歪みと符号長を考慮したACELPゲインコードブックの設計と評価*

☆大嶋 崇良 (東大院・情報理工), 鎌本 優 (NTT・CS 研), 守谷 健弘 (NTT・CS 研),
小野 順貴 (NII), 嵯峨山 茂樹 (東大院・情報理工)

1 はじめに

これまでの携帯電話の音声符号化は，符号誤り耐性を重視し，パラメータの可変長符号化を使わず，フレーム内ではほぼ固定のビット配分を行っている．しかし今後のパケットベースの通信では，これらの制約は不要となる．このため，可変長符号や柔軟なビット配分によって，さらなる情報圧縮が可能と見込まれる．本研究では，最新の ITU-T G.718 方式のゲインパラメータについて，可変長符号を適用した場合に符号長と歪みを効率良く抑制するようなコードブックの設計を目指す．前回発表 [1] で我々は，エントロピー制約ベクトル量子化 (Entropy-Constrained Vector Quantization; ECVQ) [2] の枠組みを適用し，seg. SNR の改善が可能であることを示した．本発表では，歪みと符号長の関係を考慮し，歪み尺度を見直すことで，品質改善を目指した．また，客観音質評価実験によってその性能を評価した．

2 G.718 における ACELP 方式の概要

まず，G.718 で用いられている ACELP 符号化の大まかな仕組みを説明する [3][4]．G.718 では，12 kbps の低ビットレート符号化において四つの処理モードを設けており，その内主に有声音を処理する二つの処理モードとして，Voiced-Coding (VC) モードと Generic-Coding (GC) モードがある．前者は定常性の高い有声音，後者はそれ以外の様々な有声音の処理にそれぞれ適用され，ほとんどの入力音声はこの二つのモードのいずれかで処理される．線形予測分析処理は，長さ 20 ms のフレーム単位で行われ，処理モードはフレーム単位で切り換わる．ただし予測残差信号についてはより短い 5 ms のサブフレーム単位で符号化される (Fig. 1)．

Fig. 2 に表されているように，ACELP において予測残差信号は，前サブフレームからの複製である Adaptive codebook とそれを修正する Algebraic codebook によって符号化され，元信号との誤差を最小とする各 codebook のゲインが同時に計算される．復号化の際には，二つの成分がそれぞれのゲインで足し合わされ，最終的に予測残差信号が合成される．

このゲインパラメータは，ベクトル量子化された二次元のテーブルを用いて符号化されている．そのテーブルは 5 bit サイズで，VC と GC の両モードで同一のものが用いられている．

3 エントロピー制約ベクトル量子化

ベクトル量子化とは，多次元の連続サンプルを有限個の代表ベクトルで置き換える方法である．そのテーブルの学習には，一般的に LBG アルゴリズム [5] が用いられる．これは k-means をベースとしたアルゴリズムで，あるテーブルサイズの下で歪みの総和が最小となるようにテーブルが設計されるため，固定長符号の条件下では最適な量子化器が得られる．しかしこの場合，量子化器の設計段階ではエントロピー自体は考慮されていないため，可変長符号を適用し

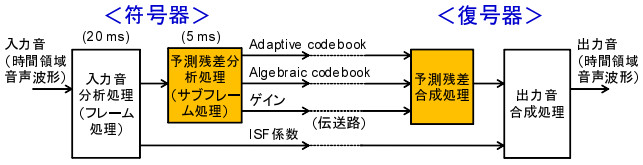

Fig. 1 ACELP 符号化の仕組み

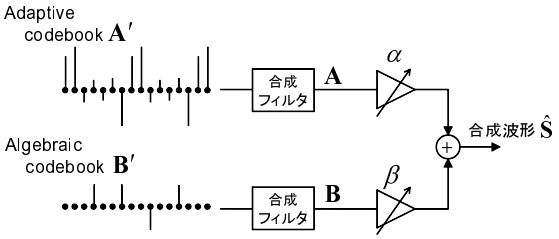

Fig. 2 励起信号のモデル化

た場合に平均符号長に対して最も効率良く歪みを抑制できている保証はない．そこで，符号長に対して最も効率良く歪みを抑制するためには，量子化器の設計の段階において符号長と歪みを同時に最適化する必要があると考えられる．

