

Golomb-Rice 符号化のための最適スペクトル包絡表現*

☆杉浦亮介¹, 鎌本優², 原田登², 亀岡弘和², 守谷健弘² (¹東大院 情報理工, ²NTT)

1 はじめに

近年, 携帯電話等の音声通信に用いるための音声音響符号化方式として, 入力音源に合わせて量子化・圧縮を行う領域を適応的に変える手法が有力視されており, 3GPP Extended Adaptive Multi-Rate WideBand (AMR-WB+) や MPEG-D Unified Speech and Audio Coding (USAC), 3GPP Enhanced Voice Services (EVS) 等でこの考え方が採用されている [1, 2, 3]. ここで用いられる領域は大きく二つある. 時間領域での符号化は音声の統計的性質を利用して音声に特化しており, 周波数領域での符号化は聴覚的な性質を利用して音声以外の音, 特に音楽のように様々な音が混在する音源に対して有効である. 故に上記の領域の切り替えの枠組みにおいては, 音声信号は時間領域で, それ以外の信号は周波数領域で主に符号化することにより, どのような音源が入力されても音質良く圧縮できると期待されている. しかしながら, 周波数領域での符号化はその性質上, 遅延を許さなければ音質を上げ難く, 音声通信のように低遅延が要求される条件下では特に効率よく信号を表現する必要がある. そこで本研究では, 周波数領域での低遅延な符号化である Transform Coded eXcitation (TCX) に焦点を当て, 高圧縮・低遅延の条件下での音質向上を目指す.

TCX ではスペクトルを量子化した後, その包絡情報を利用してエントロピー符号化を行い圧縮する. そこで本稿では, そのエントロピー符号化に合わせて最適化されたスペクトル包絡の表現法と, その包絡を低演算量で抽出する方法を提案する.

2 周波数スペクトルの符号化

本稿では, TCX の中でも修正離散コサイン変換 (MDCT) で表されるスペクトルを量子化し, エントロピー符号化するもの [4] を基に議論を行う.

この方式では, マスキング効果を利用して聴覚的な量子化誤差を小さく抑えるために, 各周波数 $k(0 \leq k \leq N-1)$ において線形予測係数 $\{a_n\}$ から算出されるスペクトル包絡を平滑化した値に比例した

$$w_k s = s |1 + \sum_n \gamma^n a_n e^{-j \frac{\pi n k}{N}}|^{-1}, \quad (0 < \gamma < 1) \quad (1)$$

を量子化幅としてスカラー量子化を行う. ここで $\{w_k\}$ が平滑化された包絡の形状を示しているが, これは人

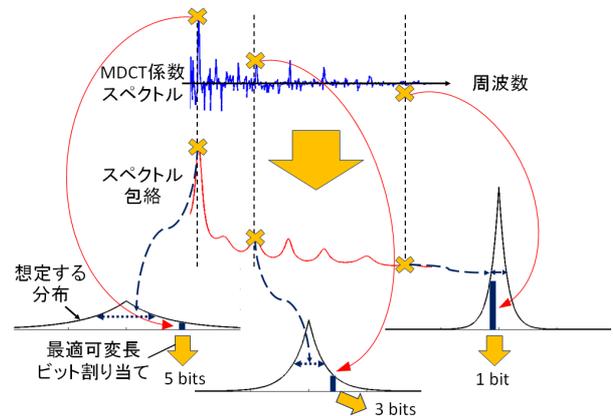


Fig. 1 包絡情報を用いたエントロピー符号化のイメージ.

の聴覚のマスキング閾値を大まかに近似しているとされており, $\gamma = 0.92$ を用いることで聴感上の歪の小さい量子化ができることが実験的に知られている [1, 2, 3, 4].

量子化後はこれをエントロピー符号化により圧縮するが, そこで包絡の情報

$$H_k = \sigma |1 + \sum_n a_n e^{-j \frac{\pi n k}{N}}|^{-1} \quad (2)$$

を用い, 図 1 のように包絡の値が大きい周波数では分散の大きい分布を, 値が小さい周波数では分散の小さい分布を各 MDCT 係数にそれぞれ仮定し, その仮定に対して最適なビットをそれぞれ割り当てることで効率よく圧縮を行っている. ただし, σ は線形予測分析の予測利得を表す.

