日本音響学会誌 71 巻 11 号（2015），pp. 624–631

小特集──スパース表現に基づく音響信号処理──

スパース表現に基づく音声音響符号化*

亀岡弘和（日本電信電話株式会社/東京大学）**・鎌本　優，杉浦亮介（日本電信電話株式会社）***

43.60.Ek

1. はじめに

情報圧縮の概念は言語でたとえることができる。例えば「I」や「is」のように日常で頻繁に使用する単語は，「algorithm」のように余り使用しない単語に比べて通常は短い。このため，日常会話の文は平均的に長くならずに済んでいる。これと同様，頻繁に起こるイベントには短い符号長，滅多に起こらないイベントには長い符号長の符号語を割り当てることでデータ全体を短い符号長で効率的に表現することができる。このようにデータ中の各イベントの出現確率に応じて符号長を決定する符号化方式の枠組をエントロピー符号化といい，Huffman 符号，算術符号，後述の Golomb-Rice 符号 [1, 2] などがその例である。

線形予測符号化（Linear Predictive Coding; LPC）に基づく音声音響信号の可逆圧縮符号化 [3, 4] では，まず線形予測分析により所与の信号の予測誤差を算出し，その際求まる予測係数と共に予測誤差を量子化（整数値化）及び符号化して伝送した後，受信側で復号化し，元の信号を復元する方式が取られる。線形予測分析により得られる予測誤差の振幅は 0 付近に集中する傾向にあるため，予測誤差の符号化にエントロピー符号化を用いることで全体の符号長を抑えられる点がこの方式の特徴である。また，Transform Coded eXcitation（TCX）と呼ぶひずみのある周波数領域符号化 [5] では，音響信号のスペクトルを量子化し，エントロピー符号化を用いて各周波数成分を符号化する方式が取られる。対象とする音源の種類や区間によって周波数成分の確率分布は帯域によって異なるため，線形予測分析により得られるスペクトル包絡の値を（実際には未知の）各帯域の周波数成分の分散と見なすことによりエントロピー符号化で各成分を効率的に圧縮できる点がこの方式の特徴である。

本稿では，エントロピーと符号化の関係について略説した上で，上記の LPC による時間領域及び周波数領域の音声音響信号符号化をエントロピー最小化問題として捉えた高効率な符号化アプローチを紹介する。また，エントロピー符号化に Golomb-Rice 符号を用いる場合には符号化対象に対しスパース性が暗に仮定されることを示し，エントロピー符号化とスパース性の関係についても言及する。

2. エントロピーと符号化

情報理論においては，イベントが起こる頻度の偏りを表すのにエントロピーという尺度が用いられる。エントロピーは対象とするイベントのランダムさを意味し，例えばイベント X を「整数値」とした場合は

$$H(X) = -\sum_{x\in\mathbb{Z}} p(x)\log p(x) \qquad (1)$$

と定義される。ただし，$p(x)$ は整数 x が出現する確率を表す。また，エントロピーは，符号化の際に必要な平均符号長の理論的な下限を表すことが知られている。つまり，x に対して割り当てる符号語の長さを $-\log p(x)$ とするのが最も効率的な符号化ということを意味する。

式 (1) よりエントロピーは x が従う確率分布 $p(x)$ に依存することが分かる。そこで，$p(x)$ がどういう分布のときにこの値が大きくなる（又は小さくなる）かを考えよう。二つの確率密度関数 $p(x)$ と $f(x)$ との間の近さは Kullback-Leibler（KL）ダイバージェンス

* Speech and audio coding with sparse representations.

** Hirokazu Kameoka (Nippon Telegraph and Telephone Corporation, Atsugi, 243–0198/The University of Tokyo) e-mail: kameoka.hirokazu@lab.ntt.co.jp

*** Yutaka Kamamoto and Ryosuke Sugiura (Nippon Telegraph and Telephone Corporation)

$$\mathrm{KL}(p\|f) := \sum_{x\in\mathbb{Z}} p(x)\log\frac{p(x)}{f(x)} \tag{2}$$

で測ることができる。この規準は p と f が一致する場合にのみ 0 となり，p が f から離れれば離れるほど大きい値をとる。今，$f(x)$ を一様分布 $f(x)\propto 1$ とすると，$\log f(x)$ は定数，$\sum_x p(x)=1$ であるから式 (2) は

$$\mathrm{KL}(p\|f) = \sum_{x\in\mathbb{Z}} p(x)\log p(x) + \text{const.} \tag{3}$$

となり，定数項を除けば負のエントロピーと等しくなる。$\mathrm{KL}(p\|f)$ は $p(x)$ が一様分布に近いほど小さい値をとるので，逆にエントロピーは対象とする確率変数が従う確率分布が一様分布に近いほど大きな値をとる。一様分布に従うということはすなわちランダムであるということなので，確かにエントロピーはイベントのランダムさを意味した尺度になっていることが分かる。X が正規分布に従う確率変数のとき，分散が小さいほど（$p(x)$ が一様分布から遠ざかることからも予想されるように）エントロピーは小さくなる。実際，分散 σ^2 の正規分布に従う確率変数のエントロピーは $\frac{1}{2}\log(2\pi e\sigma^2)$ で与えられ，確かに σ^2 が小さいほどエントロピーは小さくなることを示している。このことは，所与の信号が正規分布に従う系列ならば，分散の小さい別の表現に変換することで高効率な符号化が可能であることを意味する。LPC による時系列信号の可逆圧縮符号化方式は正にこの原理に基づくものである。

