

Vector Quantization of LSP Parameters Using Kalman-Neuro-Training Method

メタデータ	言語: jpn 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/3010

論文

カルマンニューロ法による LSP パラメータのベクトル量子化

森田 義則[†] 園 博幸^{††*} 船田 哲男^{††}

Vector Quantization of LSP Parameters Using Kalman-Neuro-Training Method

Yoshinori MORITA[†], Hiroyuki SONO^{††*}, and Tetsuo FUNADA^{††}

あらまし 電話音声の低ビットレート符号化にはコードブックを用いたベクトル量子化が有効であるが、計算量、記憶容量が多いという問題がある。この改善策としてニューラルネットワークを用いたベクトル量子化を提案する。ニューラルネットワークを用いる場合の問題点は学習精度を上げることであり、LSP パラメータの量子化に、カルマンフィルタを用いた学習法の一つであるカルマンニューロ法を用いる。その結果、24 ビット/フレームの量子化によりスペクトルひずみで 1.41 dB という良好な結果を得た。また、カルーネン・レーベ変換と比べた結果、スペクトルひずみでほぼ同じ性能を得た。

キーワード 音声符号化、カルマンフィルタ、ニューラルネットワーク、LSP、ベクトル量子化

1. まえがき

移動体通信の需要が増えている。限られた周波数帯域でより多くのサービスを行うためには個々の占有周波数を狭くする必要がある。中でも、携帯電話、PHS のような音声通信の需要はますます増加していくと考えられ、そのため音声の情報圧縮による狭帯域通信が求められている。

低ビットレートの音声符号化には、一般に、音声情報を音源情報と声道情報を分け、それぞれを高能率に符号化する方法が有効である。本研究では、声道情報を表す LSP パラメータをフレーム単位で圧縮することを検討する。

LSP パラメータのベクトル量子化 (VQ) に関しては、LBG アルゴリズム [1] などによるコードブックを用いたベクトル量子化があるが、ビット数が多くなると必要とする計算量や記憶容量が多いという問題点が残されている。このため、多段ベクトル量子化 [2], [3] などが提案されている。

ニューラルネットワーク (NN) を用いたベクトル量子化はこの点で有利であり、計算量、記憶容量とも約

1 けた小さくなることが期待できる。しかし、ニューラルネットワークの代表的な学習法である誤差逆伝搬法 (BP 法) [4] には学習回数が非常に大きくなること、誤差を所定の誤差に到達させるにはさまざまなテクニックを使う必要があることの問題点がある。以前の研究において [6], 5 層のニューラルネットワークでは十分に学習が収束せず、スペクトルひずみを小さくできなかった。そのため、ベクトル量子化器としては十分な性能を発揮させることができなかった。ニューラルネットワークのより有効な学習法が必要である。

この誤差逆伝搬法の問題点を改善するために、カルマンフィルタを用いた学習法が提案され、学習誤差が誤差逆伝搬法より小さくなったと報告されている [7], [8]。特に、カルマンニューロ学習法 (KNT 法) は収束精度が誤差逆伝搬法をはるかにしのぐほどに改善されたと報告されている。

本研究では、更にひずみの低減を目指してカルマンニューロ学習法を用いたニューラルネットワークによるベクトル量子化を行い、その性能を示す。また、ニューラルネットワークによる情報圧縮を入力から出力への恒等写像と考えると、カルーネン・レーベ変換 (KL 変換) を応用する方法がある。カルーネン・レーベ変換は変換後の特徴成分が互いに無相関となるように変換するものであり、特徴の集中度を最大限に高めるという意味で最適な変換方式である。この変換と特徴成分の最適ビット配分のスカラ量子化を組み合せた

[†] 石川工業高等専門学校電気工学科、石川県
Ishikawa National College of Technology, Ishikawa-ken, 920-03
Japan

^{††} 金沢大学工学部、金沢市
Faculty of Engineering, Kanazawa University, Kanazawa-shi, 920
Japan

* 現在、シャープ K. K.

