

論文 / 著書情報
Article / Book Information

論題(和文)	スペクトル空間のクラスタ化に基づく教師なし話者適応化法
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1988年春季講演論文集, , , pp. 83-84
Citation(English)	, , , pp. 83-84
発行日 / Pub. date	1988, 3

古井 貞照 (NTT ヒューマンインタフェース研究所)

1. まえがき

新しい話者が発声した少数の音声サンプルを用いて、認識系をその話者に自動的に適応させることは、不特定話者の大語彙や連続音声を対象とした音声認識において不可欠であると言える。このような目的のため、最近、ベクトル量子化の符号帳を話者に適応化する方法が提案されているが[1][2]、これらの方法は適応化に用いる学習音声の発声内容が既知であることが前提になっている。この前提を取り除いて、任意の音声を用いて適応化することができれば、その用途は大きく広がると思われる。

ここではこのような観点から、白木らが提案しているセグメント符号化における話者適応化法[3]を、不特定話者を対象とした単語音声認識系に適用し、その有効性を確認した結果を報告する。なおこのような方法では対応

できない個人差として、時間構造の個人差があるが、ここではこれに対してはマルチテンプレートとDPマッチングによって対処することとした。

2. 教師なし話者適応化法

図1に、ここで検討した話者適応化法と、それを含むSPLIT法[4]に基づく単語音声認識系のブロック図を示す。各単語を播音韻標準パターン符号の複数の系列で表現した単語辞書は話者に共通とし、播音韻標準パターンを各話者に適応化する。各話者の学習音声としては、複数の単語あるいは短い文章を発声した音声を用いるが、未知入力音声をそのままプールのしておいて用いることもできる。この適応化法のポイントは、学習音声スペクトルと標準パターンを階層的にクラスタ化し、対応する各クラ

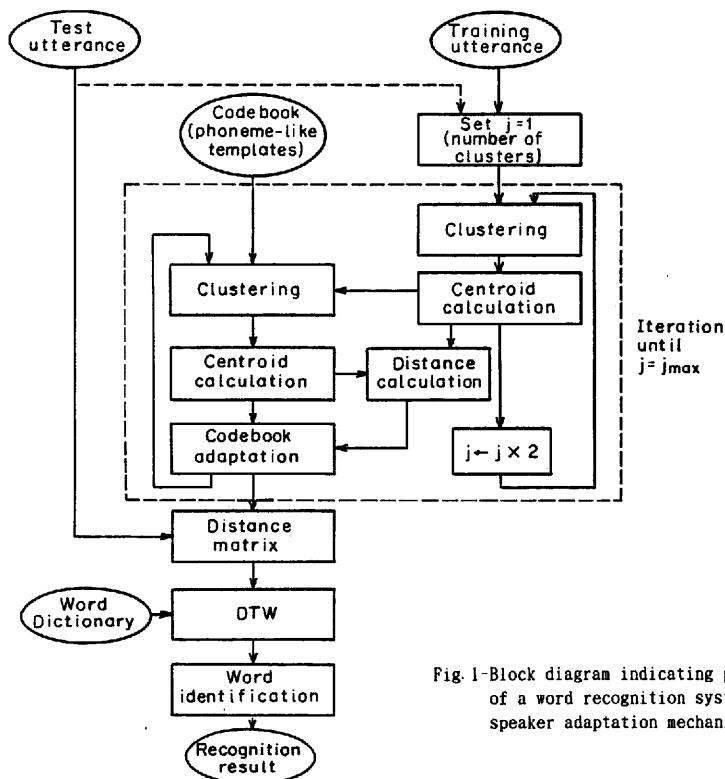


Fig. 1-Block diagram indicating principal structure of a word recognition system including the speaker adaptation mechanism.

* Text-Independent Speaker Adaptation Based on Hierarchical Spectral Clustering.
by Sadaoki Furui (NTT Human Interface Laboratories)

スタ重心間のずれにもとづいて、スペクトルの大局的な話者特性から局所的な話者特性へと徐々に適応化を進めていくところにある。

まずクラスタ数 $j=1$ とし、学習音声全フレームの重心と全標準パターンとの重心のずれを求める。全標準パターンをこのずれに相当するだけ移動させて重心を一致させる。次に $j=2$ とし、学習音声をクラスタ化する。各クラスタの重心を求め、これらの重心を用いて全標準パターンをクラスタ化する。学習音声と標準パターンの対応するクラスタの重心間のずれを求め、これらのずれをもとに、隣接クラスタ間の連続性を保ちながら標準パターンを移動する。この処理を、クラスタ数(j)を2倍に増やしながらか $j=j_{max}$ となるまで繰り返し、標準パターンと学習音声の階層的マッピングを進める。最終状態の標準パターンを用いて、SPLIT法による入力単語音声の認識を行う。