3. 線形予測分析

本章では線形予測分析の原理を概説する。所与の離散時間信号を $s_1, s_2, \ldots, s_T$ とする。線形予測分析は，時刻 t の信号の標本値 s_t を時刻 t より過去の標本値 $s_{t-1}, s_{t-2}, \ldots, s_{t-P}$ の線形結合で予測することを目的とし，

$$\mathcal{J}(\boldsymbol{a}) = \sum_t \left(s_t - \sum_{p=1}^{P} a_p s_{t-p}\right)^2 \tag{4}$$

を最小化する「予測係数」$\boldsymbol{a}=(a_1,\ldots,a_P)^{\mathsf{T}}$ を求める最適化問題として定式化される。この最適化問題は，$\boldsymbol{s}=(s_1,\ldots,s_T)^{\mathsf{T}}$ を自己回帰過程

$$s_t = \sum_{p=1}^{P} a_p s_{t-p} + \epsilon_t \tag{5}$$

$$\epsilon_t \overset{iid}{\sim} \mathcal{N}(\epsilon_t; 0, \sigma^2) \tag{6}$$

から生成された観測値系列と仮定した場合の $\boldsymbol{a}=(a_1,\ldots,a_P)^{\mathsf{T}}$ の最尤推定問題と等価である [6]。これを以下で確認する。ただし，$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) \propto \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$ とし，式 (6) は ϵ_t が独立に同一（平均 0，分散 σ^2）の正規分布に従うことを意味する。$\boldsymbol{\epsilon}=(\epsilon_1,\ldots,\epsilon_T)^{\mathsf{T}}$ とし，

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & & & & 0 \\ -a_1 & \ddots & & & \\ \vdots & \ddots & \ddots & & \\ -a_P & & \ddots & \ddots & \\ & \ddots & & \ddots & \ddots \\ 0 & & -a_P & \cdots\ -a_1 & 1 \end{bmatrix} \tag{7}$$

と置くと，式 (5) は

$$\boldsymbol{\Psi}\boldsymbol{s} = \boldsymbol{\epsilon} \tag{8}$$

のように書ける。$\boldsymbol{\Psi}$ は逆行列を持つので，$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ 及び式 (8) より

$$\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{s}; \boldsymbol{0}, \sigma^2\,\boldsymbol{\Psi}^{-1}\,\boldsymbol{\Psi}^{-\mathsf{T}}) \tag{9}$$

が言える。$|\boldsymbol{\Psi}|=1$ より，$\boldsymbol{s}=(s_1,\ldots,s_T)^{\mathsf{T}}$ が観測された下での $\boldsymbol{a}$ の対数尤度は

$$\log p(\boldsymbol{s}|\boldsymbol{a}) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_t \left(s_t - \sum_{p=1}^{P} a_p s_{t-p}\right)^2 \tag{10}$$

となり，$\boldsymbol{a}$ によらない項を除けば式 (4) の正負を逆転したものと等しい。以上より確かに式 (4) を $\boldsymbol{a}$ に関して最小化することと式 (10) を $\boldsymbol{a}$ に関して最大化することは等価であることが分かる。

$\mathcal{J}(\boldsymbol{a})$ を $a_1,\ldots,a_P$ に関してそれぞれ偏微分して 0 と置き，連立させると，

$$\begin{bmatrix} r_{1,1} & \cdots & r_{1,P} \\ \vdots & \ddots & \vdots \\ r_{P,1} & \cdots & r_{P,P} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} r_{0,1} \\ \vdots \\ r_{0,P} \end{bmatrix} \tag{11}$$

$$r_{q,p} = \sum_t s_{t-p}s_{t-q} \tag{12}$$

という形を得る。よって $\mathcal{J}(\boldsymbol{a})$ を最小化する $\boldsymbol{a}$ は上式を解くことで得られる。ここで，s_t が弱定常でエルゴード的であれば $v_{|p-q|}=r_{q,p}$ は s_t の自己相関関数となり，式 (11) は

$$\begin{bmatrix} v_0 & \cdots & v_{P-1} \\ \vdots & \ddots & \vdots \\ v_{P-1} & \cdots & v_0 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_P \end{bmatrix} \quad (13)$$

と書ける。この特殊な形の連立方程式を Yule-Walker 方程式といい，Levinson-Durbin アルゴリズムにより効率的に解くことができる。

以上で求めた $\hat{\boldsymbol{a}}$ を $\boldsymbol{\Psi}$ に代入すれば，式 (8) より $\boldsymbol{s}$ から予測誤差系列 $\boldsymbol{\epsilon}$ を得ることができる。逆に，予測係数 $\hat{\boldsymbol{a}}$ と予測誤差系列 $\boldsymbol{\epsilon}$ のペアから，

$$\begin{aligned} s_1 &= \epsilon_1 \\ s_2 &= \epsilon_2 + \hat{a}_1 s_1 \\ s_3 &= \epsilon_3 + \hat{a}_1 s_2 + \hat{a}_2 s_1 \\ &\vdots \end{aligned} \quad (14)$$

