Vector Quantization of LSP Parameters with Layer Neural Networks

電子情報通信学会論文誌 '95/9 Vol. J78-A No. 9

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

船由 哲男†(正員)　　村本 武志†(准員)

Vector Quantization of LSP Parameters with Layer Neural Networks

Tetsuo FUNADA†, *Member* and Takeshi MURAMOTO†, *Associate Member*

† 金沢大学工学部, 金沢市
Faculty of Engineering, Kanazawa Univ., Kanazawa-shi, 920 Japan

あらまし　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について実験的な検討を行った．中間層ユニット出力を量子化しながらの学習，コードブックを介して学習する際のコード番号表示にGrayコードを用いる効用，5層ネットワークを応用するときの重み係数の初期値設定法について示す．

キーワード　階層型ニューラルネットワーク，ベクトル量子化，LSPパラメータ，Gray符号

1. まえがき

移動通信システムの急速な普及に伴い，限られた無線周波数においてより多くの通信回線を確保するために，低ビットレートでの音声符号化技術の開発が期待されている．そのためには，駆動音源とスペクトル包絡を分離して量子化する方法が効果的と言える．本研究では音声のスペクトル包絡特性を表すLSPパラメータを，階層型ニューラルネットを用いてフレーム単位でベクトル量子化することを目的とする．

LSPパラメータのベクトル量子化に関しては，LBGアルゴリズムなどによるクラスタリング法[1]が挙げられるが，ビット数が多くなると計算量や記憶量を多く必要とする．そのため多段ベクトル量子化[2],[3]，ベクトル・マトリクス量子化の方法[4]などが提案されている．本研究では，種々の構成の階層型ニューラルネットに対し，LSPパラメータのベクトル量子化の性能比較に関する実験的な検討を行った[5],[6]．階層型ネットワークで音声波形を直接ベクトル量子化する研究は文献[7]で見られるが，本研究ではLSPパラメータのベクトル量子化を対象としている．

最初に3層の階層型ニューラルネット（中間層1層）で恒等写像を実現するネットワークを構成し，その中間層ユニットの出力をスカラ量子化することでベクトル量子化とその復号を実現する．また，学習時に中間層出力を量子化しながら学習することで量子化ひずみの軽減ができることを示す．

次に，ベクトル量子化および復号化をそれぞれ個別に行う2種類の3層階層型ニューラルネット（中間層1層）を，別途作成したコードブックを用いて学習し，量子化ひずみを調べる．その際コードブックのコード番号の2進表現の仕方によるひずみの違いについても調べる．

更に結果を改善するため，これら二つの階層型ニューラルネットを一つにまとめ5層の階層型ニューラルネット（中間層3層）として再度学習を行い，量子化ひずみを比較する．

2. 実験方法の概要

ニューラルネットの学習に用いた音声は，男女各2名が発声した天気予報などの朗読文8文（計約84秒）を8kHzでサンプリングして収録した．また，ニューラルネットへの入力には，フレーム長20ms，フレーム周期20msとして8次で分析した計4208組のLSPパラメータを用いた．実験結果の評価には，式(1)で定義されるスペクトルひずみ（ネットワークへ入力したLSPパラメータのパワースペクトル $S_o(f,i)$[dB]と，ベクトル量子化した後復号して得られたLSPパラメータのパワースペクトル $S_d(f,i)$[dB]の差の周波数積分）によって行う．

スペクトルひずみ[dB]

$$= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{W}\int_0^W \{S_o(f,i)-S_d(f,i)\}^2 df} \tag{1}$$

ここで，iはフレーム番号，Nはフレーム総数を表す．学習内データでの評価は4208フレームすべてのLSPで，また学習外データでの評価は学習には用いなかった音声データから計算した503フレームのLSPで行った．

