TR-I-0202

不特定話者音声認識に関する検討
Study of Speaker Independent Speech Recognition

上田佳央、　服部浩明
Yoshio UEDA　Hiroaki HATTORI

概要

不特定話者音声認識に関する基礎検討について報告する。始めに不特定話者用コードブックの作成方法およびコードブックサイズについて検討を行った。作成方法では、話者全員の音声を用いて一度にクラスタリングを行う方法(1段階作成法)、話者ごとにコードブックを一度作成した後、各コードワードの出現頻度を考慮して、全話者のコードブックを用いて再クラスタリングを行う方法(2段階作成法(重み付き))、2段階作成法において頻度を考慮しない方法(2段階作成法(重み無し))に付いて比較を行った。その結果、2段階作成法(重み付き)によって1段階作成法と同程度のディストーションを持つコードブックが得られることがわかった。また、コードブックサイズは、WLR,pow,dcepの各パラメータについて1024,64,512とすることで特定話者並み(サイズ256,64,256)のディストーションが得られた。

つぎに、得られたコードブックを用いて、不特定話者音韻認識実験を行った。このとき、学習時と認識時にそれぞれHard VQとFuzzy VQを用いた場合の四通りについて実験を行い、不特定話者におけるHardVQとFuzzyVQによる認識率の差はあまりなく、/b,d,g,m,n,N/ の認識において約70%の音韻認識率が得られた。

ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所

目次

1.はじめに

ATRでは、これまでに、HMMに基づく特定話者音声認識およびFuzzyVQによる話者適応化方法に関する研究を行ってきた。その結果少数の学習サンプルにより話者の適応化が可能であることを示した。

一方、何ら前処理を必要としない不特定話者認識は今後の重要な課題の一つであり、研究を行う必要がある。

本報告では、不特定話者音声認識の第一段階として、不特定話者用コードブック作成方法の比較とコードブックサイズの検討を行い、得られたコードブックによる音韻認識評価実験を行った。

2.コードブック作成方法の検討

2.1 コードブックの作成方法

コードブックは、2分割LBGアルゴリズムをもちいるが、作成方法として次の三つの方法を比較する。

Ⅰ 学習データをそのまますべて用いる1段階作成法。

Ⅱ 一度、特定話者用コードブックを作成し、そのコードブックを頻度に応じた重み付けを考慮して不特定話者用コードブックを作成するという次のような二段階作成法。

①各話者の学習データの全フレームから特定話者用コードブックを作成し、得られた各コードにクラスタリング時の所属フレーム数に比例した重みを与える。

②この重みを考慮しながら、各話者のコードを再度クラスタリングすることによって、最終的な不特定話者用のコードブックを作成する。

Ⅲ Ⅱにおいて重み付けを行わないもの。

また、不特定話者用コードブックサイズの検討を行う。

作成方法およびコードブックサイズの検討を行う際には、コードブックを作成したのとは別の単語のディストーションを用いて評価を行った。

2.2 音声資料

コードブック学習データ及び評価データを表1に示す。この音声資料を12kHzでサンプリングし、窓長21.3msec、周期3msecのハミング窓で切り出し、14次のLPC分析を行う。

分析により、WLR(自己相関係数)、pow(パワー)、dcep(ケプストラムの51msecにわたる回帰係数)の3種類の特徴量を抽出し、コードブックを作成する。

表1 音声データA

話者	男性･女性 各8名、計16名
コードブックの 学習データ 評価データ	音韻バランス216単語の 前半100単語 後半100単語

2.3 ディストーションの評価式

WLR,pow,dcepのディストーションは、以下の式で評価する。

ベクトルxとコードブックcの距離

$$D_{WLR}(x,c) = \sum_{i=1}^{p} (Cx_i - Cc_i)(Rx_i - Rc_i) \quad -(1)$$

$$D_{pow}(x,c) = Px/Pc + Pc/Px - 2 \quad -(2)$$

$$D_{dcep}(x,c) = (\sum_{i=1}^{p} (\Delta Cx_i - \Delta Cc_i)^2)^{1/2} \quad -(3)$$