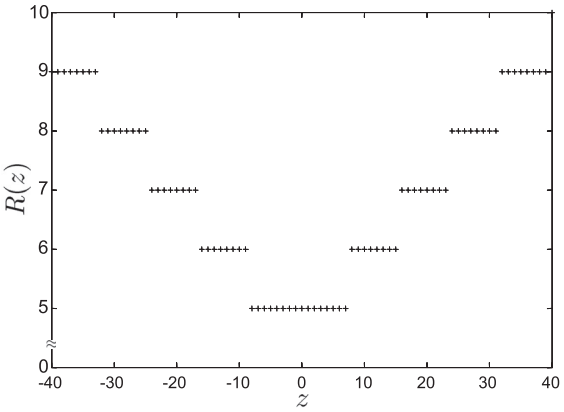
のように $\boldsymbol{s}$ を逐次的に復元することができる。通常，予測次数 P は信号の全体の長さ T に比べてかなり小さく設定されるので，$\hat{\boldsymbol{a}}$ に必要な符号長は $\boldsymbol{\epsilon}$ のそれに比べれば無視できるほど小さい。よって，$\boldsymbol{\epsilon}$ に必要な符号長が $\boldsymbol{s}$ のそれより十分小さければ元の情報を失うことなくデータを圧縮することができる。もし $s_1, \ldots, s_T$ が実際に正規分布に従う系列であれば，その線形変換である $\epsilon_1, \ldots, \epsilon_T$ もまた正規分布に従う系列となる。この場合，線形予測分析により得られる $\epsilon_1, \ldots, \epsilon_T$ は $s_1, \ldots, s_T$ に比べて分散が小さくなっているので，2 章で述べたようにエントロピー符号化で高効率に符号化することが可能である。

4. Golomb-Rice 符号長を規準とした時間領域符号化

4.1 動　　機

LPC による時系列信号の可逆圧縮符号化の国際標準 MPEG-4 Audio Lossless Coding（ALS）[3, 4] では $\boldsymbol{\epsilon}$ の符号化に Golomb-Rice 符号が採用されている。Golomb-Rice 符号は整数 z を除数 $r \in \mathbb{N}$ で除算した際の商と剰余をそれぞれ符号化したものであり，符号化と復号化の処理の計算量が小さく済むという特長を持つ。$\lfloor \cdot \rfloor$ を切捨て整数化演算子とすると，整数 z に対する Golomb-Rice 符号長 $R(z)$ は

$$R(z) = \begin{cases} \left\lfloor \dfrac{z}{2^{r-1}} \right\rfloor + r + 1 & (z \geq 0) \\ \left\lfloor \dfrac{-z-1}{2^{r-1}} \right\rfloor + r + 1 & (z < 0) \end{cases} \quad (15)$$



図–1　整数 z に対する Golomb-Rice 符号長 $R(z)$ $(r = 4)$

で与えられる。r を Rice パラメータという。2 章で述べたように，エントロピー符号化で x を符号化する場合，一般に x の符号長を $-\log p(x)$（ただし，$p(x)$ は x の出現確率）とするのが最も高効率であるので，Golomb-Rice 符号を用いる場合は，$R(x) = -\log p(x)$ のとき，すなわち（整数化演算子を無視すれば）$p(x)$ が平均が 0 の Laplace 分布のときに最も高効率となる。しかし従来の線形予測分析では $\boldsymbol{\epsilon}$ が正規分布に従うことを仮定して $\boldsymbol{a}$ を最尤推定するため，$\boldsymbol{\epsilon}$ の符号長を最短にする方法にはなっていなかった。従って，$\epsilon_1, \ldots, \epsilon_T$ の二乗和の代わりに直接 Golomb-Rice 符号長を最小化するように $\boldsymbol{a}$ を推定することができれば，標準準拠の拘束から外れることなく従来の線形予測分析より高い圧縮性能を達成できる可能性がある。亀岡らはこのような問題意識の下，予測誤差の Golomb-Rice 符号長を規準とした線形予測分析手法を提案している [7]。以下で，その方法を紹介する。

4.2 スパース性規準に基づく線形予測分析

$\boldsymbol{s} = (s_1, \ldots, s_T)^{\mathsf{T}}$ の生成プロセスとして，ϵ_t が Laplace 分布に従う場合

$$s_t = \sum_{p=1}^{P} a_p s_{t-p} + \epsilon_t \quad (16)$$

$$\epsilon_t \overset{iid}{\sim} \mathrm{Laplace}(\epsilon_t; 0, b) \quad (17)$$

を仮定した $\boldsymbol{a} = (a_1, \ldots, a_P)^{\mathsf{T}}$ の最尤推定問題を考える。ただし，$\mathrm{Laplace}(x; \mu, b) = \frac{1}{2b} \exp(-|x-\mu|/b)$ とし，式 (17) は ϵ_t が独立に Laplace 分布に従うことを意味する。このとき $\boldsymbol{\epsilon} \sim \prod_t \mathrm{Laplace}(\epsilon_t; 0, b)$ となり，式 (8) の関係式と $|\boldsymbol{\Psi}| = 1$ という事実を用いて，確率密度関数の

変数変換により

$$\boldsymbol{s} \sim \prod_t \mathrm{Laplace}([\boldsymbol{\Psi}\boldsymbol{s}]_t; 0, b) \tag{18}$$

が言える。ただし，$[\cdot]_x$ はベクトルの x 番目の要素を表す。従ってこの場合，所与の信号 $\boldsymbol{s}$ の下での $\boldsymbol{a}$ の対数尤度は

$$\log p(\boldsymbol{s}|\boldsymbol{a}) = -T\log(2b) - \sum_t \frac{1}{b}\left|s_t - \sum_{p=1}^{P} a_p s_{t-p}\right| \tag{19}$$

