

ACELP におけるピッチシャープニングの特性評価*

☆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

8 kbit/s 程度の音声通信では、人間の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する ACELP 音声符号化アルゴリズムが広く使用されている [1]。ACELP では、代数符号帳に対してピッチシャープニングと呼ばれるピッチ周期化処理を行うことで、音声品質を向上させている。本研究では、ITU-T G.718 Layer 1 [2] に実装されている広帯域用 ACELP 方式を評価プラットフォームとし、まず代数符号帳ピッチ周期化の必要性を確認する。そして、音声品質客観評価法である ITU-T P.862.2 PESQ [3] を用いてピッチシャープニングの特性評価を行い、ピッチシャープニングのフィルタ係数最適化を試みる。

2 評価プラットフォーム

2.1 G.718 Layer 1 ACELP 方式の概要

ACELP では、入力信号から線形予測分析により一定間隔のフレーム毎に声道特性を求め、フレームを更に分割したサブフレーム単位で、予測残差を最小化するように声道特性フィルタと励起信号の組み合わせを求めて符号化する。励起信号は、Fig. 1 で表されるように線形予測残差を聴覚重み付けされた領域で最小二乗近似する適応符号帳と代数符号帳の和により表現される。適応符号帳は過去の励起信号からピッチ周期を基にして合成され、代数符号帳は振幅 ± 1 の数個のパルス列を有限個の候補から選択することにより求められる。

G.718 Layer 1 広帯域用 ACELP 方式では、入力信号を 12.8 kHz にダウンサンプリングし、音声の短時間の状態を識別しながら、Voiced Coding (VC)、Generic Coding (GC)、Transition Coding (TC)、Unvoiced Coding (UC) の 4 つのモードを 20 ms のフレーム毎に切り替えて、5 ms のサブフレーム単位で符号化する。本研究では、頻繁に使用される VC と GC モードを対象とする。

2.2 ピッチシャープニングの概要

長さ N 、時系列 $n = 0, \dots, N - 1$ のサブフレームにおいて、代数符号帳のターゲット信号 $r[n]$ は、入力信号 $x[n]$ 、非量子化適応符号帳利得

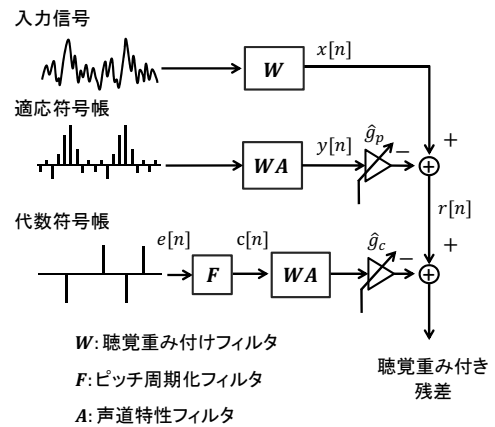


Fig. 1 ACELP の基本構造

g_p 、フィルタ処理された適応符号帳 $y[n]$ より、

$$r[n] = x[n] - g_p y[n]$$

と表すことができる。 $y[n]$ は周期成分を予測するよう合成されるが、 $r[n]$ にはピッチ周期性が残ることが知られている [4]。そこで ACELP では、

$$c[n] = \begin{cases} e[n], & n = 0, \dots, T_0 - 1 \\ e[n] + \beta c[n - T_0], & n = T_0, \dots, N - 1 \end{cases}$$

と表されるピッチシャープニングによって、代数符号帳 $e[n]$ をピッチ周期化した $c[n]$ により $r[n]$ の近似を行う。ここで、 T_0 は現在のサブフレームのピッチ周期、 β は非負のフィルタ係数を表す。G.718 では $\beta = 0.85$ という固定値が用いられているが、サブフレームごとに $r[n]$ のピッチ周期性の程度は異なると考えられるため、適応化・最適化することにより音質を改善できる可能性がある。そこで本研究では、ピッチシャープニングのフィルタ係数値に着目し改良を検討する。

2.3 研究用コーデックについて

本研究では、ピッチシャープニングによる性能差を明確化するため、G.718 Layer 1 ACELP 方式のピッチ周期分析範囲の下限値を 34 サンプル (約 376Hz) から 26 サンプル (約 492Hz) に拡張した MPLR-G.718 (Modified Pitch Lag Range G.718) を評価用コーデックとする。

*Experimental Evaluation of Pitch Sharpening in ACELP. 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)

Table 1 相関係数 R のモード比較

	VC モード	GC モード
相関係数 R	0.555	0.386

3 代数符号帳ピッチ周期化の必要性

3.1 相関係数によるピッチ周期性の調査

代数符号帳のターゲット信号 $r[n]$ のピッチ周期性の程度を、周期間の相関により調査する。ピッチ周期 T_0 がサブフレーム長 N 未満のサブフレームで処理モードごとに $r[n]$ の $0, \dots, (N-1) - T_0$ サンプルのベクトル成分と $T_0, \dots, N-1$ サンプルのベクトル成分の相関係数、

$$R = \frac{\sum_{k=0}^{(N-1)-T_0} r[k]r[k+T_0]}{\sqrt{\sum_{k=0}^{(N-1)-T_0} r[k]^2} \sqrt{\sum_{k=T_0}^{N-1} r[k]^2}}$$