ベクトル量子化との比較結果も示す。

2. ニューラルネットワークによるベクトル量子化

使用したニューラルネットワーク (NN) の構造は図 1 のような 5 層の階層型であり、各ユニットの入出力変換にはシグモイド関数を用いる。入力および出力が 1 フレーム分の LSP パラメータであり、この NN は与えられた入力に対して同じ値を出力する恒等写像 NN である。第 2 中間層ユニット出力の量子化はユニットごとにスカラ量子化を行い、これによって入力された LSP パラメータをベクトル量子化し、また第 2 中間層ユニット出力の量子化後の値を第 3 中間層に入力し復号した LSP パラメータを得る。よってこの NN は図 2 のように第 2 中間層が 2 層に分割され、符号化部および復号化部がそれぞれ 3 層の NN を有する形であるといえる。

NN の学習には重み係数の初期値をランダムに与え、誤差逆伝搬法およびカルマンニューロ学習法の両者で比較を行う。このとき、第 2 中間層ユニット出力を量子化して次の層へ入力しながら学習する「量子化学習」を行う。この量子化学習はひずみの低減に有効である [6]。

本研究では、DoD-CELP 方式 [9] による合成音声とそん色のない程度の合成音声を 34 ビット未満のビット数で得ることを目的の一つとしている。DoD-CELP

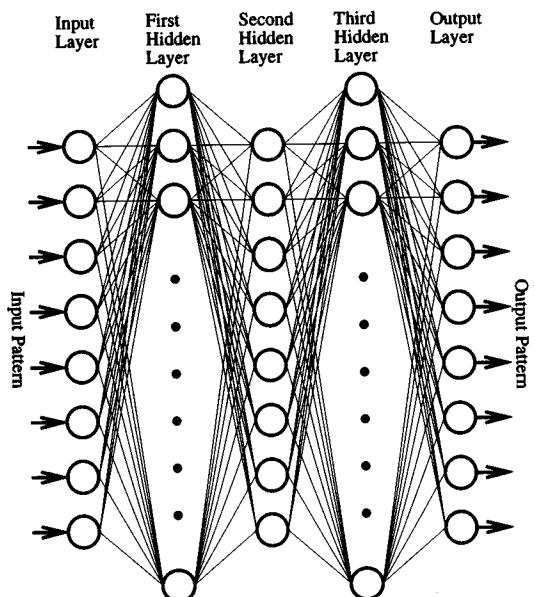


図 1 5 層の階層型ニューラルネットワーク
Fig. 1 5 layers neural network.

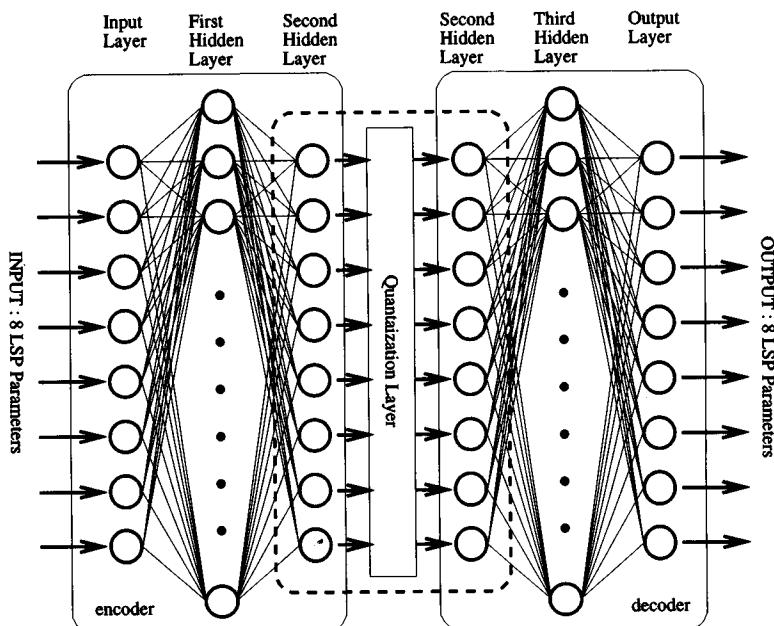


図 2 量子化ニューラルネットワーク
Fig. 2 Vector quantization using neural network.

方式の簡単な諸元を表 1 に示す。本研究ではフレーム当たり 24 ビットを想定する。

以前の研究において [10]、中間層の出力分布に基づいて量子化レベルを決定する非一様量子化法ではスペクトルひずみに改善がみられなかったので、ここでは一様量子化とし、量子化された出力値が 0 から 1 で等間隔となるように量子化を行う。

一様量子化による量子化学習では、量子化ビット数を一定とした場合、各ユニットの量子化ビット数を少なく抑えてユニット数を多くしても、また、ビット数を増やしてユニット数を少なくしてもひずみは改善されない。中間の値にひずみが最小となる組合せが存在する。8 次の LSP パラメータの場合、中間層のユニット数を 8 以上としても、また、各ユニットの量子化ビット数を 6 以上としてもひずみは減らないことを確認しているので、ここではフレーム当たり 24 ビットとするために、ユニット数を 8、ビット数を 3 とした。文献 [10] のいくつかの組合せに対する結果を表 2 に示す。