となり，$\boldsymbol{a}$ の最尤推定問題は予測誤差の絶対値和

$$\mathcal{J}(\boldsymbol{a}) = \sum_t \left|s_t - \sum_{p=1}^{P} a_p s_{t-p}\right| \tag{20}$$

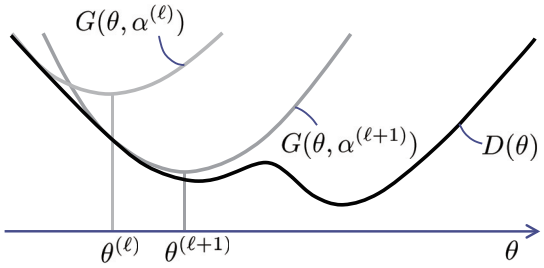
を $\boldsymbol{a}$ に関して最小化する問題に帰着する。すなわち，$\boldsymbol{\epsilon}$ の ℓ_1 ノルムが最適化規準となる。

ℓ_1 ノルムはスパースさを測る規準の一つであり，スパースなベクトルを得るための最適化や正則化の規準として様々な場面で応用されている。上述の目的関数は，予測誤差系列がスパースになるように予測係数を求めることで符号長を小さくすることができることを示している。

4.3 補助関数法の原理

式 (20) を最小化する $\boldsymbol{a}$ は解析的には得られないが，補助関数法と呼ぶ方法論 [8, 9] により式 (20) を単調減少させる反復アルゴリズムを導くことができる。詳細は後述するが補助関数法とは，目的関数値が丁度下界となっているような関数（補助関数と呼ぶ）を設計し，目的関数の代わりにその関数を反復的に降下させることで目的関数を間接的に降下させていく方法である。不完全データの下で確率モデルのパラメータを推定する方法として知られる Expectation-Maximization（EM）アルゴリズムは補助関数法の特殊ケースに相当する。EM アルゴリズムと同様，補助関数法はいかなる最適化問題にも適用可能というわけではないが，ある要件を満たす補助関数が設計できれば効率の良い最適化アルゴリズムが導ける場合がある。以下に，補助関数法の原理を示し，補助関数が満たすべき要件を示す。

定義 1 (補助関数)**.** θ をパラメータとする目的関数 $D(\theta)$ に対し，$D(\theta) = \min_\alpha G(\theta, \alpha)$ が成り立つとき，$G(\theta, \alpha)$ を $D(\theta)$ の**補助関数**と定義する。また，α を**補助変数**と呼ぶ。このとき，次の定理が成り立つ。



図–2 補助関数法によるパラメータ更新のイメージ

定理 1. 補助関数 $G(\theta, \alpha)$ を，α に関して最小化するステップと，θ に関して最小化するステップ

$$\alpha \leftarrow \operatorname*{argmin}_\alpha G(\theta, \alpha) \tag{21}$$

$$\theta \leftarrow \operatorname*{argmin}_\theta G(\theta, \alpha) \tag{22}$$

を繰り返すと，目的関数 $D(\theta)$ の値は単調減少する。

Proof. 反復計算のステップ数を ℓ とし，$\theta = \theta^{(\ell)}$，$\alpha = \alpha^{(\ell)}$ から $\theta = \theta^{(\ell+1)}$，$\alpha = \alpha^{(\ell+1)}$ に更新されたときに，$D(\theta)$ が増加しないことを示す。$\alpha^{(\ell+1)} = \operatorname{argmin}_\alpha G(\theta^{(\ell)}, \alpha)$ なので，補助関数の定義より $D(\theta^{(\ell)}) = G(\theta^{(\ell)}, \alpha^{(\ell+1)})$ である。また，式 (22) より，明らかに $G(\theta^{(\ell)}, \alpha^{(\ell+1)}) \geq G(\theta^{(\ell+1)}, \alpha^{(\ell+1)})$ である。更に補助関数の定義より $G(\theta^{(\ell+1)}, \alpha^{(\ell+1)}) \geq D(\theta^{(\ell+1)})$ であるから，結局，$D(\theta^{(\ell)}) \geq D(\theta^{(\ell+1)})$ である（図–2 参照）。 □

4.4 反復アルゴリズムの導出

以上より，次の 2 点を満たす補助関数を設計できれば補助関数法を適用することができる。

1) $\operatorname{argmin}_\alpha G(\theta, \alpha)$ が解析的に求められる。
2) $\operatorname{argmin}_\theta G(\theta, \alpha)$ が解析的に求められる。

そこで式 (20) の目的関数に対し，以上の要件を満たす補助関数を設計する。絶対値関数 f

$$f(z) = |z| \tag{23}$$

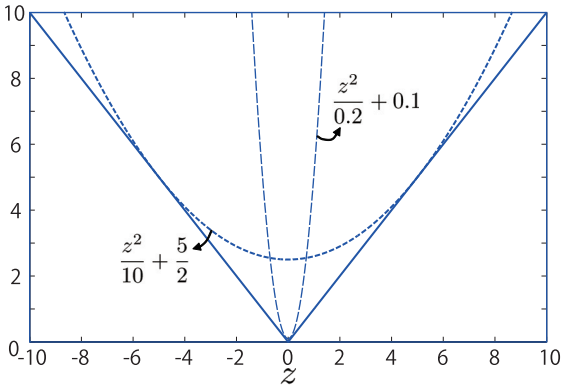
に対し，任意の w の下で $f(z)$ に $z = \pm w$ で接する 2 次関数

