復号信号の特徴に応じたACELP用ポストフィルタの制御*

☆千葉大将（筑波大），守谷健弘（ＮＴＴ・ＣＳ研），鎌本優（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的な品質を向上させている [1]。ITU-T G.718 で使われている ACELP 方式 [2] では低周波数帯域のみピッチ強調する Bass post-filter を用いているが、適用する周波数帯域は固定されている。本研究では、ピッチ周期情報を用いて Bass post-filter のピッチ強調周波数帯域を処理フレームごとに可変にする手法を提案し、音質向上を試みる。提案手法の効果を確かめるために、広帯域用 PESQ [3] による客観評価を行う。

2　G.718 における Bass post-filter

G.718 の 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、復号音声 $\hat{s}[n]$ に対して 10 ms の処理フレームごとに行う。まず復号信号 $\hat{s}[n]$ から、補正されたピッチ周期 τ が強調された中間信号 $s_p[n]$ を、

$$s_p[n] = 0.5\hat{s}[n-\tau] + 0.5\hat{s}[n+\tau] \qquad (1)$$

という処理により生成する。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、

$$\hat{s}[n+L] = \hat{s}[n+L-\tau] \qquad (2)$$

としてピッチ周期 τ で外挿し補う。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 $\hat{s}[n]$ と $\hat{s}_f[n]$ よりピッチ強調用信号 $r[n]$ を、

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

として生成する。ここで、

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

であり、ゲイン制御パラメタ α は $\hat{s}[n]$ と $s_p[n]$ より処理ごとに $0 \le \alpha \le 0.5$ の範囲で求める。この $r[n]$ 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、

$$r_{LP}[n] = \sum_{k=0}^{32} b[k]r[n-k] \qquad (5)$$

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、

$$\hat{s}_{out}[n] = \hat{s}[n] + r_{LP}[n] \qquad (6)$$

として生成し、聴感的な音質を向上させている。

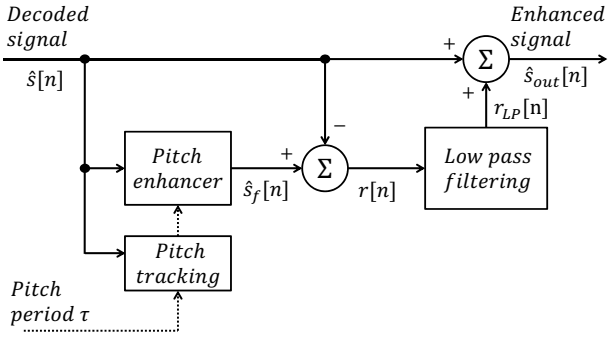

Fig. 1　Bass post-filter used in G.718.

3　ピッチ周期に依存した Bass post-filter 処理の提案

3.1　Bass post-filter 処理の可変化の検討

従来の Bass post-filter は固定の低域通過フィルタを使用しているが、音声によって低域通過フィルタの帯域を変化させることで、より効果的に音質を向上させることができると考えられる。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試みる。

3.2　予備実験

音質を向上させる Bass post-filter の低域通過フィルタの帯域と音声のピッチ周期情報の関係を、PESQ による MOS (Mean Opinion Score) 値 (以下、PESQ score) を用いて調べた。PESQ score は 5 段階主観評価を模擬した客観評価値であり、数値が大きいほど音質が良いことを示す。予備実験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。低域通過フィルタはカットオフ周波数が $50, 100, \cdots, 2000$ Hz である 40 個の 32 次 FIR フィルタを用意した。ここで、Bass post-filter の処理を行わない場合はカットオフ周波数を 0 と表記することにした。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 [4] 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 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 Fig.2 に示す。この結果から、基本周波数が比較的高い

*Adaptive ACELP post-filter depends on pitch-lag of decoded signal. by CHIBA Hironobu (Univ. of Tsukuba), MORIYA Takehiro (NTT), KAMAMOTO Yutaka (NTT), HARADA Noboru (NTT), MIYABE Shigeki (Univ. of Tsukuba), YAMADA Takeshi (Univ. of Tsukuba), MAKINO Shoji (Univ. of Tsukuba)

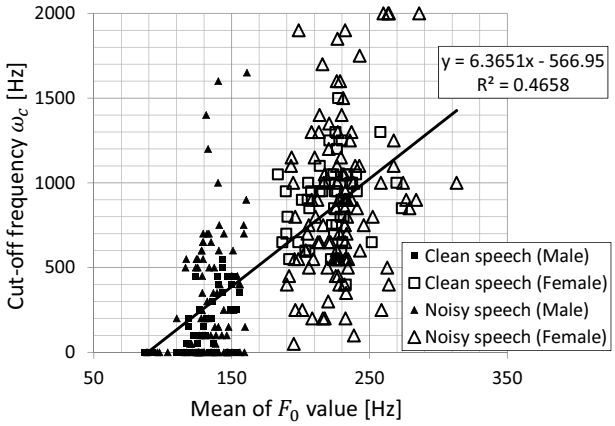

Fig. 2 Relationship between the cut-off frequency that provided maximum PESQ score and the mean of F_0 value.

