ニューラルネットワークを利用したベクトル量子化と音声への応用に関する研究

氏　　名	森田　義則
学位の種類	博士（工学）
学位記番号	博甲第675号
学位授与の日付	平成16年3月31日
学位授与の要件	課程博士（学位規則第4条第1項）
学位授与の題目	ニューラルネットワークを利用したベクトル量子化と音声への応用に関する研究
論文審査委員(主査)	船田　哲男（工学部・教授）
論文審査委員(副査)	中山　謙二（工学部・教授），村本　健一郎（工学部・教授）， 木村　春彦（工学部・教授），平野　晃宏（工学部・講師）

学位論文要旨

Abstract

The occupied bandwidth of individual telecommunication devices in the field of mobile radio communication must be narrow in order to effectively exploit the limited frequency band. Vector quantization using a codebook has good performance in low bit-rate telephone-band speech codings. However, it has problems that consume many computations and memory storages. This thesis gives vector quantization using a multi-layer neural network. One of problems of this approach is how to reduce the learning error. The Kalman-Neuro-Training(KNT) method was used for quantization of LSP parameters. In experimental studies, the spectral distortion results in 1.41dB by 24-bit quantization for clean speech. In real applications, speech quality is easily deteriorated by background noise as the bit rate is decreased. Therefore, low bit rate speech coding, which has robustness against background noise, is required. Neural network vector quantization (NNVQ) is examined as a robust LSP encoder. We compare four types of binary patterns of a hidden layer, and clarify the dependency of quantization distortion on the bit pattern. By delayed decision (selection of low-distortion codes in decoding, i.e., EbD method) the LSP error can be decreased by 22%. For noisy speech, the performance of the EbD method is better than that of the conventional VQ codebook mapping method. Finally, we examine the LSP error for speech having different SNRs from that used in training. The experimental results show that training using SNR between 30 and 40 dB is appropriate. Considering these results, we conclude that the EbD is the best method for noisy LSP quantization.

1　はじめに

業務用通信や非常時通信等の移動体通信における音声通信の需要増に対応するため、音声の高能率符号化が求められている。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に分け、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、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、必要とする演算量やメモリ量が多いという問題点が残されている。

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、演算量、メモリ量とも約1けた小さくなることが期待できる。

同時に、ニューラルネットワークのもつ非線形な写像能力を利用し、雑音抑圧を考慮した符号化が可能であることを示す。

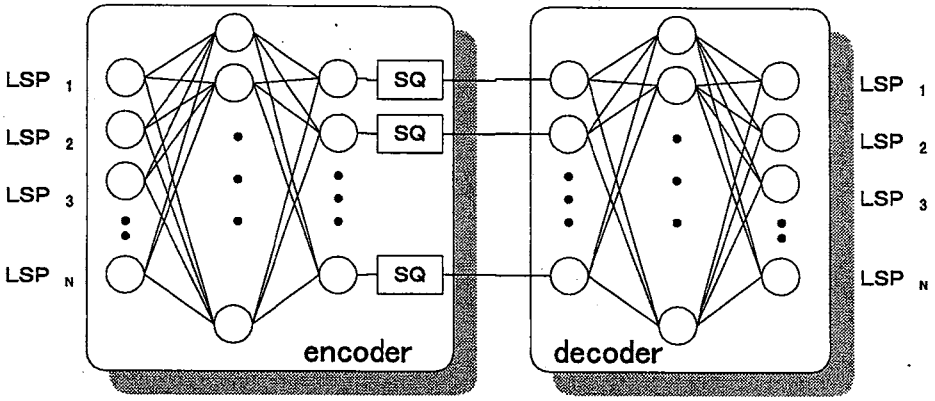
2 LSPパラメータのベクトル量子化

使用したNNの構造は図1の5層の階層型で、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データをベクトル量子化する。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。NNの学習には、誤差が小さくなることが期待できる、カルマンフィルターを用いたカルマンニューロトレーニング（KNT）法を用いた。