また、以前の研究では [5], [6]、5 層の NN 全体を一括して学習しても十分にスペクトルひずみを小さくできなかった。そこで、5 層の NN を二つの 3 層の NN に分け、それぞれの NN を別々に学習させた。このとき、教師信号として LSP パラメータをベクトル量子化したときのコード番号に対応する 2 進表記が必要となるので、学習用の LSP パラメータからコホーネンの自己組織化特徴地図 (KSFM) [11] を利用してコードブックを作成した。学習が終了したそれぞれの 3 層

表 1 DoD-CELP の諸元
Table 1 Specifications of DoD-CELP.

サンプリング周波数	8 kHz
LSP 分析次数	10
LSP 量子化方式	スカラ量子化
LSP 量子化ビット数	34

表 2 ユニット数、ビット数の組合せとスペクトルひずみ。
u は中間層のユニット数、b は 1 ユニット当りのビット数を示す

Table 2 Spectral distortions relative to combination of units and allocated bits (u=the number of units, b=bits per unit).

(u, b)	スペクトルひずみ [dB]
(3, 8)	3.36
(4, 6)	2.88
(6, 4)	2.12
(8, 3)	1.68
(12, 2)	2.36

の NN を結合して 5 層の NN を構成し、3 層での学習後の重みを初期値として、更に 5 層一括して学習を行うとよりひずみを小さくできた。このような学習方法は各ユニット出力をグレーコードで 2 値化する場合には非常に有効であるが [6]、今回のような各中間層出力を 3 ビットで量子化する (24 ビットコード) 場合には適用できない。

そこで、本研究では 5 層の NN を一括して学習してもひずみを十分に小さくできることが期待されるカルマンニューロ学習法を適用する。

3. カルマンニューロ学習法

3.1 アルゴリズム

カルマンニューロ学習法 (KNT 法) [8] では、ニューラルネットワーク (NN) の学習を次のような非線形システム

$$1) \text{ 状態式} \quad \mathbf{w}_{t+1} = \mathbf{I}\mathbf{w}_t \quad (1)$$

$$2) \text{ 観測式} \quad \mathbf{y}_t = h(\mathbf{w}_t; \mathbf{x}_t) + \mathbf{v}_t \quad (2)$$

の状態推定問題として考える。ここで、 \mathbf{w}_t は結合重み変数ベクトル、 \mathbf{x}_t 、 \mathbf{y}_t はそれぞれ NN の入力および出力ベクトル、 \mathbf{I} は単位行列であり、 $h(\cdot)$ は NN の構造と各ユニットの非線形特性で決まる関数で時不変とする。 \mathbf{v}_t はガウス性の雑音ベクトルであり、次式を満たすとする。

$$E[\mathbf{v}_t] = \mathbf{0}$$

$$E[\mathbf{v}_t \mathbf{v}_s^T] = \mathbf{R}_t \delta_{ts}$$

$$\mathbf{R}_t > 0 \text{ (正定値性)}$$

$E[\cdot]$ は期待値を表す。

式 (2) の観測式は \mathbf{w}_t について非線形となっているので、このままではカルマンフィルタを適用できない。そこで、非線形関数 $h(\mathbf{w}_t; \mathbf{x}_t)$ が滑らかであるという仮定の下に、式 (2) を推定値 $\hat{\mathbf{w}}_{t/t-1}$ のまわりに泰ラー展開し、高次の項を無視して線形化する。この線形システムにカルマンフィルタを適用することにより、次の拡張カルマンフィルタの式を得る。

1) フィルタ方程式

$$\hat{\mathbf{w}}_{t/t} = \hat{\mathbf{w}}_{t/t-1} + \mathbf{K}_t [\mathbf{y}_t - h(\hat{\mathbf{w}}_{t/t-1}; \mathbf{x}_t)] \quad (3)$$

$$\hat{\mathbf{w}}_{t+1/t} = \hat{\mathbf{w}}_{t/t} \quad (4)$$

2) カルマンゲイン \mathbf{K}_t

$$\mathbf{K}_t = \mathbf{P}_{t/t-1} \mathbf{H}_t^T [\mathbf{H}_t \mathbf{P}_{t/t-1} \mathbf{H}_t^T + \mathbf{R}_t]^{-1} \quad (5)$$

ここで

$$\mathbf{H}_t = (\partial h / \partial \mathbf{w}_t) |_{\mathbf{w}_t = \hat{\mathbf{w}}_{t/t-1}} \quad (6)$$

