質向上が期待される。

3.3 ピッチ強調帯域の可変性

前回発表 [6] では、Bass post-filter 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依存させて可変することで音質向上の可能性を確認した。本稿では、前節のピッチ強調利得上限値の可変性との

* A new post-filtering method based on an adaptive band width for pitch enhancement and an adaptive gain depending on pitch-lag of a decoded ACELP signal. by CHIBA Hironobu (Univ. of Tsukuba), KAMAMOTO Yutaka (NTT), MORIYA Takehiro (NTT), HARADA Noboru (NTT), MIYABE Shigeki (Univ. of Tsukuba), YAMADA Takeshi (Univ. of Tsukuba), MAKINO Shoji (Univ. of Tsukuba)

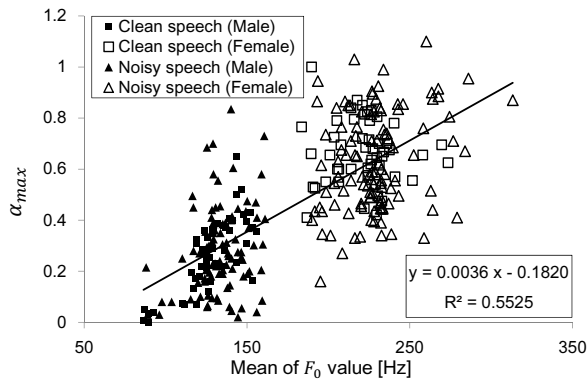


Fig. 2 Relationship between the upper limit of Bass post-filter gain that provided maximum PESQ-MOS and the mean of F_0 value.

組み合わせを考慮し、カットオフ周波数 ω_c と基本周波数 F_0 において、

$$\omega_c = 6.2179F_0 - 535.38 \quad (10)$$

となる前回発表とは異なる回帰直線を導出した。ここで、使用したデータセットは前節と同様である。

3.4 ピッチ強調帯域と利得上限値可変の実装

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、

$$\hat{F}_0 = \frac{F_s}{\tau} \quad (11)$$

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 (9) (10) 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し音質向上を試みた。ただし、 $\hat{\alpha}_{max} < 0$ の時は $\hat{\alpha}_{max} = 0$ とした。また、カットオフ周波数の可変は、

$$\hat{\omega}_c = 50 \left\lfloor \frac{6.2179\hat{F}_0 - 535.38}{50} \right\rfloor \quad (12)$$

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。さらに、 $\hat{\omega}_c = 0$ の時は Bass post-filter の処理を行わないこととした。

4 評価実験

4.1 実験条件

評価実験では、3.1 節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 2 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。評価実験用コーデックは G.718 Layer 1 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) と、ピッチ強調利得上限値を可変化する手法 (Adaptive α_{max}) と、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 Fig. 3 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-MOS 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

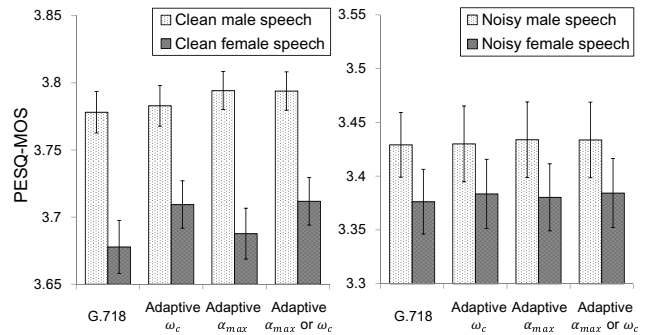


Fig. 3 Objective evaluation of speech quality with wideband PESQ.

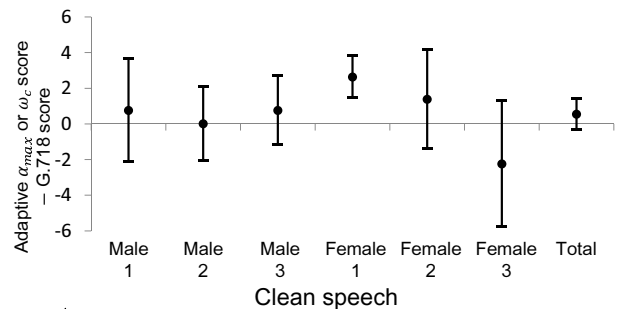


Fig. 4 Subjective evaluation of speech quality with MUSHRA.

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、雑音重畳音声では PESQ-MOS にほぼ差はない。

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-MOS の差が大きかった 6 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 Fig. 4 に示す。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。提案法は女性クリーン音声ファイル Female 1 で統計的に有意に音質が向上していることを確認できた。

5 まとめ

本稿では、ITU-T G.718 Bass post-filter のピッチ強調利得をピッチ周期に依存させて処理ごとに可変にする手法を提案した。また、前回発表で提案した、ピッチ周期に依存させてピッチ強調帯域を可変にする手法と組み合わせると主観評価した結果、クリーン音声の音質を向上させることを確認できた。

参考文献

- [1] J. Chen *et al.*, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Trans. on Speech and Audio Proc.*, vol. 3, no. 1, pp.59-71, Jan. 1998.
- [2] ITU-T Recomm. G.718, June 2008.
- [3] ITU-T Recomm. P.862.2, July 2007.
- [4] ITU-R Recomm. BS.1534-1, Mar. 2003.
- [5] H. Kawahara *et al.*, "Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity," *Proc. EUROSPEECH*, vol. 6, pp.2781-2784, Sep. 1999.
- [6] 千葉 他, "復号信号の特徴に応じた ACELP 用ボストフィルタの制御," *音講論 (秋)*, pp.319-320, Sep. 2013.