使用したデータは、日本音響学会連続音声データベースの音声から分析した、学習用として、２３，１３９個のLSPパラメータ、評価用として、２２，０６１個のLSPパラメータであり、評価はスペクトルひずみ(SD)により行った。

$$SD[dB] = \sqrt{\frac{1}{M}\sum_{i=0}^{M-1}\frac{1}{W}\int_0^W \{S_o(f,t) - S_d(f,t)\}^2\, df}$$

結果を表1に示す。最下段のDoDは、「DoD-CELP」符号化を用いたときのひずみであり、本研究の基準である。従来のコードブックを用いたVQ（2VQ）は最もひずみが少ない。スカラー量子化（SQ）や、従



SQ:スカラ量子化

図 1: 量子化ニューラルネットワーク

表 1: スペクトルひずみの比較

方式	学習内 [dB]	学習外 [dB]
2VQ	0.56	1.25
SQ	1.34	1.51
BP	1.50	1.68
KL	1.23	1.44
KNT	1.21	1.41
DoD		1.432

表 2: 演算量、メモリ量の比較

方式	演算量	メモリ量 [語]
2VQ	196614	65536
SQ	120	64
BP	1644	572
KL	304	143
KNT	1644	572

来の誤差逆伝播法を用いたVQ（BP）はひずみが大きい。KNT法を用いたニューラルネットワークベクトル量子化（KNT）は、DoDとほぼ同じひずみが得られ、またカルーネン・レーベ変換（KL）よりも、わずかではあるが、小さなひずみが得られた。

表2に演算量とメモリ量を比較した結果を示す。コードブックを用いたベクトル量子化は、演算量とメモリ量が非常に多く、一方、ニューラルネットワークを用いたベクトル量子化（BP、KNT）は、約1桁小さくなることが期待できる。

3 雑音音声のベクトル量子化

ビットレートが低くなるに従い、背景雑音による音声品質の劣化が起こりやすくなるので、雑音下での符号化性能が注目されている。そこで、雑音を加えた音声で雑音抑圧を考慮したベクトル量子化を検討した。このとき、5層のニューラルネットワークを用いる場合の問題点として：

- 多層のニューラルネットワークの学習が困難である。
- 誤差逆伝播法では、中間層に任意のパターンを形成させることができない。

が挙げられる。そのため、符号化および復号化ニューラルネットワークをそれぞれ個別に学習する方法を提案する。このとき、4種類の中間層符号パターンを検討し、比較を行う。更に改善する方法として、Encode by Decode法（ディレイド・デシジョン法）を適用した結果、従来のコードブックを用いる方法に比べて、雑音下では良い性能を示すことがわかった。

使用したデータは、ATR研究用日本語音声データベースSet C連続発声B50の分析より、学習データ用の１８，８０８個のLSPパラメータ、また、評価データには、日本音響学会研究用連続音声データベースより７，２４６個のLSPパラメータを用いた。誤差の評価は、平均２乗誤差の平方根を用いた。結果を、図3に示す。４種類の２進パターン、純２進符号(Pure)、グレイ符号(Gray)、ジョンソンカウンター符号(JC)と重み一定符号（CW；表3）の順にひずみが小さくなった。

表3: 重み一定符号

code number	constant weight code
0	0 0 0 0 ... 0 0 0 1
1	0 0 0 0 ... 0 0 1 0
2	0 0 0 0 ... 0 1 0 0
⋮	⋮
13	0 0 1 0 ... 0 0 0 0
14	0 1 0 0 ... 0 0 0 0
15	1 0 0 0 ... 0 0 0 0