3) 推定誤差共分散行列 $\mathbf{P}_{t/t}$

$$\mathbf{P}_{t/t} = \mathbf{P}_{t/t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t/t-1} \quad (7)$$

$$\mathbf{P}_{t+1/t} = \mathbf{P}_{t/t} \quad (8)$$

4) 初期条件

$$\hat{\mathbf{w}}_{0/-1} = \bar{\mathbf{w}}_0, \quad \mathbf{P}_{0/-1} = \Sigma_0 \quad (9)$$

以上の式において、ベクトル \mathbf{y}_t を学習用教師データ ($= \mathbf{x}_t$), $h(\hat{\mathbf{w}}_{t/t-1}; \mathbf{x}_t)$ をネットワーク出力、時間 t を反復計算回数と考えれば、結合重み \mathbf{w}_t を推定でき、次の KNT アルゴリズムを得る。

Step 0 $\hat{\mathbf{w}}_0$ は $-0.05 \sim 0.05$ 間の一様乱数。 Σ_0 は出力誤差で定まる^(注1)対角行列。 $t \leftarrow 0$ 。

Step 1 最初の学習データ $(\mathbf{x}_0, \mathbf{y}_0)$ を用意。

Step 2 式(6)による観測行列 \mathbf{H}_t の計算。

Step 3 式(5)によるカルマンゲイン \mathbf{K}_t の計算。

Step 4 式(3), 式(7)により $\hat{\mathbf{w}}$ および \mathbf{P} を観測更新。

Step 5 もし $t < M$ ならば、式(4), 式(8)により $\hat{\mathbf{w}}$ および \mathbf{P} を時間更新 ($t \leftarrow t+1$) して **Step 2** へ。

Step 6 もし、 ϵ^2 が指定値以下になるかまたは学習回数が指定回数以上になれば終了。

Step 7 式(4)および式(10)により $\hat{\mathbf{w}}$ および \mathbf{P} を更新し、 $t \leftarrow 0$ として **Step 1** へ。

$$\mathbf{P}_{t+1/t} = \beta \epsilon^{-2} \mathbf{I} \quad (10)$$

ここで、 M はフレーム総数、 β は定数、 ϵ^2 は学習データの全フレームにわたる出力誤差の 2乗和である。式(10)は反復ごとに \mathbf{P} をネットワーク出力の誤差に従って設定する。以下において、NN の学習回数とは全学習データを走査したときを 1 回と数える。

3.2 KNT 法と BP 法の関係 [14]

NN の学習とは、NN の出力を望ましい出力（教師データ）に近づくように重みを更新していく過程である。 o_k を NN の出力、 t_k を教師パターンとして、NN

(注 1) : 便宜上式(10)で与える。

の誤差 E を次式で定義する。

$$E = \frac{1}{2} \sum_{k=1}^N (t_k - o_k)^2 \quad (11)$$

学習の目的は E を最小化することである。

BP 学習則では最急降下法を用いて、結合重み w_i を以下のように調節する。

$$\begin{aligned} \Delta w_i &= -\eta \frac{\partial E}{\partial w_i} \\ &= -\eta \sum_{k=1}^N \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial w_i} \\ &= -\eta \sum_{k=1}^N (t_k - o_k) \frac{\partial o_k}{\partial w_i} \end{aligned} \quad (12)$$

ここで、 N は出力ユニットの数、 η は学習率と呼ばれる定数である。

階層型 NN を学習させる拡張カルマンフィルタアルゴリズムに次の仮定

$$1) \quad \mathbf{P}_{t/t-1} = \beta \epsilon^{-2} \mathbf{I}$$

$$2) \quad [\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R}]^{-1} = a \mathbf{I}$$

を代入すると、式(3)は次式のように変形できる。

$$\Delta w_i = -a \beta \epsilon^{-2} \sum_{k=1}^N (t_k - o_k) \frac{\partial o_k}{\partial w_i} \quad (13)$$

この式と式(12)を比較すると、 $\eta = a \beta \epsilon^{-2}$ とすれば式(13)は BP 学習則と等しくなる。 ϵ^2 は誤差の 2乗の総和であり、誤差の分散に比例する。

仮定 1) より BP 法は等価的に誤差の共分散行列を対角形であると仮定していることがわかる。すなわち、BP 法ではそれぞれの重みの誤差が無相関であると仮定していることになる。しかし、誤差が逆伝搬していくので、上位層の重みは下位層の重みの修正に影響を与えることを考慮すると、上位層と下位層の重みの誤差間には相関があると考えるのが自然である。しかし、BP 法はこの相関がないものと仮定している。

