

対数圧伸符号化された音声信号のロスレス符号化のための線形予測分析*

堀 豊(東大・工), 守谷 健弘, 原田 登, 鎌本 優 (NTT・CS 研),
小野 順貴, 嵯峨山 茂樹 (東大院・情報理工)

1 はじめに

本稿では, IP 電話の用途でパケット量の増加が予想される G.711 準拠 [1] の対数圧伸 PCM で符号化された音声信号をロスレス符号化するための線形予測分析法を提案する. 対数圧伸信号のロスレス符号化については, 先行研究 [2] により, 対数圧伸信号に対してそのまま線形予測を行うよりも, リニア PCM に変換してからリニア PCM 領域で線形予測を行い, 予測値を対数圧伸符号化して残差を求める方が, 圧縮率が向上するという知見が得られている. 本稿では, この枠組みの中で, エントロピー符号化される予測残差のエントロピーが最小となるようリニア PCM 領域での線形予測係数の求め方を提案し, 評価実験の結果を示す.

2 対数圧伸信号と線形予測分析法によるロスレス符号化の基本原理解

2.1 対数圧伸符号化

本研究で扱う対数圧伸信号は, 以下の変換 $f(\cdot)$ を施した信号のことである.

$$f(x) := \text{sgn}(x[n]) \frac{\log(1 + \mu|x[n]|)}{\log(1 + \mu)} \quad (1)$$

ただし, $x[n]$ (整数) は時刻 n (整数) における正規化されたリニア PCM 信号であり, μ は正の定数である.

2.2 線形予測分析

線形予測分析は時刻 n の信号の振幅値 $x[n]$ を p 時刻前までの振幅値の線形和によって予測する手法である. 予測値を $\tilde{x}[n]$ とするとき

$$\tilde{x}[n] = Q \left[\sum_{i=1}^p a_i x[n-i] \right] \quad (2)$$

となる. ただし, $Q[\cdot]$ は整数化を表す記号であり, $\{a_i\}_{i=1}^p$ は線形予測係数である. ロスレス符号化においては, 線形予測係数の情報と予測残差 $e[n] := x[n] - \tilde{x}[n]$ の情報を伝送することで展開時に元の信号の復元が可能となる.

2.3 エントロピー符号化

音声の線形予測分析の場合, 予測残差の振幅頻度分布は 0 付近で大きくなり対称な分布となることが多い. そこで, 残差の符号化にはこのような分布の偏りを利用した Golomb-Rice 符号などのエントロピー符

号化がしばしば用いられる. 従って, 残差の振幅の分散を小さくし, エントロピーを小さくすることで, 符号長を短くできることが期待できる. 本稿では, 以後このような観点で論じる.

3 対数圧伸信号の線形予測分析

3.1 提案法の枠組み

対数圧伸信号を $y[n]$ (整数) とおき, リニア PCM に変換された値を $x[n] := f^{-1}(y[n])$ とおくと, 本研究では信号の予測値 $\tilde{y}[n]$ を

$$\tilde{y}[n] = f \left(\sum_{i=1}^p a_i x[n-i] \right) \quad (3)$$

で求める枠組みを考える. 線形予測係数 $\{a_i\}_{i=1}^p$ と予測残差 $e[n] := y[n] - \tilde{y}[n]$ の情報を伝送することで, ロスレス符号化が可能となる.

3.2 残差エントロピーと残差パワー

前節で定義した残差 $e[n]$ が白色雑音で, 振幅頻度分布として分散 σ_a^2 の対称な分布 $P_a(x)$ を持つことを仮定する. また, $P_a(x)$ の分散だけを σ_b^2 に変えた分布を $P_b(x)$ とするとき, 2つの分布 $P_a(x), P_b(x)$ のエントロピー $H(A), H(B)$ の関係は次のようになる.

$$H(B) = H(A) + \log \left(\frac{\sigma_b}{\sigma_a} \right) \quad (4)$$

従って, 残差のエントロピーは分散の減少と共に単調に減少し, 残差パワー σ^2 の最小化と残差エントロピーの最小化は等価であることが示される.