この手法は包絡の情報を用いない場合と比べて効果はあるものの, 従来の線形予測分析で得られる包絡はこのエントロピー符号化に用いるために設計されているわけではないため, 最適な圧縮を実現するとは言えない. そこで本稿では, エントロピー符号化として Golomb-Rice 符号 [5] を使用し, Golomb-Rice 符号に対して最適な包絡を得る手法を導出する.

3 Golomb-Rice 符号長に基づく包絡表現

3.1 符号長最小化問題の定式化の指針

Golomb-Rice 符号とは, 非負の整数である Rice パラメータ r によってビット割り当ての仕方を変えるこ

*"Optimal representation of spectral envelope for Golomb-Rice coding." by Ryosuke Sugiura¹, Yutaka Kamamoto², Noboru Harada², Hirokazu Kameoka² and Takehiro Moriya² (¹The University of Tokyo, ²NTT).

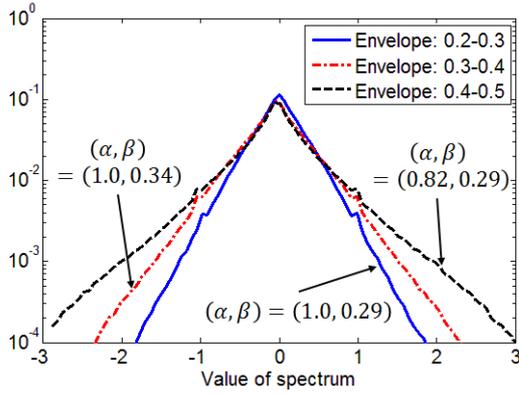


Fig. 2 包絡の値がそれぞれ0.2から0.3, 0.3から0.4, 0.4から0.5をとる周波数でのMDCT係数のヒストグラム。 α, β は[6]により推定した一般化ガウス分布の形状パラメータと分散。 $\alpha = 1$ ならラプラス分布。

とのできる可変長符号であり、ラプラス情報源に対して最適な符号であることが知られている[5]。図2にあるように、符号化の対象であるMDCT係数はラプラス分布に従っていると考えると差し支えないことから、以下ではGolomb-Rice符号について議論を行う。

Golomb-Rice符号においてはRiceパラメータが図1の分布の分散に対応しており、前節のエントロピー符号化の枠組みは、各周波数においてRiceパラメータを包絡値に従って変えることに相当する。このGolomb-Rice符号の符号長最小化の意味で最適な包絡表現を求める際に考えなければならない要素は、1.) 包絡のモデル、2.) 包絡とRiceパラメータの関係、3.) 包絡の抽出法、の3点である。1.), 2.) を決めれば3.) に対応する最適化問題が立てられるが、包絡のモデルの決め方によってはその最適化問題を解くのが難しくなり、演算量がかかりすぎる可能性もある。そこで本稿では、包絡の抽出法が従来の線形予測分析の手法となるべく変わらないように1.), 2.) を決めつつ、Golomb-Rice符号の符号長を最小化する手法を導く。そこで重要となるのが、線形予測分析の周波数領域での解釈[7]である。線形予測分析は周波数領域では、信号のパワースペクトルと式(2)で表される全極型の包絡の二乗値との板倉斎藤距離を最小化する係数 $\{a_n\}$ を求める問題と等価であり、この係数 $\{a_n\}$ はLevinson-Durbinアルゴリズムにより高速に求められる。つまり、包絡の抽出に対応する最適化問題を全極型の包絡との板倉斎藤距離最小化の形に帰着できれば、その問題は線形予測分析同様、簡単に解くことができる。

3.2 包絡表現法と符号長の最小化

各周波数 k におけるMDCT係数を x_k 、それを量子化したものを $y_k (= x_k / (w_k s))$ 、Riceパラメータを r_k

とすると、1フレームでのGolomb-Rice符号の符号長の総和は

$$\begin{aligned} & \mathcal{L}(\{y_k\}|\{r_k\}) \\ & \approx \sum_k (1 + r_k + \frac{|y_k|}{2^{r_k}}) = \sum_k (1 + \log_2 2^{r_k} + \frac{|y_k|}{2^{r_k}}) \\ & = (\log_2 e) \sum_k \left(\frac{|y_k|}{(\log_2 e) 2^{r_k}} - \ln \frac{|y_k|}{(\log_2 e) 2^{r_k}} - 1 \right) \\ & \quad + N(1 + \log_2 \ln 2 + \log_2 e) + \sum_k \log_2 |y_k| \\ & = (\log_2 e) \sum_k D_{IS}((\log_2 e) 2^{r_k} || y_k) + C(\{y_k\}) \quad (3) \end{aligned}$$