一方，テーブルの設計において歪みの総和に加えてエントロピーを考慮するベクトル量子化の方法として，ECVQ という方法がある．ECVQ では，各サンプルの所属ベクトルの計算を行う際に，距離関数として歪みに符号長の定数倍を加えたものが用いられる．結果として，歪みの総和とエントロピーの重み付け和が最小化される．しかし，この方法では歪みと符号長の関係は考慮されておらず，また，定数も実験的に定められているため，最も効率の良い最適な量子化と言うことはできない．

4 歪みと符号長の関係を考慮したゲインコードブックの設計

そこで本研究では，量子化歪みと符号長の関係を考慮することで，最適な距離関数の設計を行う．

簡単のため，スカラー量子化の場合を考えてみると，1 bit を増やすごとに平均二乗誤差が 2^{-2} 倍になることから，0 bit 時の歪みを D_0 とすると，b bit 時の歪み D_b は，

$$D_b = 2^{-2b} D_0 \tag{1}$$

と表すことができる．ベクトル量子化に拡張して考えてみると，ベクトルの次元間に相関がない場合，ベクトルの N 個の次元に b bit が均等に分配されると考えることができ，

$$D_b = 2^{\frac{-2b}{N}} D_0 \tag{2}$$

と書き換えることができる．そして歪みの減少分の対数をとると，

$$\begin{aligned} \log_{10} \frac{D_b}{D_0} &= \log_{10} 2^{\frac{-2b}{N}} \\ &= -\frac{2}{N} \log_{10} 2 \times b \\ &= -\lambda b \end{aligned} \tag{3}$$

*Design of ACELP gain codebook based on the criteria of both distortion and code length. by OSHIMA Takayoshi (the University of Tokyo), KAMAMOTO Yutaka (NTT CS Lab.), MORIYA Takehiro (NTT CS Lab.), ONO Nobutaka (NII), SAGAYAMA Shigeki (the University of Tokyo)

このようにビット数に対して線形に表現することができる．したがって，距離関数において，この λb を符号長に対するペナルティとして加えることで，歪みと符号長の双方の観点での最適化が実現されると考えられる．よって，新たな距離関数を以下のように記述することができる．

$$\log_{10} D^* = \log_{10} D + \lambda l(i) \qquad (4)$$

ここで $l(i)$ は，インデックス i のベクトルの符号長である．

また，歪みを考える上で，ゲインパラメータの幾何学的距離は音声品質とは必ずしも対応しないため，距離関数における歪みとして，ターゲット音声と量子化後合成音声との二乗誤差を用いるべきと考えられる．

$$D(\mathbf{S}, \hat{\mathbf{S}}) = \frac{\|\hat{\mathbf{S}} - \mathbf{S}\|^2}{\|\mathbf{S}\|^2} = \frac{\|\alpha\mathbf{A} + \beta\mathbf{B} - \mathbf{S}\|^2}{\|\mathbf{S}\|^2} \qquad (5)$$

ここで，$\mathbf{S}$ はターゲット音声，$\hat{\mathbf{S}}$ は量子化後の音声，$\mathbf{A}, \mathbf{B}$ はそれぞれ Adaptive codebook と Algebraic codebook のフィルタ通過後の信号，α, β は各 codebook のゲインである．

また，ゲインパラメータは二次元であるが，最終的な評価の対象とするのは音声信号でありそのベクトル長は 64 であるため，λ の式における次元数 N は 64 と考える．したがって，$\lambda = 0.009$ と計算される

学習の反復計算における各ステップは以下のようになる．まず，各サンプルについて上記の距離関数が最小となるようなベクトルに所属させる．次に，頻度の情報を更新することで，符号長の項の総和を最小化する．そして最後に，各クラスタについて重心を更新することで，歪みの総和が最小化される．

上記のアルゴリズムを用いて，ゲインテーブルの学習を行った．λ の値については，実験においては，

$$\lambda_m = 2 \times 10^{-3} m \ (m = 0, 1, 2, \cdots, 9) \qquad (6)$$

として複数の条件でテーブルを作成した．

5 性能評価実験

5.1 実験条件

作成したテーブルの性能評価実験を行った．学習データには，複数言語の clean speech，noisy speech，計約 3 時間とアカペラ曲約 3 分のデータを用いた．また評価データには，複数言語の clean speech，noisy speech，計約 40 分とアカペラ曲約 3 分のデータを用いた．

