電子情報通信学会論文誌 '92/2 Vol. J 75-D-II No. 2

自己組織化特徴写像を用いた音声の極低ビットレート擬音韻符号化システム

准員 塙 雅典† 正員 長谷川孝明†

A Pseudo-Phoneme Coding System of Speech at Very Low Bit Rate Using Self-Organizing Feature Maps

Masanori HANAWA †, *Associate Member and*
Takaaki HASEGAWA †, *Member*

† 埼玉大学工学部電気工学科, 浦和市
Faculty of Engineering, Saitama University, Urawa-shi, 338 Japan

あらまし 音声信号の統計的性質を考慮して，自己組織化特徴写像を用いて音声を極低ビットレート符号化する音声符号化システムを提案し，その有効性について述べている．

キーワード：極低ビットレート音声符号化，音声認識，韻律情報，神経回路網，線形予測

1. まえがき

ディジタル化された音声信号は，計算機において保存再生を自由に行うことが可能であり，その特徴を用いて計算機ネットワーク上での音声メールシステムなどの実現が期待できる．しかしながら，従来の符号化方式では符号化に大量のビットを要しコストの面で問題があり，極低ビットレート音声符号化システムが望まれる．

音声信号の情報源は音韻や音節などのような離散シンボルの系列であると考えられるから，このような離散シンボルに符号化することによって極低ビットレート音声符号化が実現できると考えられる．これまでに，線形予測係数をベクトル量子化するLPC-VQ方式[1]やスペクトルパターンマッチング符号化方式[2]などが提案されているが，これらの手法ではスペクトル包絡情報のみをベクトル量子化などを用いて離散シンボルに符号化し，基本周波数や利得などの韻律情報はスペクトル包絡情報とは別に毎秒100ビット以上用いて符号化する．一方，これに対して音声認識を行い言語情報のみを伝送することで，究極の極低ビットレートをねらった音声認識ボコーダが提案されている[3]．このようなシステムでは，音声を音韻や音節などの離散シンボル列に符号化し，規則合成方式によって復号音声を得る．しかし，規則合成方式では規則にしたがって生成された韻律で音声を合成するため，もとの音声の韻律は失われてしまう．そこで本論文では，音声信号の統計的性質を考慮し，自己組織化特徴写像を用いてスペクトル包絡情報と韻律情報の両方を含んだシンボルに符号化することで，もとの音声の韻律を残したまま極低ビットレートを実現するシステムを提案し，その性能評価を行う．

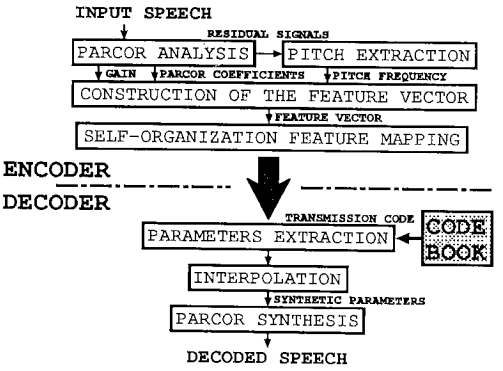

図1 擬音韻符号化システムの構成
Fig. 1 Pseudo-phoneme coding system configuration.

2. 極低ビットレート擬音韻符号化システムの提案

本章では，自己組織化特徴写像[4]を用いた極低ビットレート音声符号化システムを提案する(図1)．このシステムの特徴は，韻律情報を残したまま音声を擬音韻と呼ぶ音韻に近いシンボルに符号化し，音声認識をせずに極低ビットレートを実現する点である．

2.1 擬音韻符号化システムの提案

音声信号は，利得，基本周波数，ホルマント周波数などの情報で表現できる．従来の音声符号化システムでは，これらを表現するパラメータを別々に符号化している．しかし，利得が高くなると基本周波数も高くなり，更に，ホルマント周波数が高くなると基本周波数も高くなるというように，これらの間には相関があることが指摘されている[5]．そこで，利得や基本周波数やホルマント周波数を表現するパラメータの組合せをコードブックに保持することによって，音声信号を一つのシンボルで表現することが可能となる．このシンボルは，スペクトル的には音韻を表現するものと考えられるが，実際の音韻と1対1に対応しているわけではないので，これを擬音韻と呼ぶことにする．本論文では，音声信号をこの擬音韻に符号化することによって，極低ビットレート音声符号化を実現するシステムを提案する[6]．

