様式 C-19

科学研究費補助金研究成果報告書

平成２３年　５月２７日現在

機関番号：１６１０１
研究種目：若手研究（B）
研究期間：2008 ～ 2010
課題番号：20760229
研究課題名（和文）　ML-BEATS 法を用いた高効率音声符号化法の開発

研究課題名（英文）　A new speech coding system based on ML-BEATS

研究代表者
鈴木 基之（SUZUKI MOTOYUKI）
徳島大学・大学院ソシオテクノサイエンス研究部・准教授
研究者番号：30282015

研究成果の概要（和文）：本研究では，ML-BEATS 法を用いて音声信号中の類似区間を見つけ，それをひとつの符号として音声符号化を行うことで，極低ビットレートで品質の高い音声符号化を開発した。その際，「次元に意味を持たない」という LSP の特性を考慮してセグメント量子化を行った。従来法（G.729）と比較して，ビットレートを１フレームあたり 18bit から 12.4bit へと低減させることができたが，スペクトル歪は 1.02dB から 1.8dB へと増加してしまった。

研究成果の概要（英文）：In this research, a new speech coding system based on ML-BEATS has been developed. ML-BEATS method can split an input sequence into many sub-sequences and cluster these. This method has been applied to segment quantization. Experimental results showed the proposed method decreased bit rate, but increased spectral distortion compared with the G.729.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	900,000	270,000	1,170,000
2009 年度	1,300,000	390,000	1,690,000
2010 年度	900,000	270,000	1,170,000
年度			
年度			
総　計	3,100,000	930,000	4,030,000

研究分野：音声情報処理
科研費の分科・細目：通信・ネットワーク工学
キーワード：ML-BEATS 法，音声符号化，セグメント量子化，HMM

１．研究開始当初の背景

携帯電話やインターネットの普及に伴い，音声情報を圧縮・符号化する技術は今後ますます重要となってくると思われる。現在のところ，入力された音声をLSP 係数等の特徴量ベクトルの時系列へと変換し，それをベクトル量子化やセグメント量子化などを用いて符号化する，という方式が主流である。

この時，各符号に対応する代表ベクトル（や代表セグメント）は一般に平均ベクトルが用いられ，復号側では，各符号に対応する代表ベクトルがそのまま再生される。離散化された符号と再生するベクトルが直接対応しているため，再生されるベクトル系列は時間的に不連続となる。また，平均ベクトルをそのまま再生するだけであり，分散といった高次の

統計量は一切考慮されていない。

２．研究の目的

本研究では，HMMを用いた音声合成法で用いられている考え方を導入し，時間的に滑らかなベクトルを再生する方法を開発する。その際，HMMでモデル化されることで，高次の統計量を用いた高品質な再生を可能とする。

３．研究の方法

(1) HMM 音声合成法の利用

HMM音声合成法では，あらかじめ大量の音声データをHMMでモデル化し，それを用いて音声を合成している。モデル化の時，特徴量ベクトル中の各次元ごとに時間方向の差分を計算し，それもあわせて特徴量としている。また，HMMでモデル化することで各代表ベクトルを平均だけではなく分散も持った確率分布で表現している。音声を合成する際には，各次元の差分情報を拘束条件とした上で，なるべく確率分布からの出力確率が高いベクトル系列を計算し，そこから音声を合成している。結果として，滑らかなベクトル系列を再生することを可能としている。

そこで，提案する音声符号化法においても各次元の差分情報を加えたものを特徴量ベクトルとし，これらを確率分布でモデル化したものを代表ベクトルとする。こうしてHMM音声合成法をそのまま流用し，復号側で滑らかなベクトルを再生させる。

このため，音声データをHMMでモデル化する必要がある。HMM音声合成法では合成する発話内容との対応をとる必要があったため，音素を単位としてHMMが構築されていた。しかし今はそうした対応は不要であり，より効率的な単位でHMMを構成することが重要である。そこで，時系列中から類似部分系列を自動的に抽出するML-BEATS法を用いる。