ニューラルネットの学習には，4208フレームのLSPパラメータを用いて，サイズ256のコードブックを作成し，このコードベクトルをネットワークに与えバックプロパゲーション法で学習した．但し，**4.4**と**5.**についての結果は，4208フレームすべてのLSPを用いて学習を行った．

3. 量子化・復号化3層ネットワーク

このニューラルネットは入力，中間，出力の各層がそれぞれ8，4，8ユニットで構成される3層ネットで，近似的に恒等写像を行うように学習する[8]．学習結果を表1に示す．学習はスペクトルひずみの変化がほとんど認められなくなった10000回で打ち切った．

表1 量子化・復号化3層ネットワークのスペクトルひずみ
Table 1 Spectral distorsion of 3-layer networks for coding and decoding.

方法	学習内 [dB]	学習外 [dB]
A	2.65	3.43
B	5.89	6.64
C	3.93	4.72
D	3.24	3.64

表中の方法Aは通常の学習を行ったことを表す．方法Bは方法Aで学習したネットワークでひずみを評価する際に，中間層ユニットの出力値範囲0～1を等間隔で4レベル（2bit）に量子化した場合の結果である．すなわち1フレームのLSPパラメータを8bitでベクトル量子化したことに相当する．方法Cは，学習データに対して中間ユニットの出力値が分布する範囲を調べ，その範囲を4レベルに量子化した場合である．

更に方法Dは，学習時に中間層ユニットの出力値を4レベルに量子化しながら学習を行った結果である（以下では量子化学習と呼ぶ）．方法Dは方法Bに比べ，2.6～3dBひずみが低下しており量子化学習の効果が確認できる．そこで，この学習法を5.で述べる5層ネットワークの学習に対しても適用することにする．

4. 3層量子化と3層復号化ネットワーク

3層の階層型ニューラルネットを用い，LSPパラメータの量子化と復号化を別個に行うネットワークの学習を行った．すなわち，LSPパラメータを入力しコード番号を出力する量子化ネットワーク，および逆にコード番号を入力しもとのLSPパラメータを出力する復号化ネットワークをそれぞれ個別に学習させる．

これらのネットワークを学習するには，各LSPパラメータに対応するコード番号が与えられていなければならない．そこで，あらかじめ学習用のLSPパラメータからKohonen Self-Organizing Feature Map（KSFM）[9]を利用してコードブックを作成する．KSFMは各LSPをMap上の特定のノードに対応づけることができかつ，近接した2組のLSPに対応するノードはMap上でも近い関係を保つ特徴をもっている．ここで，Map作成にはノード数（コードブックサイズ）を256個（＝8bit）とし，ノード配列のためのMapの次元を1次元および2次元の2通りに設定した．

量子化ネットワークの構成は，入力層はLSPパラメータの次数に合わせ8ユニット，出力層はコード番号を2進符号表示するのに必要な8ユニットとした．また，中間層は10と20ユニットの2通りに設定した．復号化ネットワークも符号化ネットワークと同じ構成とした．

スペクトルひずみの評価に際しては，量子化ネットワークの出力値を4捨5入して0,1で量子化し，いったんコード番号を表す2進符号に変換する．次にこの符号を復号化ネットワークに入力し，LSPパラメータに復号する．

4.1 ノードの1次元と2次元配列

KSFMのMap上のノード配列次元を，1と2の2通りに仮定して作成されたコードブックを，量子化および復号化ネットワークの学習用教師データの作成に用い，両者を比較した．コードブックの各コード番号を2進符号に変換するため，各コード（ノード）の端点から2進符号を割り当てた．2進符号は，1次元配列の場合はコードブックサイズである8ビットを用い，0～255を割り当て，2次元配列の場合には，16×16の配列であるため，上位4ビットにx座標0～15を，下位4ビットにもy座標0～15を割り当てる．また以後，KSFMのノード配列を1次元とした場合を1次元KSFM，2次元配列とした場合を2次元KSFMと呼ぶことにする．