5.2 Segmental SNR による性能評価

学習したテーブルの性能評価結果の一部を Table 1 に示す．VC, GC 両モードにおいて，λ の値によって，G.718 のテーブルに対してハフマン符号を適用した場合と同程度の符号長でより高い Segmental SNR (seg. SNR) が得られるケース (歪削減モード)，また，同程度の seg. SNR でより短い符号長となるケース (符号長削減モード) の結果が得られた．いずれのモードにおいても，理論的に算出される λ に近い値でこれらのケースの結果が得られることが分かった．

5.3 客観音質評価実験

前節において，seg. SNR の値が改善されることを示したが，音声としての品質をより正確に評価するため，音声品質客観評価法である Perceptual Evaluation of Speech Quality (PESQ) [6] によって評価を行った．評価値は-0.5〜4.5 で，高音質であるほど数値は大きくなる．

Table 1　5 bit の ECVQ テーブルを用いた場合の seg. SNR と平均符号長

VC モード	G.718	歪削減 モード (λ_3)	符号長削減 モード (λ_4)
seg. SNR	8.04 dB	8.06 dB	8.04 dB
平均符号長	4.27 bit	4.27 bit	3.65 bit

GC モード	G.718	歪削減 モード (λ_4)	符号長削減 モード (λ_7)
seg. SNR	4.88 dB	4.96 dB	4.88 dB
平均符号長	4.39 bit	4.39 bit	3.91 bit

Table 2　G.718 テーブルを用いた場合の PESQ 値と平均符号長

	G.718 テーブル
PESQ	3.600 ± 0.032
平均符号長	4.39 ± 0.02 bit

Table 3　ECVQ テーブルを用いた場合の PESQ 値と平均符号長

		歪削減モード	符号長削減モード
VC & GC	PESQ	3.599 ± 0.033	3.593 ± 0.033
	平均符号長	4.41 ± 0.02 bit	3.81 ± 0.03 bit
VC only	PESQ	**3.604** ± 0.032	**3.603** ± 0.032
	平均符号長	4.40 ± 0.02 bit	**4.28** ± 0.02 bit
GC only	PESQ	3.593 ± 0.033	3.597 ± 0.032
	平均符号長	4.41 ± 0.02 bit	3.92 ± 0.03 bit

5 bit のテーブルのうち，前章の歪削減モードと符号長削減モードのそれぞれのケースについて，作成したテーブルを用いて PESQ 値を算出した．さらに，GC, VC 両モードに ECVQ テーブルを用いた場合と，片方のモードのみに用いた場合とで実験を行った．PESQ 値と平均符号長の，データ毎のそれぞれの平均値と 95% CI を Table 3 に示す．平均符号長は，GC, VC 両モードでの平均値である．

VC モードにのみ ECVQ を適用した場合に限り，従来法 (Table 2) に対して品質向上と情報圧縮の双方を実現できることが示された．また，その他の場合でも，PESQ 値をほとんど下げることなく符号長を抑制できることが分かった．

6 まとめ

国際標準方式の G.718 に対して可変長符号の適用を想定した新たな量子化方法の検討を行った．低ビットレートにおける二つの処理モードについて，歪みと符号長を同時に最適化するベクトル量子化を適用し，G.718 方式の 12 kbps モードにおいて，客観評価音質を低下させることなく 0.2 kbps 程度の情報圧縮が可能であることを示した．

今後は，主観評価実験による性能確認を行う予定である．

参考文献

[1] 大嶋崇良，他，“ACELP ゲインコードブックインデックスの可変長符号化” 音講論（春）, pp.353–354, Mar. 2012.

[2] Philip A. Chou, *et al.*, “Entropy-Constrained Vector Quantization,” IEEE Trans. ASSP, Vol. 37, No.1, pp.31–42, 1989.

[3] ITU-T Recomm. G.718.

[4] 守谷健弘 “音声符号化” 電子情報通信学会 1998.

[5] Yoseph Linde, *et al.*, “An Algorithm for Vector Quantizer Design,” IEEE Trans. Comm., Vol. 28, No.1, pp.84–95, 1980.

[6] ITU-T Recomm. P.862.2.