の平均値を算出する。 $r[n]$ に周期性がない場合、相関係数 R は小さい値になる。調査のために約 20 分の基本周波数が比較的高い日本人女性声優の広帯域音声データを用いた。調査するピッチ周期の範囲は、MPLR-G.718 でピッチシャープニングが行われる 26~63 サンプルとした。

3.2 調査結果

Table 1 は、サブフレーム内のピッチ周期性を示す相関係数 R の平均値を、モード別に示している。VC と GC モードの両方において R が大きな値になり、代数符号帳のターゲット信号 $r[n]$ にピッチ周期性が残ることを確認した。また、処理モードごとにピッチ周期性は異なり、特に VC モードの時には、 $r[n]$ にピッチ周期性が強く残っていることを確認した。従って、ピッチシャープニングのフィルタ係数値を処理モード別に探索し、PESQ による MOS (Mean Opinion Score) 値 (以下、PESQ 値) を最大化する規準で最適化することにより、音質の改善を試みる。

4 ピッチシャープニングの特性評価

4.1 フィルタ係数値の最適化

ピッチシャープニングのフィルタ係数値 β を、予め設定した処理モードの条件に当てはまる場合のみ G.718 での固定値 0.85 から変更し、最大 PESQ 値を取る最適な係数値 β_{opt} の探索を行う。 β を変更する処理モードの条件として、VC モードのみ変更、GC モードのみ変更、VC と GC 両方のモードで共通に変更、計 3 条件で調査した。各条件で β_{opt} を用いた MPLR-G.718 を、それぞれ VC、GC、VC・GC と呼ぶことにする。また比較対象として、G.718 同様固定の $\beta = 0.85$ を用いた MPLR-G.718 を Default と呼ぶことにする。実験には約 7 分の日本人女性声優の広帯域音声データを使用した。 β の探索範囲は 0.01 刻みで 0.0~2.0 とし、それとともに変化させる PESQ

Table 2 特性評価によるフィルタ係数の最適化

	Default	VC	GC	VC・GC
係数値 β_{opt}	0.85	1.33	1.26	1.34
PESQ 値	3.412	3.430	3.421	3.438
95% CI	± 0.029	± 0.028	± 0.026	± 0.027

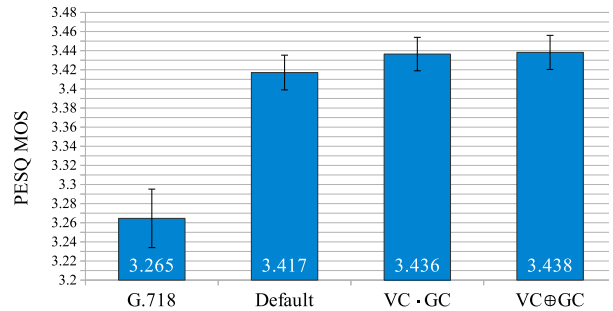


Fig. 2 フィルタ係数の最適化による性能評価

値は 5 段階主観評価を模擬したものであり、数値が大きいほど音声品質が良いことを示す客観評価値である。各 β における PESQ 値の平均値と 95% 信頼区間 (CI) を求め、 β_{opt} を探索した結果を Table 2 に示す。この結果を用いて、次節で性能評価を行う。

4.2 性能評価

評価には、前節の実験データと異なる、約 20 分の日本人女性声優の広帯域音声データを使用した。ITU-T G.718 Layer 1 (以下、G.718) と前節の Default、VC・GC に加えて、VC と GC モードを個別に探索して得た β_{opt} をそれぞれの処理モードに適用した MPLR-G.718 (以下、VC⊕GC)、計 4 つのコーデックで PESQ を用いて性能評価した結果を Fig. 2 に示す。この結果から、G.718 で用いられているピッチシャープニングの固定の β よりも高い PESQ 値を与える β_{opt} が存在し、また、処理モードを条件とした β_{opt} の最適化が音声品質の向上に有効である可能性が確認できた。

5 まとめ

本稿では、ITU-T G.718 を基にした ACELP 方式において、処理モードにより代数符号帳のターゲット信号のピッチ周期性が異なることを確認した。また、PESQ を用いてピッチシャープニングの特性評価を行いピッチシャープニングのフィルタ係数値を最適化することで、PESQ による MOS 値が向上することを確認した。今後は、主観品質評価を行うことを検討している。

参考文献

- [1] 守谷, “音声符号化,” コロナ社, 1998.
- [2] ITU-T Recomm. G.718, 06/2008.
- [3] ITU-T Recomm. P.862.2, 07/2007.
- [4] 三樹, 他, “ピッチ同期雑音励振源をもつ CELP 符号化 (PSI-CELP),” 信学会論文誌 A, Vol. J77-A No. 3, PP. 314-324, 1994.