と表すことができる。ただし、ここでは丸めは無視し、正負符号は別途符号化するものとする。また、 $D_{IS}(X|Y) = Y/X - \ln Y/X - 1$ は X の Y からの板倉斎藤距離を表す。このように、符号長のRiceパラメータに依存する項を板倉斎藤距離の形に式変形することができたので、Riceパラメータを式(2)の H_k を用いて

$$r_k \equiv \log_2 \left(\frac{(\ln 2) H_k^2}{w_k s} \right) \quad (4)$$

とモデル化することにより、符号長最小化問題は

$$\begin{aligned} & \arg \min_{\sigma, \{a_n\}} \mathcal{L}(\{y_k\}|\{r_k\}) \\ & = \arg \min_{\sigma, \{a_n\}} \sum_k D_{IS}(H_k^2 / (w_k s) || |y_k|) \\ & = \arg \min_{\sigma, \{a_n\}} \sum_k D_{IS}(H_k^2 / (w_k s) || |x_k| / (w_k s)) \\ & = \arg \min_{\sigma, \{a_n\}} \sum_k D_{IS}(H_k^2 || |x_k|) \quad (5) \end{aligned}$$

のように全極モデルの二乗とスペクトルの絶対値との板倉斎藤距離最小化問題に帰着させることができ、この解は、入力信号のスペクトルの絶対値をパワースペクトルとして持つ信号の自己相関関数に相当するもの、つまりスペクトルの絶対値の逆フーリエ変換に対してLevinson-Durbinアルゴリズムを適用することにより得られる。そしてこの際、スペクトル包絡のモデルは $\{H_k\}$ ではなく $\{H_k^2\}$ となる。つまり、本手法は包絡のモデルを $\{H_k\}$ から $\{H_k^2\}$ に変えることによって、Golomb-Rice符号に対して最適な包絡の抽出を高速解法により実現させるものである。

また、量子化幅 $w_k s$ はスペクトルの大まかな山谷を反映させたもので構わないことから、上記で得られた包絡を平滑化した

$$w_k s = s |1 + \sum_n \gamma^n a_n e^{-j \frac{\pi k}{N} n}|^{-2} \quad (6)$$

を用いて近似する。

この手法は、従来の線形予測分析と比較し、包絡のモデルが $\{H_k\}$ から $\{H_k^2\}$ に置き換わったことと、ス

ペクトルの二乗の代わりに絶対値を逆フーリエ変換して自己相関関数を得るところの違いしかないため、特に MDCT 係数をスペクトルとして使用する場合には演算量の増加はほとんどない。また、モデル係数の安定性も従来のものと同様、Levinson-Durbin アルゴリズムの中で保証することができる上、量子化も線スペクトル対 (LSP) [8] の形で行うことができる。

4 実験と結果

4.1 圧縮性能比較

前節の提案手法を評価するため、従来の TCX [4] をベースとした符号化器を作成し、Golomb-Rice 符号の対象を固定として、スペクトル包絡の抽出・表現法に従来の線形予測分析と提案法を使用した時の圧縮率の比較を複数のビットレートにて行った。この符号化器は音声通信を想定し、32 ms の遅延で符号化を行うように設計した (cf. AMR-WB+ は遅延 144 ms)。Golomb-Rice 符号の対象となる量子化スペクトルは、従来の設定でテストデータを符号化し、その結果の量子化スペクトルを用いた。また公平のため、量子化の重み $\{w_k\}$ はどの条件でもこの符号化の際に用いたもので固定した。テストデータは、RWC 音楽データベース [9] 内から無作為に選んだ 50 曲の中からそれぞれ 10 秒を切り出し、16 kHz にダウンサンプリングしたものを使用した。

この結果が図 3 である。各ビットレートにおいて左から順に、各フレームにおいて全ての周波数で同じ Rice パラメータを用いた場合、Rice パラメータを線形予測分析による包絡から求めた場合、Rice パラメータを提案手法による包絡から求めた場合の圧縮率であり、100 % は全ての周波数で最適な Rice パラメータを用いた時の理想的な記述長を表す。図のとおり、線形予測分析の包絡の情報を用いた Golomb-Rice 符号は条件によっては、Rice パラメータとしてフレーム毎に固定値を用いる場合よりも圧縮効率が良いが、線形予測分析はこの使い道に対して最適化されているわけではないので逆効果になる場合も見受けられる。これに対して、提案法による Golomb-Rice 符号はどのビットレートにおいても高い圧縮効率を示した。