これらのことより以下の 2 点で BP 法に対する KNT 法の特徴を表すことができる。

1) 誤差の分散の逆数で学習率を適応的に設定している。

2) 重みの誤差間の相関も利用した重みの修正を行っている。

このことより、KNT 法は BP 法に比べて高速な学習が期待できる。また、常に真とは限らないが、NN

の学習の高速化のための修正アルゴリズムは、よりよい局所最適解に導く場合がある [15] ことが知られているので、KNT 法は BP 法に比べて収束誤差を小さくできることが期待できる。

4. 評価実験

NN への入力データは音声波形の分析より得られた LSP パラメータである。分析条件を表 3 に示す。学習用として男女 5 名が発声した短文 15 文より 23,139 フレーム、評価用として別の男女 5 名が発声した短文 15 文より 22,061 フレームのデータを用いた。

これらの音声は、日本音響学会より提供されている連続音声データベース CD-ROM 内に、16 kHz サンプリングで収められている音声を、8 kHz にダウンサンプルしたものである。

学習結果の客観的評価基準として、次式で定義するスペクトルひずみを用いる。

$D[\text{dB}]$

$$= \sqrt{\frac{1}{M} \sum_{t=0}^{M-1} \frac{1}{W} \int_0^W \{S_o(f, t) - S_d(f, t)\}^2 df} \quad (14)$$

ここで $S_o(f, t)[\text{dB}]$ はニューラルネットワークへ入力した LSP パラメータから計算されるパワースペクトルであり、 $S_d(f, t)[\text{dB}]$ は出力された LSP パラメータから計算されるパワースペクトルである。また、 t はフレーム番号、 M はフレーム総数である。

5. 各種ベクトル量子化法の比較

5.1 KSFM による VQ

コードブックによるベクトル量子化 (VQ) では、コードブックのサイズとして計算量、記憶容量の点で 1,000 くらいまで使われることが多い。ここでは、コホーネンの自己組織化特徴地図 (KSFM) により 256 と 4096 のサイズのコードブックを作成し、LSP パラメータのベクトル量子化を行った。このときのそれぞれのスペクトルひずみは 2.41 dB, 1.50 dB となった。

表 3 LSP 分析条件
Table 3 Analysis conditions of LSP parameters.

フレーム長	30 ms
フレーム周期	30 ms
LSP 分析次数	8

5.2 2 段 VQ

今回想定している 24 ビットの量子化ビット数の場合のコードブックによる量子化を評価するために、5.1 のコードブックのサイズを 24 ビットのサイズにするのは、計算量、記憶容量の点で現実的でなく、また、学習用データ数が 23,139 個であるので意味がない。そこで、2 段ベクトル量子化のためのコードブックを KSFM により作成した。それぞれの段のコードブックのサイズは 4096 である。このときのスペクトルひずみは 0.56 dB となった。

5.3 スカラ量子化

LSP パラメータをスカラ量子化するために、学習用データより Lloyd アルゴリズムを使ってコードブックを作成した。8 個のコードブックに配分されるビット数の合計を 24 として、最もスペクトルひずみが小さくなる組合せを探したところ、LSP の各線スペクトル周波数の低周波側より順に 3, 3, 3, 3, 4, 3, 3, 2 のビット配分のとき最小となり、このときのひずみは 1.34 dB となった。

5.4 BP 法

使用した NN の構造は図 1 の 5 層の階層型で、各層のユニット数は入力層側から 8, 15, 8, 15, 8 である。第 2 中間層を 3 ビットで量子化し、全体で 24 ビットとする。学習はモーメント項を入れた BP 学習則を用い、更新則は一括学習を用いた。学習率を 10, モーメント項の係数を 0.9 とした。10 万回学習後のスペクトルひずみは 1.51 dB となった。

5.5 カルーネン・レーべ変換

学習に用いた全 LSP パラメータの共分散行列を計算し、その固有ベクトルから KL 変換のための変換行列を計算する。LSP パラメータをこの行列で変換し、その変換した 8 次の成分を特徴パラメータとみなして、それぞれの特徴パラメータを、全体のビット数が 24 となるようにビット配分を行って量子化する。量子化された特徴パラメータで KL 展開された LSP パラメータのスペクトルひずみを調べた。KL 変換では、大きい固有値に対応する固有ベクトルに対する特徴パラメータに特徴がより集中するので、集中度に応じて多めにビット配分をすることでひずみを小さくすることができる。良好なひずみを得たビット配分パターンについてのスペクトルひずみを表 4 に示す。最下段はビット数を等配分したときのひずみである。固有値の大きい順に第 1 から第 8 の特徴パラメータである。5, 4, 3, 3, 3, 2, 2, 2 のビット配分のとき最もひずみが小さ

表 4 KL 変換におけるビット配分の例とひずみ
Table 4 Examples of bit allocation and spectral distortions in KL transform.