1次元KSFMおよび2次元KSFMで得られたコードブックを用いて，量子化および復号化ネットワークの学習を行った結果，1次元KSFMで変換したコード番号を用いて学習した方が，2次元KSFMの場合よりもスペクトルひずみが約1dB小さくなった．このため，以後コードブックにはKSFMのノードを1次元配列として得られたものを用いることにする．

4.2 純2進符号とGray符号による比較

次に，1次元KSFMで作成されたコードブックの各ベクトルのコード番号を2進符号に変換する際，単純に端点のノードから純2進符号を割り当てた場合と，Gray符号を割り当てた場合についての違いを比較した[10]．

ここでGray符号とは，隣接する符号間では1ビットのみ変化する符号である．例えば，入力に与えるLSPパラメータのわずかの違いで，符号化ネットワークの出力が00000100あるいは00001100のように下位4bitめが0か1で違ったとしても，Gray符号であればコード番号が7か8のわずか1番違いに変換されるだけである．しかし，純2進符号の場合は4か12の違いとなりその差は大きい．従って，KSFMにより得られたコードブックでは，隣接するコードベクトル間

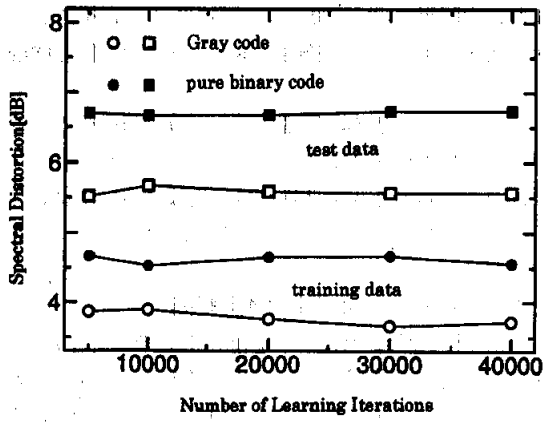

図1 純2進符号とGray符号で学習したときのスペクトルひずみ

Fig. 1 Spectral distortion trained with pure binary code and Gray code.

は類似度が大きいベクトル同士になっていることを考慮すると，Gray符号によりスペクトルひずみが軽減できることが期待される．

そこで，1次元KSFMで作成したコードブックのコード番号に純2進符号とGray符号を割り当て，量子化および復号化を行うネットワークを個別に学習し，スペクトルひずみを比較した結果，図1で示すようにGray符号を割り当てる方が，純2進符号よりも約0.9～1.2 dB小さくなった．

このことは，距離の近いLSPパラメータ同士には，表記の上でも近いコード番号を与えて教師データとすることが有効な方法であることを示している．このため以後の実験では，コード番号を表す2進符号にはGray符号を用いることにする．

4.3 サイクリック配列による学習

KSFMのノード配列には端点があり，始点と終点におけるノード間のコード番号差は大きい．しかし，1次元配列のコード番号にGray符号を割り当てる場合，始点と終点でのコード番号の2進表現は00000000か10000000であり，わずか1 bitの違いである．そこで，ノード配列を循環的に配置して（以後，サイクリック配列と呼ぶ），端点のない1次元KSFMを学習し，得られたコードブックのコード番号にGray符号を割り当てて，符号化および復号化を行うネットワークを学習することで更にひずみの軽減が期待できる．

しかし，実験の結果ノード配列がサイクリックの場合は，そうでない場合に比べて学習内，学習外データともにスペクトルひずみは約0.5 dB大きくなり，サイクリック配列の効果は認められなかった．

以上より，ノード配列が端点のある1次元KSFMから得られたコードブックのコード番号にGray符号を割り当てたときに最も良い結果が得られた．しかし，この場合でもスペクトルひずみが学習内データで3.6 dB，学習外データでは5.7 dBであり，LBG法などによる通常の8 bitベクトル量子化に比べ，それぞれ約0.8 dB，約1.7 dB大きくなった．