4.2 音質の客観評価

続いて、上記と同じテストデータを用い、16 kbps における音質の客観評価を行った。作成した符号化器が十分な性能を持っていることを示すため、まずは Perceptual Evaluation of Audio Quality (PEAQ) [10] で算出した客観評価値を AMR-WB+ のものと比較した。その結果が図 4 であり、作成した符号化器が AMR-WB+ よりも高い評価値が得られていることが

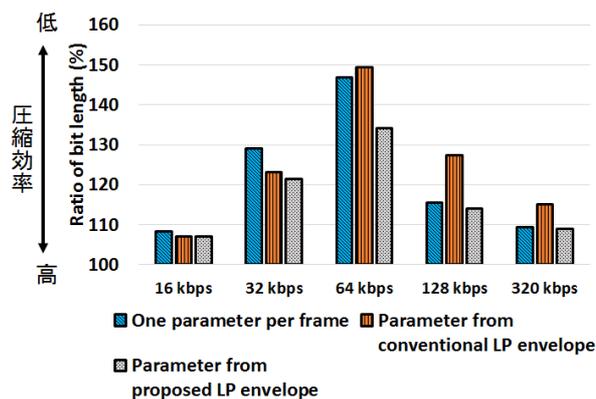


Fig. 3 Rice パラメータの割り当てによる平均記述長の比較。100 % は全ての周波数において最適な Rice パラメータを割り当てた時の平均記述長を表す。包絡の次数は 16。

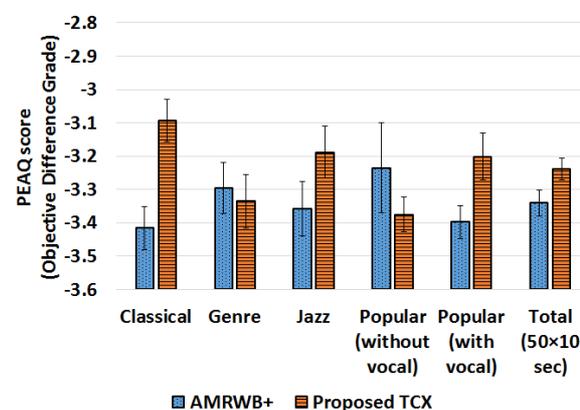


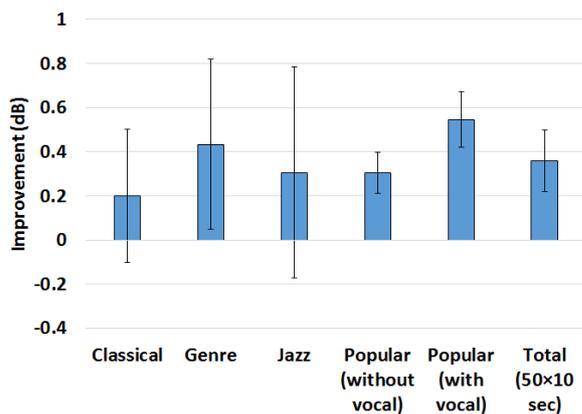
Fig. 4 作成した TCX と AMR-WB+ とのデータベース毎の PEAQ 値比較。平均と 95 % 信頼区間。

わかる。

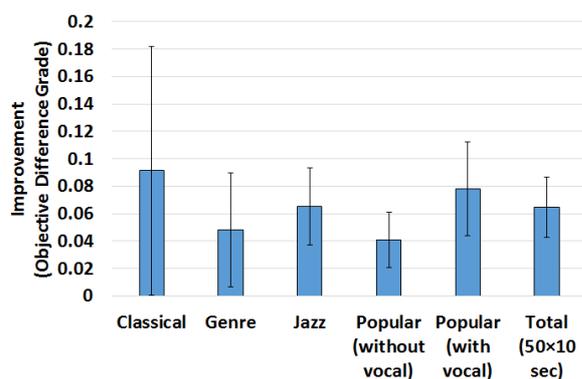
次に、作成した符号化器で線形予測分析を用いた場合と提案手法を用いた場合で量子化スペクトルの SNR と PEAQ の評価値を比較した。図 5 は提案法を導入したことによる各評価値の上昇量を表したものである。提案手法により、Golomb-Rice 符号の効率が上がり、その分同じビットレートなら量子化を細かく行えるようになって、各評価値が上昇したことが確認できる。

4.3 音質の主観評価

最後に被験者 6 名の AB テストによる音質の主観評価を [11] と同様の手法で行った。RWC 音楽データベース内の 5 曲から各 10 秒を切り出し、16 kHz にダウンサンプリングしたものを上記の符号化器で提案手法を導入したものとしていないもので 16 kbps に圧縮し、被験者にそれぞれブラインドで提示した。た



(a) 量子化スペクトルの SNR.