ビット配分	学習内 [dB]	学習外 [dB]
54333222	1.23	1.44
44333322	1.24	1.45
54333321	1.25	1.54
53333322	1.28	1.46
33333333	1.50	1.64

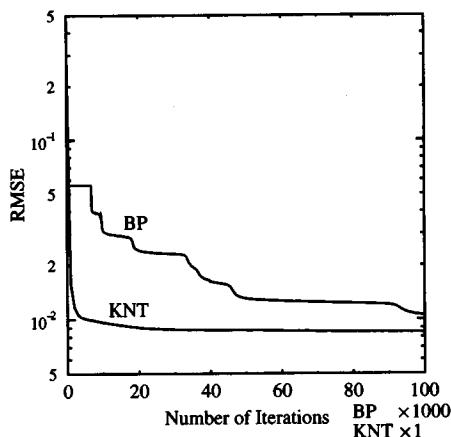


図 3 誤差曲線
Fig. 3 Learning curve.

くなり、1.23 dB となった。

5.6 カルマンニューロ法

5.4 と同じ NN をカルマンニューロ学習法 (KNT 法) で、 β を 200 として、100 回学習を行った。このときのスペクトルひずみは 1.21 dB となった。学習時の誤差の推移を BP 法と共に図 3 に示す。KNT 法の学習時における計算量は BP 法の約 3.5 倍である。

6. 考 察

5. で述べた各方式によるスペクトルひずみの結果をまとめて表 5 に示す。DoD-CELP 方式 (DoD) を用いて評価用の音声データを量子化した場合のスペクトルひずみは 1.432 dB であり、本研究の比較の基準とする。

KSFM による VQ はコードブックのサイズが 8 ビット (VQ8) または 12 ビット (VQ12) で小さいので、ひずみが大きくなる。コードブックのサイズを大きくすればひずみは小さくできるが、計算量、記憶容量が非常に大きくなってしまう。今回想定している量子化ビット数が 24 ビットの 2 段ベクトル量子化 (2VQ) の場合は 1.25 dB であった。この値は今回検討した中

表 5 スペクトルひずみの比較
Table 5 Comparison of spectral distortions.

方式	学習内 [dB]	学習外 [dB]
VQ8	2.41	2.99
VQ12	1.50	2.35
2VQ	0.56	1.25
SQ	1.34	1.51
BP3	1.50	1.68
BP5	1.51	1.88
KL	1.23	1.44
KNT3	1.49	1.65
KNT5	1.21	1.41
DoD		1.432

表 6 計算量、記憶容量の比較
Table 6 Comparison of computations and memory storages.

方式	計算量	記憶容量 [語]
VQ8	6143	2048
VQ12	98303	32768
2VQ	196614	65536
SQ	120	64
BP5	1644	572
KL	304	143
KNT5	1644	572

で最小であり、KL 変換および KNT 法に比べても約 0.2 dB 良い結果となった。学習内外の差がかなり大きいのは学習に用いたデータ数が少なかったためであろう。また、スカラ量子化 (SQ) した場合のスペクトルひずみは 1.51 dB であった。DoD-CELP よりもひずみが大きいのは、DoD-CELP の量子化ビット数が 34 ビットであり、ビット数の差のためである。

BP 法において 3 層の NN を用い、中間層を 24 ビットに量子化すると (BP3)，ひずみを 1.68 dB とできただが [10]，必ずしも十分ではない。スペクトルひずみを十分に小さくするためには、符号化、復号化 NN の両方に中間層が必要と考えられる。そこで、NN の構造を 3 層から 5 層 (BP5) に変えてひずみの低減を期待したが、ローカルミニマムに落ち込み逆にスペクトルひずみが増えている。

KL 変換 (KL) と KNT 法 (KNT5) ではそれぞれ 1.44 dB, 1.41 dB となり、基準とした DoD-CELP の 1.432 dB とほぼ同じひずみが得られた。そして、わずかではあるが KNT 法のひずみが小さくなった。NN の構造を 3 層 (KNT3) としたときのひずみは 1.65 dB である。

表 6 にいくつかの方式の計算量と記憶容量を示す。計算量の算出において、その単位として、四則演算は 1, 積和は 2, 整数化および実数化処理は 1, 指数関数

表 7 非単調増加の割合

Table 7 Percentage of nonmonotonic increasing LSPs.