(2) ML-BEATS法を用いた音声符号化

HMMを構築する単位としては，時間的に類似している部分系列が望ましい。つまり，時間軸方向に見て，同じようなベクトル系列がいろいろな場所に出現しているとすれば，それらを集めてきて1つのHMMでモデル化できればよい。このような類似部分系列を自動的に抽出し，HMMを構築する方法がML-BEATSである。

この方法は入力されたベクトル系列から類似する部分系列を抽出し，クラスタリングした上でそれぞれをHMMでモデル化する。部分系列の抽出やHMMのパラメータ推定などは，すべて「尤度最大」を基準として同時に最適化される。

このようにして得られた各HMM を符号語として登録すれば，確率モデルを用いた音声符号化が実現できる。符号化側はHMM の番号と各状態に停留するフレーム数を送信し，復号化側ではそれらの情報をもとに，HMM 音声合成法で用いられている方法を使ってベクトル系列を再生する。

(3) LSP係数の特性

音声符号化を行う際，現在最もよく行われている方法は，音声を音源（声帯）情報とスペクトル（声道形状）情報に分離し，それぞれを符号化して送信している。この時，スペクトル情報はLSP係数というパラメータを用いて表現している。

LSP係数は1フレームあたりn次元のベクトルとして表現される。その値は，線スペクトルが存在する周波数の値であり，低い周波数から，1次元目，2次元目・・・n次元目としてベクトルに表現される。

ここで，直前のフレームには近い周波数に存在していた2本の線スペクトルが，次のフレームでは1本に統合されてしまう，といったことが時々観測される。また逆に，1本の線スペクトルが次の時刻では2本に分離する，といったことも観測される。

こうした現象が起きると，今までn次元目にあった線スペクトルがn+1（またはn-1）次元目になってしまう，ということが起こる。つまり，LSP係数を表現したベクトルにおける次元とは，低い周波数側から何番目，という意味しかなく，いわゆる3次元空間の2次元目の軸の値，といったような意味はないことになる。こうしたベクトルをモデル化する時，ほぼ同じ周波数に線スペクトルが並んでいたとしても，たまたま低周波数側に1本多ければ，それ以降の次元がすべて1つずつずれるため，全く異なるベクトルとなってしまう。そこで，こうした「次元のずれ」に対応したモデル化を行う。

(4) 次元のずれに対応したモデル化

低次元に線スペクトルが追加される，または消されると，それ以降の次元はすべて1つずつずれる。こうしたことが起こると，本来同じベクトルが全く異なるベクトルとなってしまうため，モデルとマッチできず，効率のよい符号化が行えない。

そこで，あらかじめ次元のずれたベクトルを機械的に作りだし，それらをすべてモデル化しておくことで，符号化の際に次元がずれたベクトルが入力されたとしても正しく符号化できるようにする。具体的には，モデルの学習時に1～n次元のベクトルがあったとすると，それから2～n+1次元，3～n+2次元，といったベクトルを作り出し，それらをすべて1つのモデルで学習する。

この方法では，まとめて次元がずれた場合には対処可能であるが，途中の次元からずれが生じた場合はあいかわらず対処ができない。例えば，1～n次元のものがまとめて2～n+1次元へとずれればよいが，3次元目の線スペクトルが分離した，といったような場合には対処できない。

そこで，こうした場合にも対処できるよう，各次元をすべてバラバラにする方法も提案する。1次元目，2次元目，というようにすべての次元を別々にし，それら1次元ベクトルの時系列をまとめて1つのHMMでモデル化する。こうすることで，「途中」という概念がなくなるために，どこの次元からずれが始まっても問題はない。

しかし，この方法はもはや「ベクトル」量子化ではなく，「スカラー」量子化となる。一般にベクトルをそのまま量子化した法が量子化効率はよく，結果として低いビットレートで高い品質の音声が再生される。そのため，この方法は原理的にビットレートが高くなってしまう。その分，LSP係数の次元の分離や結合には強くなるため，どちらの方がよいかは，実験によって評価する。