(b) PEAQ 値.

Fig. 5 線形予測分析の代わりに提案手法を用いた場合の各評価値の上昇量. 平均と 95 % 信頼区間.

だし、提案法を入れた場合については包絡の精度を上げれば、圧縮の効率を必ず上げることが保証されているので、スペクトルの圧縮に使用するビット数の中から 3 ビットを調波成分における包絡値の補正に割り当てた。被験者は参照音と比較して音質の高いと思うほうに対して、その相対的な音質を 0, 1, 2 点で点数をつけ、それぞれの音源について評価した。

図 6 が実験の結果である。各音源において優位差は現れなかったが、合計では提案法が 5 % 優位水準で有意差をもって高い評価を得られた。これにより、提案法が主観音質も向上させる傾向があることが確認できた。

5 おわりに

本稿では、Golomb-Rice 符号に対して最適なスペクトル包絡の表現法について提案を行った。本手法は従来の包絡モデルから少し変更を加えることにより、線形予測分析と同様なアルゴリズムでもって Golomb-Rice 符号の符号長を最小化させる包絡の抽出を可能としている。Golomb-Rice 符号の比較実験の結果か

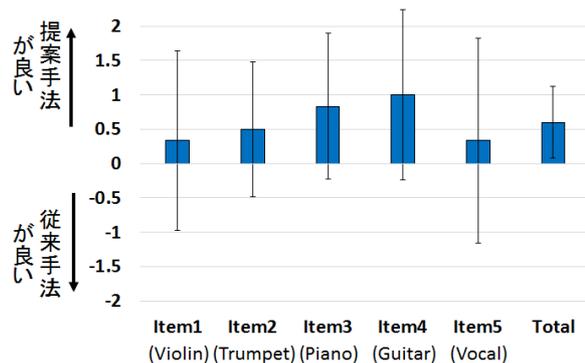


Fig. 6 主観評価実験の結果. -2 点 (従来法の音質が良い) から 2 点 (提案法の音質が良い) による評価. 平均と 95 % 信頼区間.

ら、提案法は演算量の増加がほとんどないにも関わらず、圧縮効率が線形予測分析を用いた場合に比べて向上していることがわかった。また、客観評価と主観評価により符号化器の音質向上の傾向があることも確認した。以前発表した、周波数解像度の伸縮された包絡表現 [12] との組み合わせにより、より圧縮の効率が上げられるかを検討することが今後の課題の一つである。

参考文献

- [1] 3GPP TS 26.290 Release 11, 3GPP, 2012.
- [2] ISO/IEC 23003-3:2012, Part 3.
- [3] 3GPP TS 26.441 Release 12, 3GPP, 2014.
- [4] G.Fuchs, *et al.*, EUSIPCO, pp.1264-1268, 2009.
- [5] R. F. Rice, JPLTR, vol. JPL-79-22, JPL-83-17, JPL-91-3, 1979, 1983, 1991.
- [6] S. Mallat, IEEE Trans. Pattern Anal. Mach. Intell., vol. 11, pp. 674-693, 1989.
- [7] 板倉, 博士論文, 名大院工学研究科, 1972.
- [8] 菅村, 板倉, 電子通信学会論文誌, J64-A, 8, pp.599-606, 1981.
- [9] [Online]. Available: <https://staff.aist.go.jp/m.goto/RWC-MDB/> ('14 年 12 月現在).
- [10] [Online]. Available: <http://www-mmsp.ece.mcgill.ca/Documents/Software/Packages/AFsp/AFsp.html> ('14 年 12 月現在).
- [11] M. Orger, *et al.*, ICASSP, pp. IV-545-548, 2007.
- [12] 杉浦, *et al.*, 日本音響学会講演集, 1-6-8, pp. 281-284, 2014.