2.2 システムの構成

提案するシステムでは，まず音声信号をPARCOR分析し，基本周波数をピークの位置によって表現するベクトル(ピッチベクトル)に変換し，利得，PARCOR係

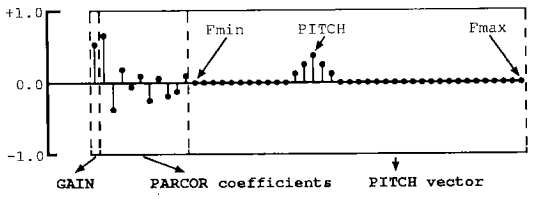

図2　特徴ベクトルの例
Fig. 2　An example of the feature vector.

数と組み合わせて特徴ベクトルを構成する(図2)．このベクトルを擬音韻に符号化するために，Kohonenによって提案された自己組織化特徴写像を用いる．このアルゴリズムは，学習によって学習用ベクトルの確率分布に従って出現頻度の高い代表的なベクトルを参照ベクトルとして獲得する．学習用音声を用いて学習を行った参照ベクトルに符号を割り当てることで，擬音韻符号化用のコードブックが得られる．符号化は，入力音声から構成した特徴ベクトルに最適整合する参照ベクトルの符号を伝送することで行う．

復号器では，伝送された符号に応じて，コードブックから音声の合成に必要なパラメータを抽出する．基本周波数は，学習された参照ベクトル中のピッチベクトルのピークの位置から決定される．このようにして抽出された各パラメータを線形補間した後，ピッチ同期合成によって音声を合成する．

3. 計算機シミュレーションによる音声符号化実験

3.1 符号化実験の諸元

提案するシステムの性能評価のために，計算機シミュレーションによる符号化実験を行った．シミュレーションの諸元を表1に示す．基本周波数は，その分布範囲を約70～250 Hzと仮定し，32段階に線形量子化を行った．参照ベクトルは無声音用と有声音用に分割し，それぞれ256個と768個用意した．参照ベクトルの総数は1,024個であることから，特徴ベクトルで表された音声を符号化するのに必要なビット数は10ビットであり，符号化周期を30 msとしたときの伝送速度は334 bit/sである．

3.2 復号音声の品質評価

評価用音声を用いて，復号音声の品質評価を行った．評価基準には，主観評価とよく対応するといわれているLPCケプストラムひずみ(CD)[7]を用いた．CDは16次のLPCケプストラムから計算した．比較のために，2,400 bit/sのPARCORボコーダについてもCDを求めた．その結果を表2に示す．提案するシステム

表1　計算機シミュレーションの諸元

音声データ	
発声者	男性1名
サンプリング周波数	10 kHz
量子化ビット数	8 bit
フレーム数	学習用　6,000個(60秒)
	評価用　6,000個(60秒)
PARCOR分析	
分析窓	ハミング窓
分析窓長	30 ms
分析周期	10，20，30 ms
分析次数	10次
ピッチ量子化	32段階
自己組織化特徴写像	
参照ベクトル数	1,024個
距離尺度	ユークリッドノルム

表2　ケプストラムひずみによる評価

フレーム周期 [ms]	ビットレート [bit/s]	CD値 [dB]
10	1,000	3.34
20	500	3.65
30	334	4.23
PARCOR	2,400	3.32

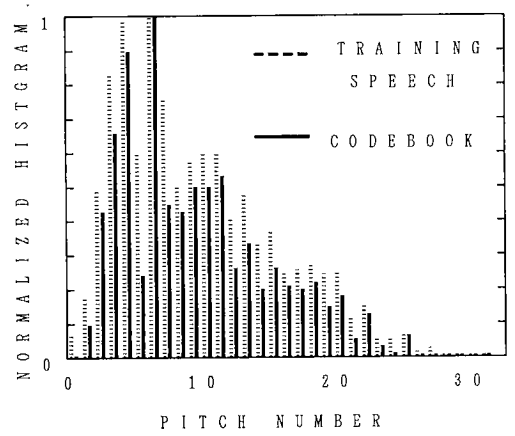

図3　基本周波数のヒストグラム
Fig. 3　Histgram of the fundamental frequency.