(5) Huffman符号化による効率的表現

前述のように1つのHMMで多数の「ずらした」ベクトルをモデル化した場合，多様なベクトルを表現することになるため，モデルの規模が大きくなり，それに伴って送信する符号長も長くなる。そこで，Huffman符号化を効率的に適用し，符号長を短くすることを試みる。

Huffman符号化は，高い頻度で使用されるものには短い符号を，低い頻度で使用されるものには長い符号を割りあてることで，実質的なビットレートを下げよう，というものである。本研究で構築するHMMは，1～3次元，といった低い次元のベクトルから，8～10次元，といった高い次元のベクトルまで一緒になってモデル化されている。実際に1～3次元のベクトルを符号化しようとした場合，モデルの一部だけが頻繁に使用されることは十分に考えられるため，Huffman符号化を行うメリットは大きいと思われる。

そこで，1～3次元，4～6次元といったように，次元別にHuffman統計をとっておき，実際に符号化する際にはそれぞれ対応する

Huffman統計を用いて符号化を行う。こうすることで，実際に符号化しようとしている次元に近いベクトルには短い符号が割りあてられていることになり，低ビットレートを実現する。

４．研究成果

(1) 実験条件

提案する方法の性能を評価するため，音声の符号化実験を行った。ML-BEATSを用いたモデルの構築には，30名が発声した音声（1,500文）を用いた。符号化実験には，別の12名が発声した600文を用いた。

LSPの分析次数は10であり，LSP係数は10次元のベクトルで表現される。これをそのまま量子化すると学習サンプルの不足による過学習が起きてしまうため，従来から1～3次元，4～6次元，7～10次元，といったように3つにベクトルの次元を分割して，それぞれ別々にベクトル量子化することが行われていた。そこで，提案する方法においても，1～3次元，2～4次元，3～5次元・・・，のように3次元で1つのベクトルとし，それらの次元を1つずつずらしながら学習用のベクトルを作成し，モデル化を行った。また，1つのベクトルにまとめる次元数を3だけではなく，4～7まで増やしていった時の性能についても評価を行った。

(2) 基本的な性能評価

まずは，次元をずらすことを行わず，ML-BEATS法を用いたモデルの構築と，HMM音声合成法のアルゴリズムを流用した音声の復号化，という基本的な枠組みの有効性を確認した。比較対象としては，従来からよく用いられ，ITU-Tによって規格化されているG.729方式を用いた。

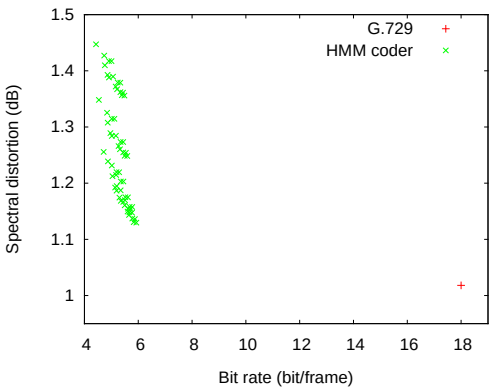

図1 提案法による音声符号化性能

結果を図１に示す。この図において，横軸はビットレート，縦軸はスペクトル歪である。どちらも共に低い方が望ましいため，グラフ上で左下にあればあるほどよい方法，ということである。また，緑の点は（パラメータを様々に変化させた）提案方法の結果，赤い点は従来法である G.729 の結果である。
このグラフを見ると，提案法は非常に低いビットレートで符号化できていることがわかる。G.729 と比較すると，およそ 1/3 程度になっている。しかし，スペクトル歪は大きくなり，音声の品質が低下していることがわかる。モデル化のパラメータを変化させることで，（ビットレートは高くなるが）スペクトル歪を低くさせることは可能であると考えられるが，学習にかかる計算時間が膨大になってしまうため，現時点ではあまり現実的ではない。