$$h(z) = \frac{z^2}{2|w|} + \frac{|w|}{2} \tag{24}$$

は $f(z)$ を下回らない（図–3 参照）ため，

$$|z| \leq \frac{z^2}{2|w|} + \frac{|w|}{2} \tag{25}$$

が成り立つ（等号は接点 $z = w$ において成立する）。この不等式を式 (20) に当てはめると，



図–3　絶対値関数とそれに接する放物線（$w=0.1,5$）

$$\mathcal{J}(\boldsymbol{a}) \le \sum_{t=1}^{T} \frac{1}{2|w_t|}\left(s_t - \sum_p a_p s_{t-p}\right)^2 + d \equiv \mathcal{I}(\boldsymbol{a},\boldsymbol{w}) \tag{26}$$

のような不等式が立てられる。ただし，$\boldsymbol{w} = \{w_1,\ldots,w_T\}$ とし，$d = \sum_{t=1}^{T}|w_t|/2$ である。$\mathcal{J}(\boldsymbol{a}) = \mathcal{I}(\boldsymbol{a},\boldsymbol{w})$ となる $\boldsymbol{w}$ は明らかに，

$$w_t = s_t - \sum_p a_p s_{t-p} \tag{27}$$

である。よって $\mathcal{J}(\boldsymbol{a}) = \min_{\boldsymbol{w}} \mathcal{I}(\boldsymbol{a},\boldsymbol{w})$ より以上の $\mathcal{I}(\boldsymbol{a},\boldsymbol{w})$ は補助関数の定義及び上述の要件 1 を満たす。次にこの補助関数が要件 2 を満たすことを確認する。$\boldsymbol{w}$ 固定のときに $\mathcal{I}(\boldsymbol{a},\boldsymbol{w})$ を最小化する $\boldsymbol{a}$ は，

$$\begin{bmatrix} r'_{1,1} & \cdots & r'_{1,P} \\ \vdots & \ddots & \vdots \\ r'_{P,1} & \cdots & r'_{P,P} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} r'_{0,1} \\ \vdots \\ r'_{0,P} \end{bmatrix} \tag{28}$$

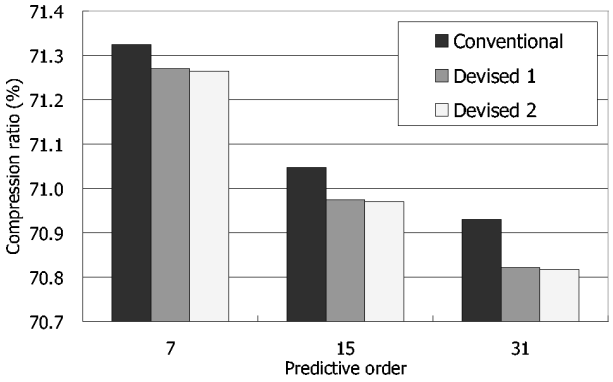
$$r'_{q,p} = \sum_t \frac{1}{2|w_t|} s_{t-p} s_{t-q} \tag{29}$$

を解くことにより得られる。式 (28) の左辺の行列は正定値対称行列であるので，$\boldsymbol{a}$ は Cholesky 分解を用いて求めることができる。以上より，$\mathcal{I}(\boldsymbol{a},\boldsymbol{w})$ は補助関数の要件を満たし，次の手順により，予測係数の絶対誤差最小解を得ることができる。

1) $\boldsymbol{a}$ を初期値設定する
2) $\boldsymbol{w}$ を式 (27) により更新する
3) $\boldsymbol{a}$ を式 (28) の解に更新し，2) に戻る

式 (20) は $\boldsymbol{s}$ に関して凸なので，収束解は大域最適解に一致する。

式 (26) は $1/|w_t|$ で重み付けされた二乗誤差を表しており，式 (27) より，式 (28) の解は 1 ステップ前に算出された予測誤差の絶対値が大きい時刻には小さい重みを，絶対値が小さい時刻には大きい重みを課す重みつき最小二乗誤差推定量に相当している。このため，以上の $\boldsymbol{w}$ と $\boldsymbol{a}$ の更新ステップは予測誤差を徐々にスパースにしていく（0 に近い値をより 0 に近づけていく）効果がある。



図–4　二乗誤差規準の線形予測分析（従来法）と Golomb-Rice 符号長を規準とした線形予測分析（提案法 1&2）による圧縮率の比較

4.5 提案法の効果

RWC 研究用音楽データベース [10] に含まれる，サンプリング周波数 44.1 kHz，16 ビットで収録されたステレオの音楽ファイル（WAV 形式）10 個（RWC-MDB-P-2001 No.33〜42）を実験データ（合計約 432 MB）として，提案法の圧縮性能の評価実験を行った。Levinson-Durbin 法に基づく二乗誤差規準の LPC（従来方式）を**従来法**，従来法で得られた予測係数を初期値として 4.4 節で述べた反復更新を 10 回行ったものを**提案法 1** と呼ぶ。また，式 (23) を $z=0$ で微分可能な $f(z)=\sqrt{z^2+\beta^2}$ に置き換えた目的関数も同様のアルゴリズムで最小化することができる [7]。これを**提案 2** と呼ぶ。従来法，提案法 1，提案法 2 それぞれについての，予測次数が 7, 15, 31 の場合の圧縮率を図–4 に示す。実験結果より，提案法は初期値から 10 回程度の反復計算で単調に収束することが確認され，従来法に比べわずかではあるが高い圧縮率を得た。