p:	ベクトルの次数
C:	ケプストラム係数
R:	自己相関係数
P:	パワー
ΔC:	ケプストラムの回帰係数

2.4 コードブックの評価結果

図1にWLRコードブックのディストーションを示す。これをみると、1段階作成コードブックと2段階作成(重み付き)コードブックのディストーションにほとんど違いは見られず、サイズが1024のとき特定話者用コードブック(サイズ256)と同程度のディストーションになることがわかる。

図2にpowコードブックのディストーションを示す。

これをみると、2段階作成コードブックはディストーションが同じ値となった。調べた結果、ディストーションだけでなく、コードブック自体がほとんど同じ値となっていることが分かった。これは、コードブックサイズに対して学習データが十分に多いため、重みを付けても付けなくてもベクトルの分布が変化しないためだとおもわれる。

また、2段階作成コードブックのディストーションが他のコードブックに比べて非常に大きい。調べた結果、2段階作成を行うと1段階目のコードブックを作成した時点で、個々の値が殆ど0のサンプルが無くなってしまい、最終的に得られるコードブック中に値が0に近いものが存在しなくなることが分かった。そのため、評価用データのなかに値が殆ど0のサンプルが存在する場合には、評価式(2)の非対称性のために、ディストーションが大きくなる。

図3はdcepコードブックのディストーションである。1段階作成コードブックと2段階作成(重み付き)コードブックでは、特定話者用コードブックとほぼ同程度のディストーションが得られている。

以上の結果より、WLRとdcepのコードブックに対しては、2段階作成(重み付き)法でも1段階作成法と同程度のディストーションを得られることが分かった。

また、コードブックサイズでは、2段階作成(重み付き)法を用いた場合WLR,dcepのサイズがそれぞれ1024,256のとき、特定話者のサイズ256,256と同程度のディストーションを得られることが分かった。

powのコードブックサイズについては、ディストーションの評価式が適切ではなかったために、この実験では正しい評価はできない。しかし、特定話者音声認識の研究結果から、サイズは64で十分であろうと思われる。

3.音韻認識実験

2段階作成(重み付き)コ　2章で得られたコードブックを用いて、音韻認識実験を行った。この時、学習時と認識時にそれぞれFuzzyVQとHardVQを用いた場合の四通りについて実験を行った。

3.1　音韻認識実験方法

2段階作成(重み付き)コードブックを用いて、つぎの条件で音韻認識実験を行う。

①16名中、15名の音声データからコードブック作成、HMM学習を行い、他の1名の認識を行う15:1のオープン実験とする。16名の話者はAセット中のFFS、FKN、FKS、FMS、FSU、FTK、FYM、FYN、MAU、MHT、MMS、MMY、MNM、MSH、MTK、MXMを用いた。認識実験はMAU、MHT、MNM、FSUの4人を未知話者として行う。

② 語彙に/b,d,g,m,n,N/の6音韻を用いる。そのうち/b,d,g/は学習時に語頭、語中をわける。

③ HardVQとFuzzyVQを学習、認識の両方に用いて4通りの実験を行う。このときFuzzyVQはFuzziness 1.6、k＝6で行う。

また、参考のため特定話者と話者適応の認識結果を示す。特定話者の実験条件は、WLR,pow,dcepのコードブックサイズをそれぞれ256,64,256として、あとの条件は同じにしたものである。話者適応の実験条件は、WLR,pow,dcepのコードブックサイズをそれぞれ256,64,256として、話者FSUをMHT,MNMに適応したものと、MAUをMHT,MNMに適応したものの平均である。適応には、コードブック作成に用いたのと同じ100単語を用いている。