以上の方法を方法Iとする。比較のために方法IIとして、クラスタ化を階層的に行わず直接 j_{max} のクラスタに分割し、標準パターンと学習音声のマッピングを行う方法も検討した。

3. 認識実験

日本語100都市名を認識対象とし、標準化周波数8kHz、フレーム周期16ms、分析次数10の条件で、LPCケプストラムにより分析を行った。特徴パラメータとして、ケプストラムのみを用いる場合(Cep)と、これにケプストラム及び対数パワーの短区間ごとの回帰係数[5]を組み合わせる場合(Cep & Δ Cep)の2種類の実験を行った。男性の話者間変動を代表する4名の音声から、1024の擬音韻標準パターンを作成し、この4名の各単語音声を擬音韻標準パターンの系列(マルチテンプレート)として蓄えた。未知音声として、この4名とは異なる男性20名の音声を用い、学習音声には、各話者が別に発声した10種類(無

作為)の単語音声を用いた。

方法Iにおいてクラスタ数を増やしながらか適応化を行ったときと、方法IIにおいてクラスタ数を種々に選択して適応化を行ったときの、単語標準パターンに対する未知音声のスペクトル歪の推移を図2に示す。学習単語セットを2種類に変えた場合の、20名の100単語に関する平均値を示した。方法Iの場合は適応化の進行とともに歪が低下している。方法IIの場合はクラスタ数に最適値があり、この条件をうまく選ばないと歪が大きくなってしまふ。階層的方法が歪を確実に低下させる上で有効であることがわかる。

図3には、適応化なし、方法Iを用いて32クラスタまで適応化、128クラスタまで適応化、話者毎に同じ100単語を別に発声した音声から各単語辞書を作成、の4種類の場合の認識誤り率を示した。適応化を行う場合については、学習単語セットを上述と同じ2種類に変えた時の平均結果を示した。ここで提案した適応化法が、誤り率を低下させる上で極めて有効性が高いことがわかる。

4. むすび

発声内容を限定しない、音声認識のための新しい話者適応化法を提案し、その有効性を示した。この方法は原理的に、SPLIT法による単語音声認識だけでなく、種々の音声認識系に広く適用することが可能である。

参考文献

- [1] K. Shikano, K-F. Li and R. Reddy: Proc. ICASSP 86, Tokyo, 49. 5 (1986)
- [2] 新美、小林: 音学講義, 2-5-13 (昭62)
- [3] 白木、菅田: 音声研資, SP87-67 (昭62)
- [4] 菅村、古井: 信学論, J65-D. 8, pp. 1041-1048 (昭57)
- [5] S. Furui: IEEE Trans. ASSP-34, 1, pp. 52-59 (1986)

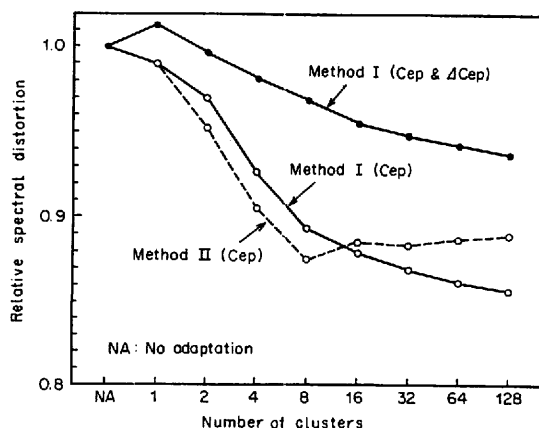


Fig. 2-Trend of spectral distortion between input speech and word reference templates with the advance of speaker adaptation.

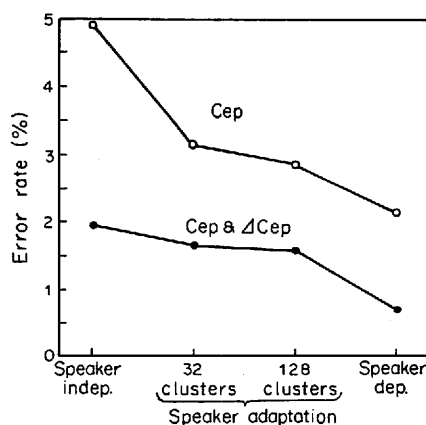


Fig. 3-Recognition error rates under four experimental conditions.