5. Golomb-Rice 符号長を規準とした周波数領域符号化

5.1 背景と動機

前章では音響信号の時間領域符号化のアプローチ例を紹介したが，本章では周波数領域符号化のアプローチ例を紹介する。前章で述べた手法は，所与の信号をスパースな表現（予測誤差）へ変換

するための変換パラメータ（予測係数）を求めることで符号長を短くするものであったのに対し，本章で述べる手法は，音響信号のスペクトルがスパースになる性質を利用する。

周波数領域での符号化方式では人間の聴覚心理学上の特性を利用して情報の圧縮を行える点が特色であり，音声信号に限らず音楽など一般的な音響信号の符号化に効果的である。例えばある周波数の音によりその近くの周波数の小さな音が聞こえにくくなる人間の聴覚のマスキング特性を利用し，聞こえにくい成分に短い符号を割り当てることで音質の劣化をほとんど感じさせることなくデータを圧縮することができる。この方式は，元の情報を一部捨てるため大幅なデータ圧縮が可能である一方で不可逆な符号化となる。また，高い周波数解像度のスペクトルを算出するためには長い分析窓を取らざるを得ないため，時間領域の符号化に比べて遅延が大きい場合が多い。その中で，TCX [5] と呼ぶ周波数領域符号化方式は，低遅延に符号化できるという特長を持ち，対象音源の種類に合わせて量子化・圧縮を行う領域を適応的に変える音声通信用の符号化方式への応用が期待されている。

TCX の中でも修正離散コサイン変換（Modified Discrete Cosine Transform; MDCT）によるスペクトルを量子化し，エントロピー符号化する方式がある。エネルギーが集中する周波数帯域は対象とする音源の種類や分析区間によって異なるため，周波数成分の確率分布は周波数に大きく依存する。各帯域の周波数成分の確率分布が既知であればそれに合わせてエントロピー符号化すれば高効率な圧縮が可能であるが，一般の音響信号を対象とした場合は未知である。そこで TCX では，後述するように線形予測分析が周波数領域では所与の信号のスペクトル包絡推定に相当していることを利用し，線形予測係数から求められるスペクトル包絡（以下，LPC 包絡）を各帯域の周波数成分の分散の推定値として用いている。また，スペクトル包絡の情報は，各帯域においてエネルギーの大きい周波数成分が存在しているかどうかを示しているため，人間の音の強さに対する弁別閾がおおよそ対数的である点とマスキング特性を利用して聴覚的な量子化誤差を小さく抑えるための帯域ごとの量子化幅の設定にも活用できる。

しかし，従来の LPC 包絡は各周波数成分のエントロピー符号化への利用を想定したものにはなっていないため，符号長をより短くできるスペクトル包絡の与え方が存在する可能性がある。杉浦らは [11, 12] で，エントロピー符号化として Golomb-Rice 符号の利用を想定し，Golomb-Rice 符号に対して最適な包絡を得る手法を導出している。以下では，この手法を紹介する。

5.2 線形予測分析によるスペクトル包絡推定

まず，線形予測分析がスペクトル包絡推定に相当していることを確認するため，線形予測分析を周波数領域における最適化問題として定式化する。3 章で述べたように線形予測分析は時間領域では式 (9) を尤度関数とした予測係数 $\boldsymbol{a}$ の最尤推定問題として定式化される。ここで，$\boldsymbol{s}$ をある長さの分析窓における離散時間信号とし，その離散 Fourier 変換を考える。$\boldsymbol{F}$ を離散 Fourier 変換行列[1]とすると，$\boldsymbol{s}$ の離散 Fourier 変換は $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{s}$ で与えられ，$\boldsymbol{x}$ の各要素は異なる周波数の成分を表す。$\boldsymbol{F}$ は直交行列で $|\boldsymbol{F}| = 1$ なので，確率密度関数の変数変換により，$\boldsymbol{x}$ は

$$\boldsymbol{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{x}; \sigma^2 \boldsymbol{F}\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{-\mathsf{T}}\boldsymbol{F}^{\mathsf{H}}) \tag{30}$$

に従う。ただし，$\mathcal{N}_{\mathbb{C}}$ は複素正規分布を表す。ここで，分析窓の両端点において信号が巡回していることを仮定し，式 (7) の代わりに $\boldsymbol{\Psi}$ を

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & & & -a_P & \cdots & -a_1 \\ -a_1 & \ddots & & & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & -a_P \\ -a_P & & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ 0 & & -a_P & \cdots & -a_1 & 1 \end{bmatrix} \tag{31}$$

のような巡回行列とする。巡回行列同士の積は巡回行列になり，また，巡回行列は離散 Fourier 変換行列により対角化されるため，

$$\begin{aligned} \sigma^2 \boldsymbol{F}\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{-\mathsf{T}}\boldsymbol{F}^{\mathsf{H}} &= \sigma^2 (\boldsymbol{F}\boldsymbol{\Psi}^{\mathsf{T}}\boldsymbol{\Psi}\boldsymbol{F}^{\mathsf{H}})^{-1} \\ &= \mathrm{diag}(\lambda_1, \ldots, \lambda_T) \end{aligned} \tag{32}$$