方式	学習内 (%)	学習外 (%)
VQ	0	0
BP3	0.35	0.46
BP5	0.056	0.095
KL	1.00	1.09
KNT5	0.039	0.054

の計算は 10 と仮定した（付録参照）[16]。コードブックを用いたベクトル量子化に対して NN を用いた方式は計算量、記憶容量とも約 1 けた小さくなることが期待できる。KL 変換と KNT 法ではスペクトルひずみはほぼ同じである。しかし、計算量、記憶容量の点で KL 変換の方が有利である。

LSP パラメータの各線スペクトル角周波数を $\omega_1, \omega_2, \dots, \omega_8$ とすると、LSP を用いた合成フィルタが安定であるためには、それらが単調増加することが必要十分条件である。すなわち、 $0 < \omega_1 < \omega_2 < \dots < \omega_8 < \pi$ という条件を満たせばよい。LBG アルゴリズムによる VQ では LSP の逆転は生じない。ところが、KL 変換や NN を用いたベクトル量子化はこの単調増加の関係が満たされない場合が起こり、この割合を調べた結果を表 7 に示す。

VQ は、例えば、LBG アルゴリズムでコードブックを作成すれば、LSP パラメータの逆転は生じない。KL 変換は、スペクトルひずみは小さくなるが、LSP 逆転の割合がかなり大きい。BP 法において 3 層 (BP3) から 5 層 (BP5) に増やしたところ、スペクトルひずみに大きな差はないにもかかわらず、非単調増加の割合は激減した。KNT 法の場合は更に割合が小さくなり、学習内で 23,139 フレーム中 9 フレーム (0.039%)、学習外で 22,061 フレーム中 12 フレーム (0.054%) であり、学習外のデータに対しても KNT 法が良い結果を示している。

7. む す び

音声の高能率符号化方式としてコードブックを用いたベクトル量子化法は十分な性能を示すが、計算量、記憶容量の点で問題がある。本論文では、コードブック検索の代わりに、ニューラルネットワークを用いた方式によりその欠点の改善を試み、24 ビットのベクトル量子化でスペクトルひずみが 1.41 dB という良好な結果を得た。また、カルーネン・レーべ変換との比較の結果、同一ビット数でほぼ同じスペクトルひずみを

得た。本方式は LSP パラメータの雑音低減、声質変換等のための任意の変換に応用することができる。

今後の課題としては、ニューラルネットワークに与える学習データを前処理することでスペクトルひずみを低下させることを検討したい。また、ニューラルネットワークの学習時において、出力の各ユニットの誤差に対して重みを付けることによるスペクトルひずみの低減について検討したい。今回は客観評価で DoD-CELP 方式とほぼ等しい値が得られたが、聴取実験による主観評価で KNT 法によるニューラルネット量子化器を評価する必要がある。

文 献

- [1] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., vol.COM-28, no.1, pp.84-95, Jan. 1980.
- [2] B.H. Juang and A.H. Gray, Jr., "Multiple stage vector quantization for speech coding," Proc. ICASSP, pp.597-600, 1982.
- [3] 保谷早苗, 板倉文忠, "木探索を用いた LSP パラメータの多段ベクトル量子化," 信学技報, SP93-69, 1993.
- [4] 甘利俊一監修, PDP モデル, 第 8 章, 産業図書, 1989.
- [5] 村本武志, 船田哲男, 八木敏夫, "5 層ネットワークによる LSP パラメータの量子化," 音響学会 H6 秋季論文集, 分冊 1, no.1-P-4, Oct.~Nov. 1994.
- [6] 船田哲男, 村本武志, "階層型ネットワークによる音声 LSP パラメータのベクトル量子化," 信学論 (A), vol.J78-A, no.9, pp.1242-1246, Sept. 1995.
- [7] S. Singhal and L. Wu, "Training feed-forward networks with the extended kalman algorithm," Proc. ICASSP, Glasgow, vol.2, pp.1187-1190, May 1989.
- [8] 村瀬治比古, 小山修平, 石田良平, パソコンによるカルマン・ニューロコンピューティング, 森北出版, 1994.
- [9] J.P. Campbell, Jr., V.C. Welch, and T.E. Tremain, "An expandable error-protected 4800 bps CELP coder," Proc. ICASSP, Glasgow, vol.2, pp.735-738, May 1989.
- [10] 園 博幸, 船田哲男, "3 層 NN による LSP パラメータのベクトル量子化と聴取実験による検討," 信学技報, SP95-52, 1995.
- [11] J. Dayhoff, Neural Network Architectures, Van Nostrand Reinhold, 1990.
- [12] 森田義則, 船田哲男, "誤差逆伝播法とカルマンニューロトレーニング法によるニューラルネットワークの学習の比較," 平7 北陸連大, no.B-57, p.159, 1995.
- [13] 森田義則, 船田哲男, "カルマン・ニューロ法による LSP パラメータの量子化," 信学技報, NC96-4, 1996.
- [14] D.W. Ruck, S.K. Rogers, M. Kabrisky, P.S. Maybeck, and M.E. Oxley, "Comparative analysis of backpropagation and the extended kalman filter for training multi-layer perceptrons," IEEE Trans. Pattern Anal. & Mach. Intell., vol.PAMI-14, no.6, pp.686-691, June 1992.
- [15] 甘利俊一, 向殿政男共編, "ニューロとファジイ," p.63, 培風館, 1994.