3.2　音声資料

音韻認識実験に用いる資料を表2に示す。

表2　音声データB

話者	男性･女性　各8名、計16名
コードブックの 学習データ 評価データ	重要語5240単語の 偶数番目の単語中の音韻 奇数番目の単語中の音韻

3.3　音韻モデル

HMMの学習に用いる音韻モデルを図4に示す。4状態3ループで「結び」がひとつある。

3.4　音韻認識実験結果

表3に音韻認識実験結果を示す。

第1位の認識率をみると、どのVQ方法を用いても約70%で、FuzzyVQとHardVQの間に有意な差は見られなかった。一般に、訓練サンプルが少ない場合にはFuzzyVQが、多い場合にはHardVQが良い結果を与えることが知られている。今回の実験では、特定に場合に比べて訓練サンプルは多いが、コードブックサイズ、表現すべき音響現象が増加しているため、サンプル数が多くなったと

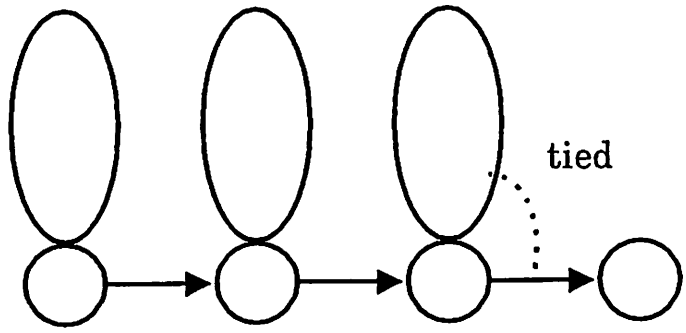

図4　HMMの音韻モデル

表3　/b,d,g,m,n,N/ の音韻認識結果
hard:HardVQ
fuzzy:FuzzyVQ　fuzziness 1.6　近傍数6
話者MAU,MHT,MNM,MSUの平均

VQ方法		認識率　(%)			参考データ	
学習	認識	1位	2位	3位	特定話者	話者適応
fuzzy	fuzzy	69.9	85.3	92.2	85.9	63.0
fuzzy	hard	69.6	86.0	93.0	88.3	-
hard	fuzzy	70.6	86.0	92.9	88.9	-
hard	hard	69.5	86.5	93.9	90.0	-

も少なくなったとも言えず、このような結果になったと思われる。
　また、特定話者と比較した場合には、約15~20%の劣化となっている。誤りの多くは /m,n/ でのものである。これらの音韻の識別は微妙なスペクトルの際を用いて行わなければならないが、多数話者の音声を用いて学習を行うことによって、その特徴がボケてしまい誤りが増えたと考えられる。
　話者適応の場合とは、実験条件が異なるため直接比較は難しい。しかし、認識率では話者適応と同等以上の結果となっており、不特定話者の音声認識においてもHMMの手法が有効であると言える。

4.まとめ
　2段階作成(重み付き)コードブックのサイズがWLR,dcepが1024,256で、特定話者用コードブックのサイズ256,256と同程度のディストーションを得た。
15:1の音韻認識実験を行った結果、話者4名の平均で約70%の認識率をえた。
音韻認識実験では、FuzzyVQとHardVQの間に有意な差は見られなかった。
不特定話者音声認識の研究としては話者16名では足りないと考えられるので、今後は、もっとデータを増やして実験を行う必要があるだろう。

謝辞

音声データベースおよび実験結果を快く提供してくださった鹿野さん、中村さん、花沢さん、研究の機会を与えてくださった樽松社長、他ATRの皆さんに心から感謝します。

参考文献

今村
疑似連続分布HMMによる不特定話者電話音声認識
電子情報通信学会技術研究報告SP89-21

古井
マルチテンプレートと教師なし話者適応化による音声認識
電子情報通信学会技術研究報告SP89-17

花沢、川端、鹿野
Hidden Markovモデルを用いた日本語有声破裂音の識別
Research Activities of the Speech Processing Department SP87-97