となる。ただし，λ_k は $\sigma^2 \boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{-\mathsf{T}}$ の固有値

$$\lambda_k = \frac{\sigma^2}{|A(e^{2\pi j(k-1)/T})|^2} \tag{33}$$

$$A(z) = 1 - a_1 z^{-1} - \cdots - a_P z^{-P} \tag{34}$$

[1] 各行に異なる周波数の複素正弦波が格納された行列

で与えられ，周波数 k における全極スペクトルの二乗を表す。式 (30) 及び式 (32) より，所与の $\boldsymbol{x}$ の下での $\boldsymbol{a}$ の対数尤度は

$$\log p(\boldsymbol{x}|\boldsymbol{a}) = -\sum_k \left(\log \pi\lambda_k + \frac{|x_k|^2}{\lambda_k} \right) \tag{35}$$

となり，$\lambda_k = |x_k|^2$ のときに最大になる。よって，式 (35) に $\lambda_k = |x_k|^2$ を代入したものから式 (35) を引いたもの

$$\sum_k \underbrace{\left(\frac{|x_k|^2}{\lambda_k} - \log \frac{|x_k|^2}{\lambda_k} - 1 \right)}_{D_{\mathrm{IS}}(\lambda_k \| |x_k|^2)} \tag{36}$$

は信号のパワースペクトル $|x_k|^2$ と全極スペクトルの二乗 λ_k の離れ具合を表す非負の尺度となる。これを板倉齋藤距離という [6]。板倉齋藤距離は非対称で，λ_k が $|x_k|^2$ を下回る場合により過大なペナルティを課す関数であるため，λ_k が $|x_k|^2$ をできるだけ下回らず $|x_k|^2$ のピークの近くを通る曲線のとき小さい値になる。これが LPC をスペクトル包絡推定と見なせる理由である。式 (36) を最小化する $\boldsymbol{a}$ は，観測パワースペクトル $|x_1|^2, \ldots, |x_T|^2$ を逆 Fourier 変換して自己相関関数を求め式 (11) を解くことで得られる。

5.3 スパース性を用いた周波数領域符号化

多くの音源のスペクトルはスパースである。そこで，対象とする音響信号の各 MDCT 係数は Laplace 分布に従うものとする。4.1 節で述べたように Golomb-Rice 符号は Laplace 分布に従う情報源に対して最適な符号であるので，以下ではエントロピー符号化に Golomb-Rice 符号を用いる場合について議論する。

式 (15) において 2^{r-1} が分布の分散に対応しており，[5] のエントロピー符号化の枠組では，各周波数において Rice パラメータ r を LPC 包絡値によって決定する点がポイントである。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，スペクトル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る可能性がある。これに対し，[11, 12] では，スペクトル包絡の抽出法が従来の線形予測分析と同形のアルゴリズムになるように 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いている。ここで重要となるのが 5.2 節で述べた線形予測分析の周波数領域での解釈である。5.2 節で述べたように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 (33) で表される全極スペクトルの二乗値との板倉齋藤距離を最小化する係数 $\boldsymbol{a}$ を求める問題と等価であり，この係数 $\boldsymbol{a}$ 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スペクトルとの板倉齋藤距離最小化の形に帰着できれば，その問題は他の線形予測分析手法同様，効率的に解くことができる。

5.4 符号長を規準とした包絡推定法

各周波数 k における MDCT 係数を x_k，それを量子化幅 $w_k s$ で量子化したものを $y_k (= x_k/(w_k s))$，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

$$\begin{aligned} \mathcal{L}(\boldsymbol{r}) &\simeq \sum_k \left(\frac{|y_k|}{2^{r_k}} + r_k + 1 \right) \\ &= (\log_2 e) \sum_k D_{\mathrm{IS}}((\log_2 e) 2^{r_k} \| |y_k|) \\ &\quad + C(\boldsymbol{y}) \end{aligned} \tag{37}$$

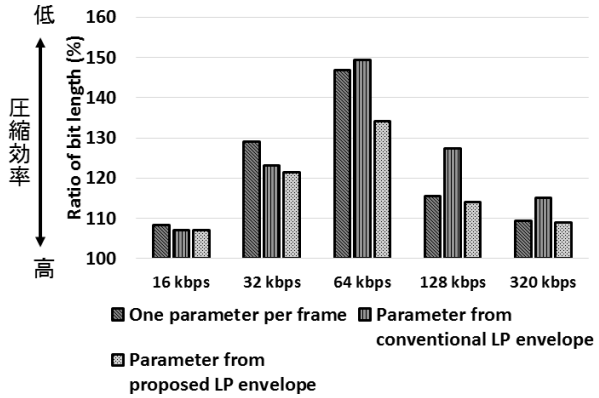
のように板倉齋藤距離を用いて表すことができる [11]。ただし，ここでは丸めは無視し，正負符号は別途符号化するものとする。ここで，Rice パラメータ r_k と全極スペクトル λ_k を