3.3 残差エントロピーの最小化規準

本節では, 残差エントロピーの最小化という規準で線形予測係数を決定する方法を述べる. 前節の議論により, フレーム内の残差パワーの最小化と残差エントロピーの最小化は等価であるから, 線形予測係数として以下の目的関数を最小化する係数を求めれば良いことがわかる.

$$\sum_{n=1}^N \epsilon^2[n] = \sum_{n=1}^N (f(x[n]) - f(\tilde{x}[n]))^2 \quad (5)$$

ただし, $\tilde{x}[n]$ はリニア PCM 領域での予測値を整数化した値で, N は定数で 1 フレームのサンプル数である. なお, この式を線形予測係数を決定するための目的関数として用いることについては, 先行研究でも示唆されている [2].

* “Linear predictive analysis for lossless coding of logarithmic companded signal” by HORI Yutaka (Faculty of Engineering, The University of Tokyo), MORIYA Takehiro, HARADA Noboru, KAMAMOTO Yutaka (NTT Communication Science Labs.), ONO Nobutaka, SAGAYAMA Shigeki (Graduate School of Information Science and Technology, The University of Tokyo)

3.4 目的関数の近似による効率的な解法

本節では、式 (5) を近似することにより、反復アルゴリズム等を用いずに最適に近い線形予測係数を求める方法について述べる。

$x[n] - \tilde{x}[n] \simeq 0$ という仮定が成り立つとき、式 (5) を以下のように近似できる。

$$(5) = \sum_{n=1}^N \left(\frac{f(x[n]) - f(\tilde{x}[n])}{x[n] - \tilde{x}[n]} \right)^2 (x[n] - \tilde{x}[n])^2 \simeq \sum_{n=1}^N (\tilde{W}(x[n]))^2 (x[n] - \tilde{x}[n])^2 \quad (6)$$

ただし、 $\tilde{W}(\cdot)$ は関数 $f(\cdot)$ の一階の微分である。式 (6) の導出時の仮定から予測値 $\tilde{x}[n]$ と $x[n]$ の差の絶対値が大きくなるにつれて、一般に近似の精度は悪くなる。そこで、近似誤差の影響を小さくするために、リニア PCM 領域での予測残差 $e[n] := x[n] - \tilde{x}[n]$ の振幅分布に仮定をおき、式 (6) において $\tilde{W}(\cdot)$ をそのまま用いる代わりに、 $\tilde{W}(\cdot)$ の期待値 $W(x) := E[\tilde{W}(x)]$ を用いる。すなわち $W(x)$ を以下のように定義する。

$$W(x) := \int_{-\infty}^{\infty} \Omega(z; x, \sigma) \tilde{W}(z) dz \quad (7)$$

ただし、 z はリニア PCM 領域の振幅値であり、本稿の実験では $\Omega(z; x, \sigma)$ として平均 x 、分散 σ^2 のガウス確率密度関数を用いた。

以上の議論により、式 (5) の近似式として

$$J = \sum_{n=1}^N (W(x[n]))^2 \left(x[n] - \sum_{i=1}^p a_i x[n-i] \right)^2 \quad (8)$$

が導かれる。式 (8) は $\{a_i\}_{i=1}^p$ に関して凸関数であるから、 J を最小化する線形予測係数は $\partial J / \partial a_j = 0$ ($j = 1, 2, \dots, p$) の解である。これを計算すると、線形予測係数は p 次の連立方程式の解となり、Cholesky 分解を用いて時間計算量 $\mathcal{O}(p^3)$ で解くことが可能である。