は符号化周期10 msすなわち1,000 bit/sで，2,400 bit/sのPARCORとほぼ同程度の品質の音声が得られている．また聴取実験を行ったところ，334 bit/sでも元の音声の韻律情報を残し，発声内容の聴取が十分に可能な音声が得られることがわかった．また，有声音用参照ベクトルに獲得された基本周波数のヒストグラムを求め，学習用音声から抽出された基本周波数のヒ

ストグラムと比較したところ，かなりよくその分布を学習していることがわかった(図3)．これらの結果から，本システムは韻律情報を残したまま極低ビットレート音声符号化を行うことが可能なシステムの一実現法であると言える．

4. むすび

韻律情報を残したまま音声信号をシンボル列に変換し，音声認識をせずに極低ビットレート音声符号化を行う擬音韻符号化システムを提案し，LPCケプストラムひずみを用いてその性能評価を行った．1,000 bit/sで符号化した音声のケプストラムひずみは2,400 bit/s PARCORとほぼ同程度であり，334 bit/sで符号化した音声ももとの音声の韻律情報を残し内容の聴取が十分に可能な音声が得られ，本システムの有効性が確認された．また，自己組織化特徴写像による学習で獲得された参照ベクトルが表現する基本周波数は，学習音声から抽出された基本周波数の分布をよく近似していることがわかった．

本システムは複数のパラメータ間の相関を利用して極低ビットレート音声符号化を実現しているため，学習に使用した音声の発声環境に依存してしまい，学習に用いた音声の発声環境から大きく異なる音声(学習外話者の音声や，意図的に韻律を変えて発声された音声)に対しては，復号音声の品質が劣化する．これまでに行った実験において，学習外話者の音声に対する復号音声は品質劣化が大きく，発声内容を聞き取ることすら困難であった．ほかにも，特徴ベクトルのユークリッド距離を最小化するように符号化するため主観的品質が最良にならない，復号音声の品質が特徴ベクトルを構成する各パラメータへの重み配分によって変化するなど，更に検討を要する点を含んでいる．これらについては現在検討中であり，別途報告の予定である．

今後は，特徴ベクトルの構成について詳細な検討を加え品質向上を図ると共に，更に極低ビットレートを実現するシステムの検討を行っていく予定である．

謝辞 日ごろから御指導頂く埼玉大学羽石操教授に深謝致します．

文献

(1) Wong D. Y., Juang B. H. and Gray A. H.: "An 800 bit/s vector quantization LPC vocoder", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-30**, 5, pp. 770-779 (Oct. 1982).

(2) 菅村 昇，板倉文忠："スペクトルパタンマッチングによる音声情報圧縮"，信学論(A)，**J65-A**, 8, pp. 834-841 (1982-08).

(3) 中川聖一，安本太一："音声認識ボコーダ(100 bit/s)"，昭63信学春季全大，SA-4-9.

(4) Kohonen T.: "Self-Organization and Associative Memory", Springer-Verlag (1989).

(5) 三浦種敏監修："新版聴覚と音声"，電子情報通信学会(1989).

(6) Hanawa M., Hasegawa T. and Hakura Y.: "A Speech Coding System at Very Low Bit Rate using Self-Organization Neural Network", Proc. ISITA '90, 29-4 (1990).

(7) Kitawaki N. and Nagabuchi H.: "Objective quality evaluation for low-bit-rate speech coding systems", IEEE J. Sel. Areas Commun., **6**, 2, pp. 242-248 (1988).

(平成3年7月8日受付，9月4日再受付)