$$r_k \equiv \log_2 \left(\frac{(\ln 2)\lambda_k}{w_k s} \right) \tag{38}$$

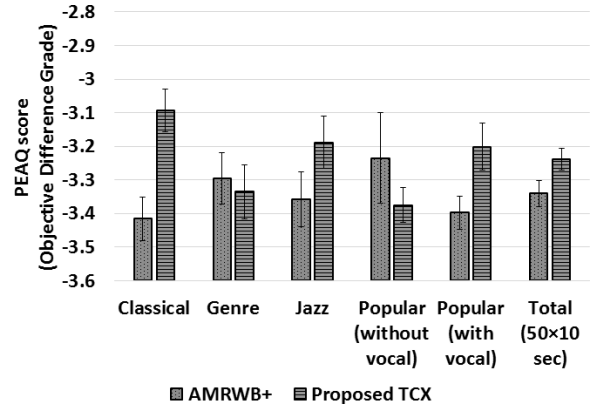
のように関連付けると，符号長最小化問題を

$$\begin{aligned} \hat{\boldsymbol{a}} &= \operatorname*{argmin}_{\boldsymbol{a}} \mathcal{L}(\boldsymbol{r}) \\ &= \operatorname*{argmin}_{\boldsymbol{a}} \sum_k D_{\mathrm{IS}}(\lambda_k \| |x_k|) \end{aligned} \tag{39}$$

のように全極スペクトルの二乗とスペクトルの絶対値との板倉齋藤距離最小化問題に帰着させることができる。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仮想的な信号の自己相関関数に相当するもの（スペクトルの絶対値の逆 Fourier 変換）を用いて立てられる式 (11) と同形の方程式を解くことで得られるため，Levinson-Durbin アルゴリズムにより高速に計算することができる。[12] では以上の定式化を Laplace 分布



図–5 Rice パラメータの割り当てによる平均記述長の比較
100%はすべての周波数において最適な Rice パラメータを割り当てたときの平均記述長を表す。包絡の次数は 16。



図–6 作成した TCX と AMR-WB+とのデータベースごとの PEAQ 値比較
平均と 95%信頼区間。

と正規分布を包含する一般化正規分布の仮定の下で一般化しており，この枠組を Powered All-Pole Spectrum Estimation（PAPSE）と呼んでいる。

5.5 提案法の効果

従来の TCX [5] 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 LPC と提案法を使用したときの圧縮率の比較を複数のビットレートで行った。RWC 音楽データベース [10] 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実験データとして使用した。図–5 に比較結果を示す。図のとおり提案法による 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。また，同じテストデータを用い，提案法と AMR-WB+の 16 kbps における音質の客観評価を行った。図–6 に，音質の客観評価値（Perceptual Evaluation of Audio Quality（PEAQ）[13]）の比較結果を示す。図のとおり AMR-WB+よりも高い評価値が得られていることが分かる。

6. おわりに

エントロピー符号の一つである Golomb-Rice 符号は Laplace 分布に従う情報源に対して最適な符号である。Laplace 分布はスパースな分布の一つであるため，Golomb-Rice 符号を用いる場合符号化対象はスパースなほど高効率な符号化が可能である。本稿では LPC による時間領域及び周波数領域の音声音響符号化を題材とし，音声音響信号のスパース性を活用して Golomb-Rice 符号長を最小化する符号化アプローチを紹介した。

文　献

[1] S.W. Golomb, “Run-length encodings,” *IEEE Trans. Inf. Theory*, **12**, 399–401 (1966).
[2] R.F. Rice, “Some practical universal noiseless coding techniques — part I–III,” *Jet Propul. Lab. Tech. Rep.*, **JPL**-79-22, **JPL**-83-17, **JPL**-91-3 (1979, 1983, 1991).
[3] T. Liebchen and Y. Reznik, “MPEG-4 ALS: An emerging standard for lossless audio coding,” *Proc. IEEE Data Compression Conference 2004*, pp. 439–448 (2004).
[4] ISO/IEC 14496-3:2005/Amd 2:2006; Audio Lossless Coding (ALS), new audio profiles and BSAC extensions.
[5] G. Fuchs, M. Multrus, M. Neuendorf and R. Geiger, “MDCT-Based coder for highly adaptive speech and audio coding,” *Proc. EUSIPCO*, pp. 1264–1268 (2009).
[6] 板倉文忠, “統計的手法による音声分析合成系に関する研究,” 博士論文, 名古屋大学大学院工学研究科 (1972).
[7] 亀岡弘和, 鎌本　優, 原田　登, 守谷健弘, “予測誤差の Golomb-Rice 符号量を最小化する線形予測分析,” 信学論, **J91-A**, 1017–1025 (2008).
[8] J.M. Ortega and W.C. Rheinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).
[9] D.R. Hunter and K. Lange, “Quantile regression via an MM algorithm,” *J. Comput. Graphical Stat.*, **9**, 60–77 (2000).
[10] 後藤真孝, 橋口博樹, 西村拓一, 岡　隆一, “RWC 研究用音楽データベース：研究目的で利用可能な著作権処理済み楽曲・楽器音データベース,” 情報処理学会論文誌, **45**, 728–738 (2004).
[11] 杉浦亮介, 鎌本　優, 原田　登, 亀岡弘和, 守谷健弘, “Golomb-Rice 符号化のための最適スペクトル包絡表現,” 音講論集, 1-2-5, pp. 233–236 (2015.3).
[12] R. Sugiura, Y. Kamamoto, N. Harada, H. Kameoka and T. Moriya, “Optimal coding of generalized-Gaussian-distributed frequency spectra for low-delay audio coder with powered all-pole spectrum estimation,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **23**, 1309–1321 (2015).
[13] [Online]. Available: http://www-mmsp.ece.mcgill.ca/Documents/Software/Packages/AFsp/AFsp.html (参照 2015-10-05).