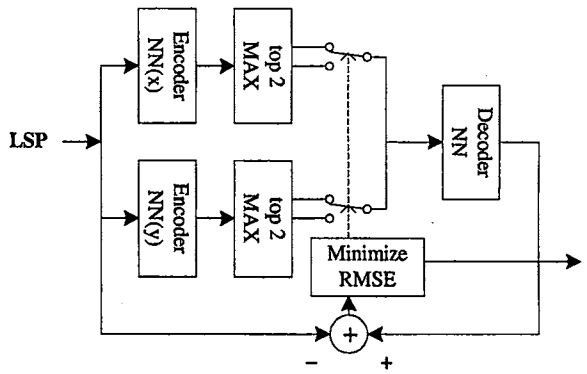


図2: EbD法による符号化

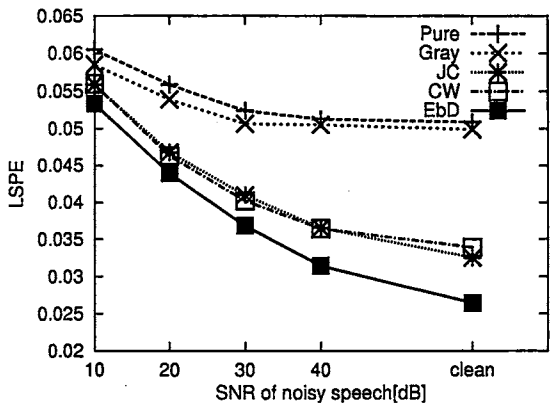


図3: 各種符号のLSP誤差による比較

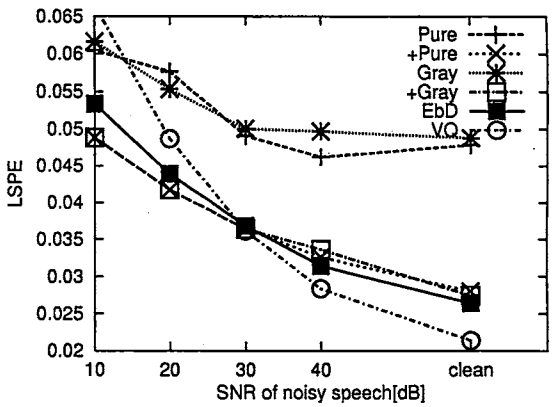


図4: 各種方式の比較

また、EbD法（図2）を適用した場合と、5層のニューラルネットワークを用いた場合(+Pure, +Gray)，および従来のコードブックを用いた場合(VQ)を図4に示す。EbD法は低雑音下では良い性能を示す。

4 まとめ

本論文では、ニューラルネットワークを利用したベクトル量子化と、その音声への応用について検討した。得られた主な結果を示す。

1. 「ニューラルネットワーク・ベクトル量子化」により、演算量、メモリ量を、従来のコードブックを用いたベクトル量子化に対して、約1桁減らせる。
2. クリーンな音声に対して、24ビット／フレームのレートで1.41dBの良好なスペクトルひずみを得た。
3. 重み一定符号とEbD法の組み合わせを新しく提案し、この方法は、提案法の中では、雑音下で最適な符号化法である。

学位論文審査結果の要旨

平成16年2月3日に第1回学位論文審査会を開催、2月4日に口頭発表、その後に第2回学位論文審査会を開催し慎重審議の結果、以下のとおり判定した。音声の特徴量であるLSPパラメータのベクトル量子化（VQ）に関しては、コードブック（CB）を用いたベクトル量子化法が用いられるが、雑音耐性に弱いことや、ビット数が多くなると、必要とする演算量やメモリ量が多くなるという問題点がある。いっぽう、ニューラルネットワーク（NN）を用いたベクトル量子化（NNVQ）には、ビット数が多くなるほど演算量、メモリ量ともCBを用いる方法に比べ相対的に少なくなる利点がある。そこで森田氏は音声符号化にNNVQを適用する際のVQひずみの低減、および雑音耐性向上を目的とした基礎的な研究を行った。まず、雑音がない場合は、「DoD-CELP」符号化を用いたときのスペクトルひずみをベースラインに設定し、他の量子化法による結果と詳細に比較した結果、コードブックを用いたVQが最もひずみが少ないこと、およびカルマンニューロ法で学習したNNVQはDoDとほぼ同じひずみで量子化できることを示している。さらに、雑音が重畳した音声をNNによって符号化するため、NNの中間層に各種の符号パターンを用いたときの性能の違いを比較実験している。ジョンソンカウンター符号（JC）や重み一定符号（CW）を用いるとひずみ軽減ができることや、さらにCWでEncode by Decode法（EbD）を適用すると、雑音下では最も良い性能を示すことを結論づけている。以上のように本論文はニューラルネットワークによる音声符号化に関する有用な結果を得ていることから、博士（工学）に値するものと判定した。