4.4 原LSPデータによる学習

これまでの実験においては，学習用データとしてKSFMで作成したコードブックのコードベクトルである256組のLSPデータを用いていた．ここでは，コードブック作成に用いた原データ（4208フレームのLSPパラメータ）を学習用データとして，再度3層の符号化および復号化ネットワークを学習し，スペクトルひずみを調べた．なおこの場合4208フレームの各LSPパラメータと，それをコードブックにより2進符号化したものを組にして学習用教師データに用いた．このときのコードブックは，端点ノードをもった1次元KSFMにより作成し，そのコード番号にGray符号を割り当てたものである．

スペクトルひずみは中間層ユニット数が10の場合で4.25 dB，20の場合で4.1 dBとなり，256組のパラメータを用いた場合よりも1 dB～1.5 dBひずみが大きくなった．これは，コードブックにはないLSPパラメータを用いて学習するため，そのLSPパラメータをこのコードブックを参照して2進符号化する際の誤差が加わるためと考えられる．

5. 5層ネットワークによる量子化

3層符号化および3層復号化ネットワークの結果を更に改善するため，符号化，復号化二つのネットワークを一つにまとめ，5層ネットワークとして再度学習を行うことによりひずみの低減を図った．

5層ネットワークで再度学習を行う際の重みの初期値として，通常行うランダムな初期値を用いる代わりに，3層符号化および復号化ネットワークにおける学習後の重みを用いることにする．3層符号化および復号化ネットワークではそれぞれ中間層が1層しかないのに対し，5層ネットワークでは中間層が3層に増えるため，更にひずみが小さくなることが期待される[11]．この実験では4.4で得られた重みを結合し，学習用データとしては4208フレームのLSPパラメータを用いた．

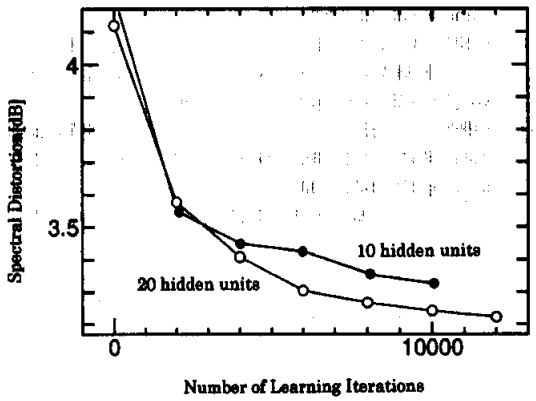

図2 5層ネットワーク 4208 フレームで学習したときのスペクトルひずみ

Fig. 2 Spectral distortion with a 5-layer neural network and LSP parameters of 4208 frames.

入出力層の素子数は LSP パラメータの分析次数に合わせ 8 とし，第 2 中間層は 8 bit 量子化に必要なユニット数 8，また第 1 および第 3 中間層ユニット数は 10 あるいは 20 の 2 通りとした．学習の際には，3. で述べた量子化学習（第 2 中間層ユニットの出力値を小数第 1 位で 4 捨 5 入し 0, 1 へ 2 値化）を行った．このときのスペクトルひずみを図 2 に示す．

第 1，第 3 中間層素子数が 10 の場合，5 層ネットワークの重み係数の初期値として，中間層素子数がそれぞれ 10 の 3 層符号化および復号化ネットワークを 5000 回学習した後のそれぞれの重み係数を結合したものを用いた．このとき，3 層符号化および復号化ネットワークでは 4.25 dB までしか下がらなかったスペクトルひずみが，5 層ネットワークとして再度学習することにより 3.3 dB（学習外で 4.1 dB）まで低下した．

同様に第 1，第 3 中間層素子数が 20 の場合，スペクトルひずみが 3.2 dB（学習外で 3.8 dB）まで下がり，5 層ネットワークとして再度学習することの効果が確認された．