なお、提案法において $W(x)$ を定数とおく場合は、よく知られている共分散法 [3] となる。

4 評価実験

4.1 実験方法

提案法の性能評価のために実験を行った。実験では、式 (5) で表される 1 フレームあたりの二乗残差総和の平均 (実験 1) と、Golomb-Rice 符号化による残差の符号長の平均 (実験 2) をそれぞれ $p = 2, 8, 15, 25$ の場合について求めた。比較のために、共分散法 [3] でも同内容の実験を行った。実験データは、サンプリング周波数が 8kHz で、中間に無音のポーズを挟んだ 2 文章からなる 8 秒間の音声に対数圧伸符号化したものを用いた。

4.2 結果と考察

Fig.1 の横軸は実験データのインデックスである。

実験 1 ではほぼ全てのデータに対して、予測次数に依らず提案法が共分散法よりも小さくなった。実験 2

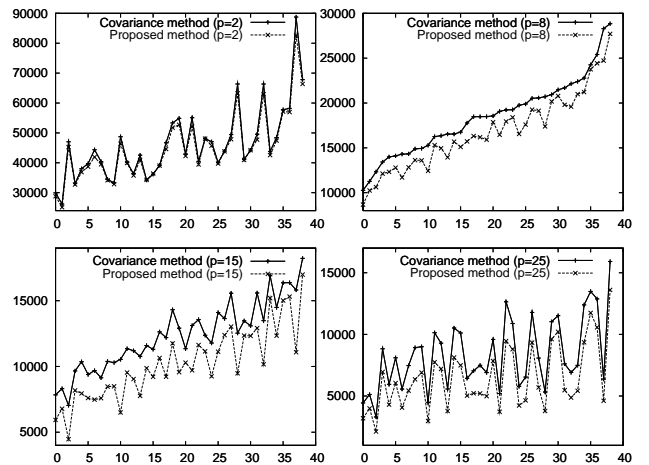


Fig. 1 フレームごとの二乗残差総和の平均

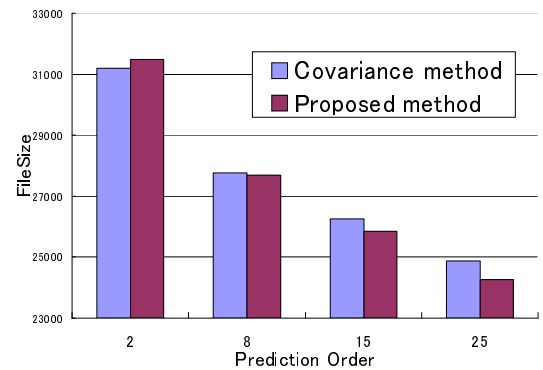


Fig. 2 Golomb-Rice 符号化後の残差の符号長の比較

では、 $p = 15, 25$ の場合に、全てのデータに対して共分散法よりも符号長が小さくなった。一方、 $p = 2, 8$ の場合には提案法の優位性は確認できなかった。

本稿で目的関数とした実験 1 の値と、実験 2 の符号長が $p = 2, 8$ の場合に必ずしも整合しなかった原因として、予測次数が小さい場合に近似の仮定が成り立たない可能性や提案法の残差の頻度分布が Golomb-Rice 符号の仮定する分布と異なる可能性などが考えられる。

5 おわりに

本稿では、対数圧伸符号化された音声信号をロスレス符号化するための線形予測分析法を提案した。評価実験により、予測次数が大きい場合は提案法がロスレス符号化に有効であることが確認された。予測次数が小さい場合についても今後さらに考察を行い検討をすすめる予定である。

参考文献

- [1] ITU-T Recommendation G.711, "Pulse code modulation (PCM) of voice frequencies," ITU-T, 1993.
- [2] Ghido, F., Tabus, I., "Accounting for companding nonlinearities in lossless audio compression," *Proc. ICASSP*, vol. 1, pp. 261-264, 2007.
- [3] 守谷健弘, 「音声符号化」, 電子情報通信学会, 1998.