- [16] 山内二郎, 宇野利雄, 一松 信, 電子計算機のための数値計算法 III, pp.38-44, 培風館, 1972.

付 錄

符号語のサイズを L ビットとして, NN とコードブックの場合の計算回数と記憶容量を求める。

NN の場合, 結合重みの数を W , 入力層を除いたユニット数を U とする。第 2 中間層の各ユニット出力を B ビットでスカラ量化すると仮定すると, 第 2 中間層のユニット数は L/B となる。各ユニットは以下の計算を行う。

$$O = f \left(\sum_i w_i \cdot I_i + \theta \right) \quad (\text{A-1})$$

但し, O, w_i, I_i, θ はそれぞれ出力, 結合重み, 入力, オフセットである。括弧内の計算は積和が W 回となる。記憶容量は重み係数とオフセットの数であるから, $W + U$ 語となる。関数 f は,

$$f(x) = \frac{1}{1 + e^{-x}} \quad (\text{A-2})$$

である。よって, 指数関数, 和, 商がそれぞれ U 回である。スカラ量化は, 値を $(2^B - 1)$ 倍して小数点以下を四捨五入することにより行うとすると, 計算回数は量子化, 逆量子化で和, 積, 商, 整数化処理, 実数化処理がそれぞれ L/B 回となる。従って, 符号化, 復号化に必要な計算回数, 記憶容量は以下のようになる。

和	$U + L/B$
積	L/B
積和	W
商	$U + L/B$
整数化処理	L/B
実数化処理	L/B
指数関数	U
記憶容量	$W + U$

コードブックの場合, 符号語のサイズが L ビットであるから, コードブックのサイズは 2^L となる。各ベクトルは 8 個の要素よりなるので, ベクトル間の遠さの尺度を以下の式

$$\sum_i (\omega_i - \omega'_i)^2 \quad (\text{A-3})$$

で求めるすると, 最も近い要素を探す計算回数は差が 8×2^L 回, 積が 2^L 回, 積和が 7×2^L 回, 比較が $2^L - 1$ 回必要となる。必要な記憶容量は 8×2^L 語である。復号時は計算, 記憶容量とも必要ないので, 計算回数, 記憶容量は以下のようになる。

差	8×2^L
比較	$2^L - 1$
積	2^L
積和	7×2^L
記憶容量	8×2^L

KNT5 の場合, $W = 526, U = 46, L = 24, B = 3$ となり, 計算量は 1644, 記憶容量は 572 となる。VQ8 の場合, $L = 8$ より, 計算量は 6143, 記憶容量は 2048 となる。

(平成 8 年 7 月 25 日受付, 12 月 27 日再受付)

森田 義則 (正員)



昭 53 富山大・工・電子卒。昭 55 同大大学院修士課程了。富山大技官を経て、現在石川高専講師。音声情報処理の研究に従事。日本音響学会会員。

園 博幸 (正員)



平 6 金沢大・工・電気情報卒。平 8 同大大学院修士課程了。同年、シャープに入社。音声情報処理の研究に従事。

船田 哲男 (正員)



昭 41 金沢大・工・電子卒。昭 46 名大大学院博士課程了。昭 46 金沢大・工・講師。現在同大教授。生体情報処理、音声情報処理の研究に従事。共著「情報科学の基礎」、「数値解析の基礎」など。IEEE、日本音響学会、日本 ME 学会、情報処理学会各会員。