また比較のため，5 層ネットワークの学習を行う際の重み係数の初期値として，ランダムな値を用いて学習を行ったときの学習後のネットワークのスペクトルひずみは，第 1，第 3 中間層素子数が 10 の場合は 3.8 dB，20 の場合は 4.0 dB となり，重みの初期値として 3 層符号化および復号化ネットワークの学習後の重みを結合して用いるよりもひずみが大きくなった．3 層の結果を 5 層の初期値として用いる場合でも，ネットワークは最適な重みに収束するわけではないがひずみを低下する効果があり，この初期値設定法は実際的なネットワークを構成する手法として意義がある．

6. むすび

本研究では，音声の低ビットレート符号化への応用を目的とし，階層型ニューラルネットを用いて，音声のスペクトル包絡を表す LSP パラメータのベクトル量子化を試みた．まず，恒等写像を実現するネットワークをベクトル量子化するには，中間層ユニットの出力を量子化しながら学習することが有効であることを示した．次に，あらかじめ，Kohonen の Self-Organizing Feature Map を利用して，3 層符号化および 3 層復号化ネットワークの学習のためのコードブックを作成した．その際，Map 上のノード配列を 1 次元とし，またコード番号は Gray 符号で 2 進表現することによりひずみを小さくすることができた．更に，5 層ネットワークをベクトル量子化に適用するには，その重みの初期値をランダムに与えるよりは，あらかじめ学習を行った 3 層符号化および 3 層復号化ネットワークの重みを用いることにより，スペクトルひずみを低減することができた．

ネットワークによる方法は，同じビット数では通常のベクトル量子化法に比べスペクトルひずみは若干大きいものの，演算量，メモリ必要量が少なくなるという利点があり，更にひずみ軽減のための方法を展開することが今後の課題である．また，本研究では 8 ビットで量子化した場合のスペクトルひずみだけで評価を行ってきたが，まだまだひずみが大きいため実際的な適用はできない．今後，合成音の聴取実験による主観評価により，どの程度のビット数が必要となるかを検討したい．

謝辞 本研究を進めるにあたり，移動通信システム（株）牧野忠由，八木敏男両氏から貴重なコメントを頂いた．ここに深謝します．

文 献

[1] R.M. Gray, "Vector quantization," IEEE ASSP Magazine, pp.4–29, April 1984.

[2] B.H. Juang and A.H. Gray, Jr., "Multiple stage vector quantization for speech coding," IEEE Proc., ICASSP-82, 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, pp.39–46, 1993.

[4] 大室 仲，間野一則，守谷健弘，"LSP パラメータのベクトル・マトリクス量子化，" 信学技報，SP91-70, pp.29–36, 1991.

[5] 牧野忠由，船田哲男，"低ビットレート音声符号化技術へのニューラルネットの応用，" 日本音響学会，3-P-10, 1991.

[6] 村本武志，船田哲男，八木敏男，"5 層ネットワークによる

LSPパラメータの量子化," 日本音響学会, 1-P-4, 1994.
[7] 森島繁生, 小松一樹, 片山泰男, 原島 博, "ニューラルネットに基づく音声情報圧縮," 信学技報, SP88-142, pp.61-66, 1988.
[8] J.L. Elman and D. Zipser, "Learning the hidden structure of speech," J. Acoust. Soc. Am, vol.83, no.4, pp.1615-1626, 1988.
[9] J. Dayhoff, Neural Network Architectures, Van Nostrand Reinhold, 1990.
[10] 船田哲男, 金寺 登, 田中秀治, "2進符号を教師信号にもつ階層型ニューラルネットの学習," 信学論 (D-II), vol.J73-D-II, no.9, pp.1574-1577, 1990.
[11] 船橋賢一, "3層ニューラルネットワークによる恒等写像の近似的実現についての理論的考察," 信学論 (A), vol.J73-A, no.1, pp.139-145, 1990.

(平成7年4月3日受付, 5月12日再受付)