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を確認できた。また、カットオフ周波数 ω_c と基本周波数 F_0 において、

$$\omega_c = 6.3651F_0 - 566.95 \tag{7}$$

となる近似直線が得られた。この時、決定係数は $R^2 = 0.4658$ であった。この式 (7) 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 Bass post-filter の低域通過フィルタ可変化の実装

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

$$\hat{F}_0 = \frac{F_s}{\tau} \tag{8}$$

のように求めることができる。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、Bass post-filter における低域通過フィルタのカットオフ周波数 $\hat{\omega}_c$ を式 (7) より、

$$\hat{\omega}_c = 50\lfloor \frac{6.3651\hat{F}_0 - 566.95}{50} \rfloor \tag{9}$$

として処理フレームごとに低域通過フィルタを変化させて音質向上を試みた。ここで、$\lfloor \cdot \rfloor$ は床関数を表し、$\hat{\omega}_c$ の値域は $0 \le \hat{\omega}_c \le 2000$ として上限・下限値で丸める。また、式 (9) で表すように 50 Hz ごとに量子化された低域通過フィルタを用いた。すなわちカットオフ周波数が $50, 100, \cdots, 2000$ Hz である 40 個の 32 次 FIR フィルタを用意した。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。このように、10 ms の処理フレームごとに式 (5) における低域通過フィルタのインパルス応答 $b[k]$ をピッチ周期で切り替えることで、Bass post-filter における低域通過フィルタの可変化を実装した。

3.4 評価実験

評価実験では、予備実験で使用したデータセット (Closed test) と、それとは異なる評価用データセット (Open test) を用いた。評価用データセットも、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。

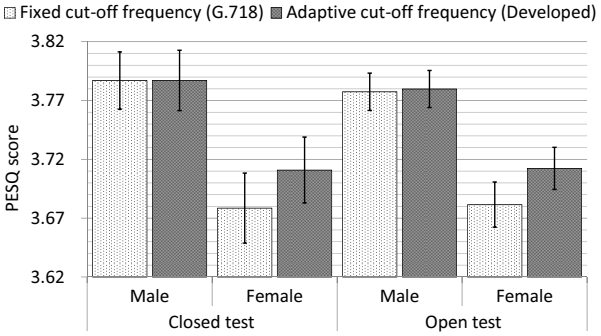

Fig. 3 Result of clean speech.

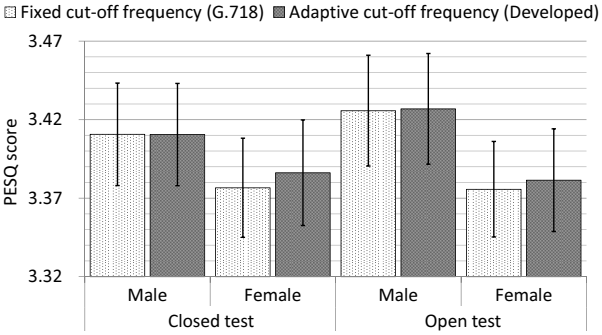

Fig. 4 Result of noisy speech.

クリーン音声の評価結果を Fig.3、雑音重畳音声の評価結果を Fig.4 に示す。図中のエラーバーは 95% 信頼区間を示している。この結果から、男性または雑音重畳音声ではカットオフ周波数が固定されている従来法 (G.718) とピッチ周期に依存した提案法 (Developed) で PESQ score にほぼ差はない。しかし、3.2 節で予想した結果通り、女性クリーン音声では統計的に有意な差を示す程度の音質向上を確認できた。

4 まとめ

本稿では、ITU-T G.718 における Bass post-filter の低域通過フィルタを音声のピッチ周期情報により可変化する手法を提案した。また、カットオフ周波数が固定されている従来法と比較して基本周波数が比較的高い女性クリーン音声では PESQ score が向上することを確認した。今後は、主観評価実験による音質改善の確認を検討している。

参考文献

[1] Juin-Hwey 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] Hideki 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.