

# ACELP ゲインコードブックインデックスの可変長符号化\*

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 (国際電気通信連合電気通信標準化部門) G.718 方式のゲインコードブックのインデックスに対して可変長符号の適用を試みる.

## 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 符号化の大きな仕組みを説明する [1] [2]. 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, VC (Voiced-Coding) モードと GC (Generic-Coding) 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 に表されているように, ACELP において予測残差信号は, 基本的に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 が計算され, 元信号との誤差を最小とする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.

## 3 ゲインテーブルの改良の検討

### 3.1 モード別ゲインテーブルの学習

本研究では, Fig. 2 に示されるような予測残差に関する三つのパラメータに対して情報圧縮の可能性を検討しているが, 本稿ではその中からゲインパラメータへの適用について論じる. ゲインパラメータは, Adaptive codebook と Algebraic codebook の各ゲインから成る二次元パラメータであり, G.718 ではこれをベクトル量子化 (Vector Quantization; VQ) による 5 bit のテーブルを用いて符号化している. そして, VC, GC の二つのモードでは Fig. 3 に示される共通テーブルが用いられている.

そこで, 両モード間でのゲイン分布の違いの有無を確認するため, 英語の約 6 分間の広帯域音声データを用いて, 各モードでのテーブル index の頻度を比

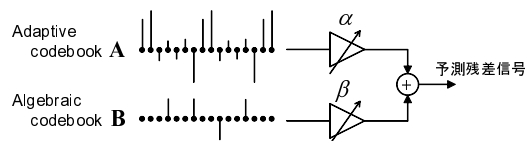


Fig. 1 励起信号のモデル化

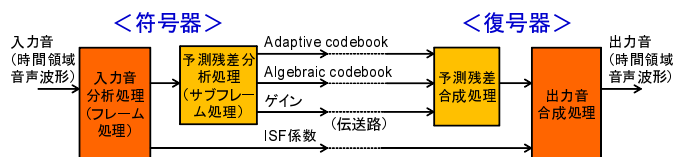


Fig. 2 ACELP 符号化の仕組み

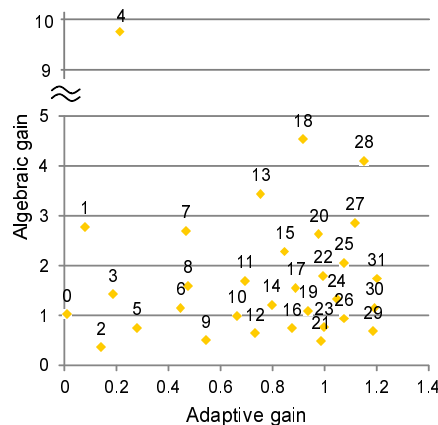


Fig. 3 G.718 ゲインテーブル

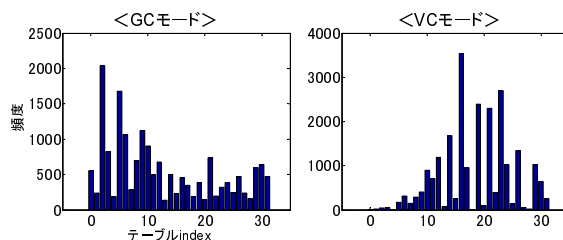


Fig. 4 GC, VC モードでのテーブル index の頻度の比較

較し, Fig. 4 に示した.

両者のヒストグラムは大きく異なっており, これを踏まえて, LBG アルゴリズム [3] を用いて GC, VC モード別のゲインテーブル (5 bit) の学習を行った. LBG アルゴリズムとは k-means 法と代表点繰り返し 2 分割の組み合わせであり, 本学習では, 各サンプルの所属計算の際の距離として, ゲインテーブル平面上の幾何学的距離ではなく, サブフレーム毎の符号化前の残差信号  $\mathbf{E} = (e_1, e_2, \dots, e_{64})$  との二乗誤差を用いた. Adaptive codebook を  $\mathbf{A}$ , Algebraic codebook を  $\mathbf{B}$  とし, それぞれゲインを  $\alpha, \beta$  とすると, 距離は以下のように書くことができる.

\* Variable-length coding of the indices of gain codebook for ACELP, by OSHIMA Takayoshi (The University of Tokyo), KAMAMOTO Yutaka, MORIYA Takehiro (NTT), ONO Nobutaka (NII), SAGAYAMA Shigeki (The University of Tokyo)

Table 1 モード別テーブル学習による SNR の比較

|            | GC モード         | VC モード          |
|------------|----------------|-----------------|
| G.718 テーブル | 6.28 dB        | 10.33 dB        |
| モード別学習テーブル | <b>6.32 dB</b> | <b>10.38 dB</b> |

Table 2 モード別テーブル学習によるエントロピーの比較

|            | GC モード          | VC モード          |
|------------|-----------------|-----------------|
| G.718 テーブル | 4.64 bit        | 4.04 bit        |
| モード別学習テーブル | <b>4.68 bit</b> | <b>4.71 bit</b> |

$$d(\mathbf{E}, \hat{\mathbf{E}}) = (\hat{\mathbf{E}} - \mathbf{E})^T (\hat{\mathbf{E}} - \mathbf{E}) \\ = (\alpha\mathbf{A} + \beta\mathbf{B} - \mathbf{E})^T (\alpha\mathbf{A} + \beta\mathbf{B} - \mathbf{E}) \quad (1)$$