(3) LSP 係数の特性への対応結果

次に，学習時に次元をずらしたベクトルを機械的に作り出し，それらをすべて用いて HMM を学習した結果を示す。

この図において，色の違いは１つのベクトルの次元数の違いを表している。また，赤い点の集合は，図１の提案法の結果である。

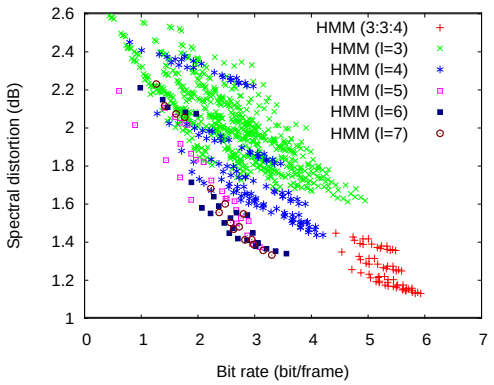

図 2 次元のずれを考慮した結果

この結果を見ると，次元のずれに対処をすることで，より低ビットレートだが，スペクトル歪の大きいモデルとなってしまっていることがわかる。このことから，LSP 係数の特性をうまく吸収できた，とはいえない結果であったといえる。これについては，今回の対処法が効果的ではなかったのか，それとも，そもそも LSP 係数の特性による性能劣化が低かったのか，別途検討を行う必要がある。

(4) 1 次元時系列のモデル化

最後に，1 次元のスカラー時系列としてモデル化した時の結果を示す。この時は，12.4bit/frame で，1.8dB のスペクトル歪となってしまった。やはりスカラー量子化することで量子化効率が落ち，ビットレート，スペクトル歪ともに上がる結果となってしまった。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　4 件）

(1). Motoyuki Suzuki, Masashi Adachi, Minoru Kohata, Akinori Ito, Shozo Makino, Fuji Ren, An HMM-based segment quantizer and its application to low bit rate speech coding , Proc. International Congress on Acoustics, CD-ROM, 査読有, 2010 年
(2). 足立 征士, 鈴木 基之, 任 福継, LSP係数の性質を考慮した音声符号化法の改善, 情報処理学会全国大会講演論文集, Vol. 2, pp.205-206, 査読無, 2010 年
(3). 足立 征士, 鈴木 基之, 任 福継, ML-BEATS法を用いたLSP係数のセグメント量子化法の検討, 電気学会 電子・情報・システム部門大会論文集, pp. 723-725, 査読無, 2009 年
(4). 大越 真裕美, 鈴木 基之, 大河 雄一, 伊藤 彰則, 牧野 正三, 混合重み再学習を用いた単語モデルによる連続音声認識, 日本音響学会 2009 年春季研究発表会講演論文集, pp. 177-178, 査読無, 2009 年

〔学会発表〕（計　4 件）

(1). Motoyuki Suzuki et al., An HMM-based segment quantizer and its application to low bit rate speech coding , International Congress on Acoustics, 2010 年 8 月 26 日, Sydney Convention Centre (Australia)
(2). 足立 征士他, LSP 係数の性質を考慮した音声符号化法の改善, 情報処理学会 創立 50 周年記念全国大会, 2010 年 3 月 9 日, 東京大学（東京都文京区）
(3). 足立 征士他, ML-BEATS 法を用いた LSP 係数のセグメント量子化法の検討, 電気学会 電子・情報・システム部門大会, 2009 年 9 月 4 日, 徳島大学（徳島市）
(4). 大越 真裕美他, 混合重み再学習を用いた単語モデルによる連続音声認識, 日本音響学会 2009 年春季研究発表会, 2009 年 3 月 17 日, 東京工業大学（東京都目黒区）

６．研究組織

(1)研究代表者

鈴木　基之（SUZUKI MOTOYUKI）
徳島大学・大学院ソシオテクノサイエンス研究部・准教授
研究者番号：30282015

(2)研究分担者

（　　）
研究者番号：

(3)連携研究者

（　　）
研究者番号：

(4)研究協力者

足立　征士（ADACHI MASASHI）
徳島大学・大学院先端技術科学教育部システム創成工学専攻・博士前期課程学生