花沢、川端、鹿野
HMM音韻認識におけるモデル学習の諸検討
Research Activities of the Speech Processing Department SP88-22

花沢、川端、鹿野
HMM音韻認識におけるセパレートベクトル量子化の検討
Research Activities of the Speech Processing Department 2-P-21

中村、鹿野
ファジイベクトル量子化を用いたスペクトログラムの正規化の検討
Research Activities of the Speech Processing Department SP87-123

中川
確率モデルによる音声認識
電子情報通信学会

武田ら
研究用日本語音声データベース利用解説書
ATR Technicak Report

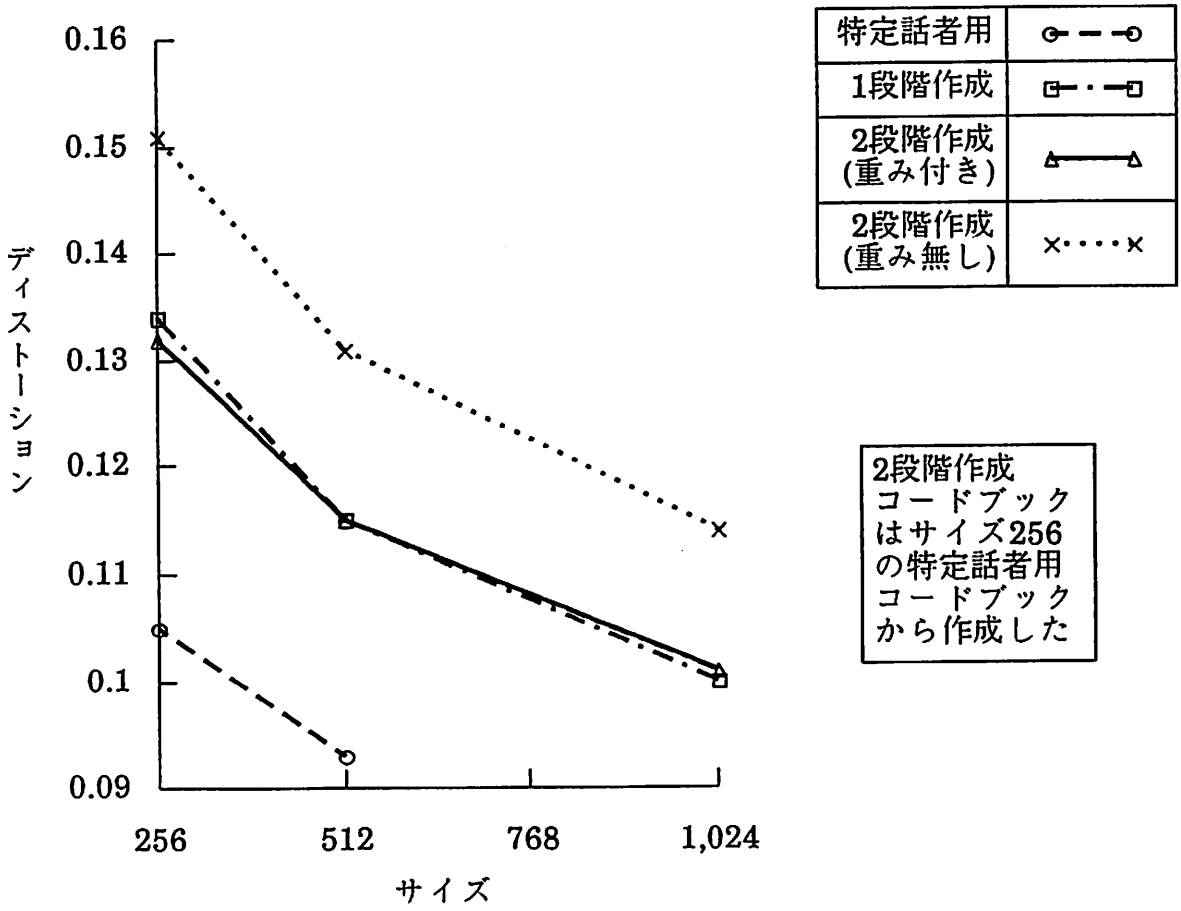

図1　WLRコードブックのディストーション

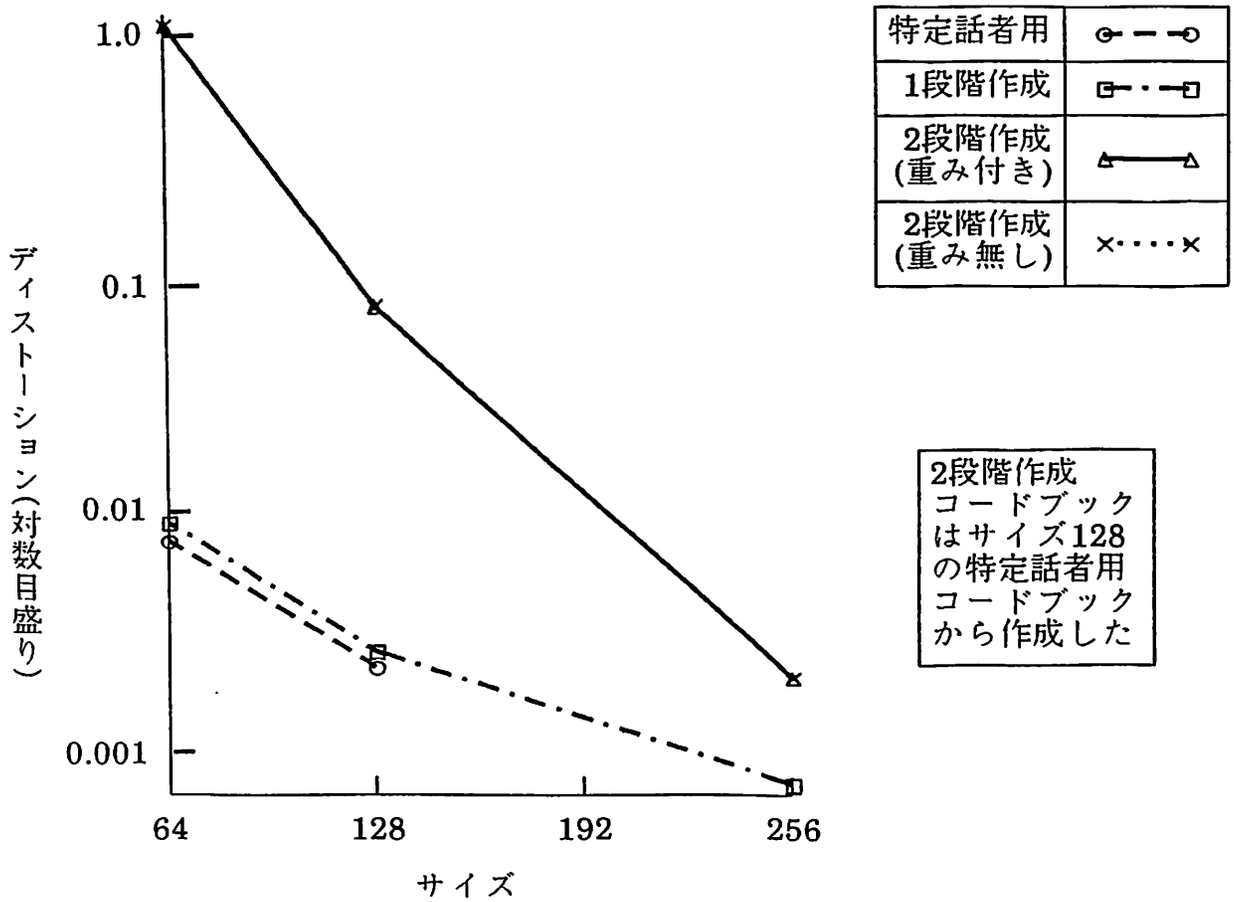

図2　powコードブックのディストーション

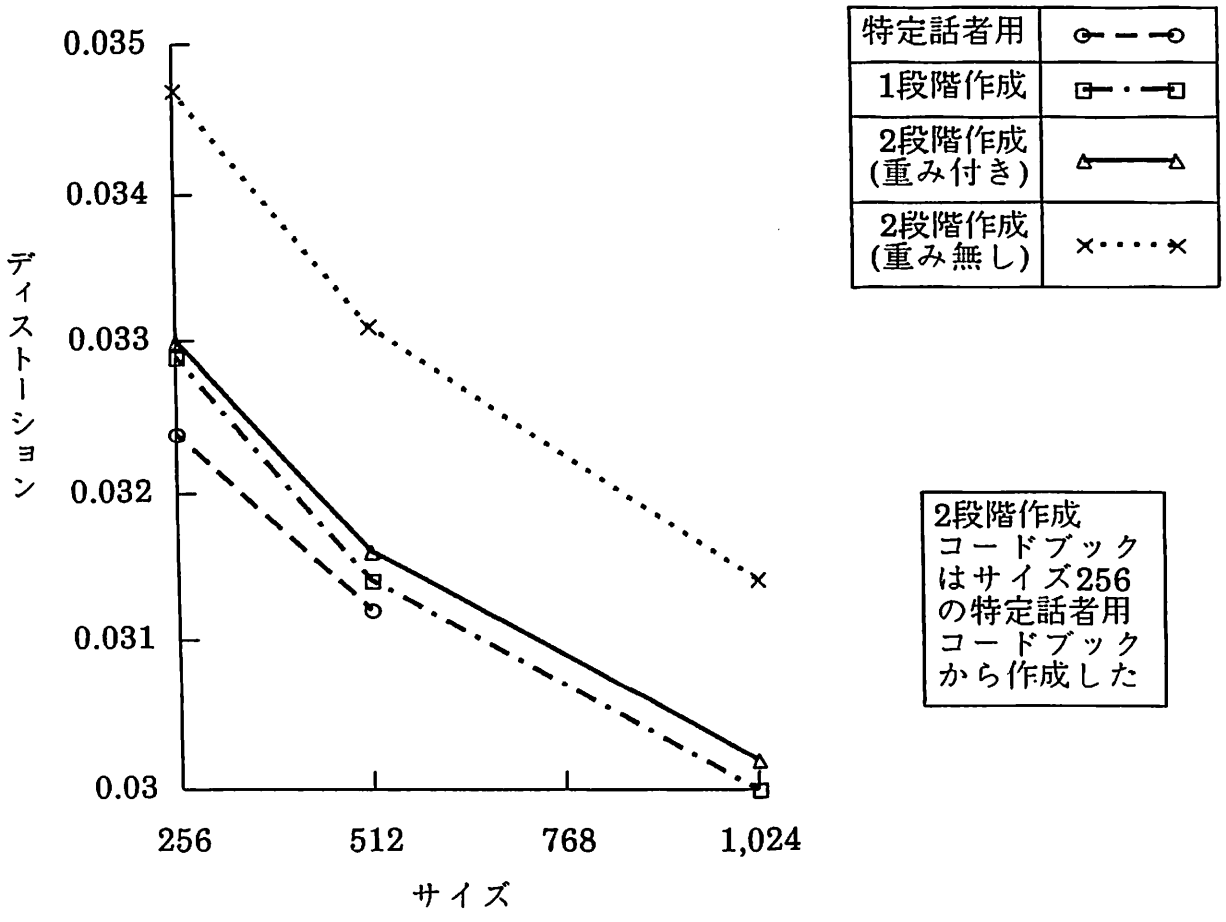

図3　dcepコードブックのディストーション