学習用データには、約 4 時間 30 分の複数言語の音声データを、評価用データには、前述の英語の約 6 分間の音声データを使用し、その内 GC、VC の各モードで処理されたフレームをサンプルデータとした。そして、学習されたゲインテーブルの音質評価結果を Table 1 に示す。両モードともに、G.718 の場合と比較して SNR が向上することが分かった。

### 3.2 エントロピー制約ベクトル量子化の適用

前節で学習したモード別のゲインテーブルを用いた場合のエントロピーは、Table 2 のようになっている。SNR の値が向上した一方で、エントロピーは G.718 の場合と比較して大きくなってしまったことが分かる。

ここで、何らかの形で index の頻度に偏りを与えられれば、あまり音質を落とすことなくエントロピーを抑えることができるのではないかと考えられる。そこで、エントロピー制約ベクトル量子化 [4] (Entropy-Constrained VQ; ECVQ) の適用を検討した。

ECVQ では、サンプル毎の所属ベクトル計算の際に、歪みだけでなく符号長の定数倍の項も加えたものを距離関数とする。また、サンプル毎のエネルギー差による符号長項の影響差を解消するため、歪みの項はサンプル毎に元信号エネルギーで正規化した。ここでは簡単のため、以下のような近似を行った。

$$d'(\mathbf{E}, \hat{\mathbf{E}}) = d(\mathbf{E}, \hat{\mathbf{E}}) + \lambda_m \times (\text{符号長}) \\ \approx d(\mathbf{E}, \hat{\mathbf{E}}) + \lambda_m |\log_2 p(i)| \quad (2)$$

こうすることで、この距離を全サンプルについて平均すると

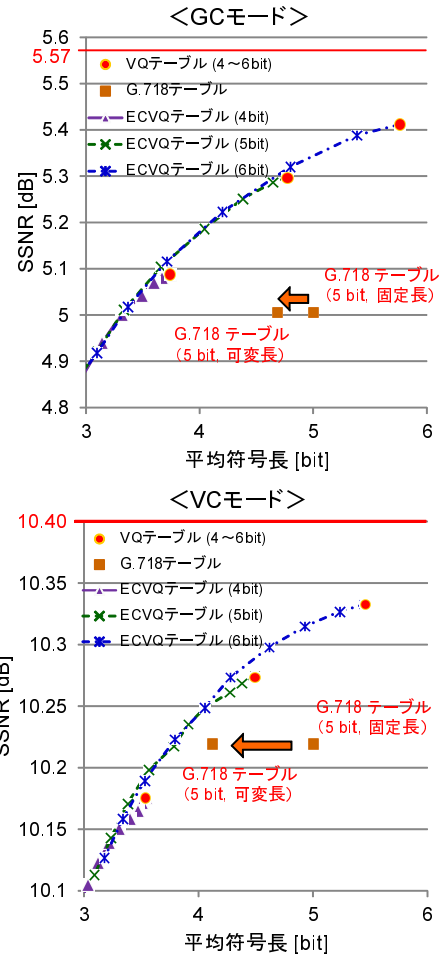
$$\bar{d}(\mathbf{E}, \hat{\mathbf{E}}) = \bar{d}(\mathbf{E}, \hat{\mathbf{E}}) + \lambda_m \sum_i p(i) |\log_2 p(i)| \quad (3)$$

のように第二項はエントロピーの定数倍となり、結果としてエントロピー制約がかかることになる。本学習では、前節と同様の方法で学習した 4~7 bit のゲインテーブルを初期テーブルとして、 $\lambda_m$  を以下のように更新して反復計算を行った。

$$\text{GC: } \lambda_m = 4 \times 10^{-3} m \quad (m = 0, 1, 2, \dots, 9) \quad (4)$$

$$\text{VC: } \lambda_m = 4 \times 10^{-4} m \quad (m = 0, 1, 2, \dots, 9) \quad (5)$$

学習したテーブルの性能評価結果を Fig. 5 に示す。評価尺度には、サンプル毎の SNR の平均値であるセグ



※赤線はゲインをベクトル量子化しない場合のSSNR

Fig. 5 G.718 のテーブルと ECVQ を用いたテーブルの平均符号長対 SSNR の比較

メンタル SNR (SSNR) を用いた。横軸がハフマン符号を適用した場合の平均符号長、縦軸が SSNR であるが、LBG アルゴリズムを用いた場合よりもプロットが左上にせり出していることが分かり、G.718 のゲインテーブルと比較して、SSNR の改善と情報圧縮の双方を実現できることが分かった。

## 4 まとめ

本稿では、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改良の検討を行った。低ビットレートにおける二つの処理モードについて予測残差信号ゲインテーブルを新たに各モード別に学習し、性能向上の可能性を示した。さらに、各モードについて ECVQ を適用することで、G.718 方式に対して SSNR の改善と平均ビットレート抑制の双方が実現可能であることを示した。

## 参考文献

- [1] ITU-T Recomm. G.718.
- [2] 守谷健弘 “音声符号化” 電子情報通信学会 1998
- [3] Yoseph Linde, *et al.*, “An Algorithm for Vector Quantizer Design,” IEEE Trans. Comm., Vol. 28, No.1, pp.84-95, 1980.
- [4] Philip A. Chou, *et al.*, “Entropy-Constrained Vector Quantization,” IEEE Trans. ASSP, Vol. 37, No.1, pp.31-42, 1989.