論文 / 著書情報
Article / Book Information

論題(和文)	音声認識のための入力環境の適応化
Title(English)	
著者(和文)	高木啓三郎, 篠田浩一, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成5年度春季研究発表会講演論文集, Vol. 1-4-22, No. , pp.
Citation(English)	, Vol. 1-4-22, No. , pp.
発行日 / Pub. date	1993,

1—4—22

音声認識のための入力環境の適応化 *

⊙ 高木 啓三郎　篠田 浩一　渡辺 隆夫
（ＮＥＣ　C&C情報研究所）

1. はじめに

音声認識の実用化に際し、標準パタン作成時のマイクを含めた入力環境と認識対象音声の入力環境との違いにより音声の認識性能が大幅に低下することが問題となる。

このような認識率低下の要因として、付加的な雑音による騒音レベルの違いとマイク固有の特性の違いによる非線形なスペクトルの変形が考えられる。前者に対してはスペクトルサブトラクション(SS)法[5]が有効であることが知られている。本稿では、これに加え後者に対して筆者等の提案したタスク適応方式[1]を適用し、少数話者の発声を用いて不特定話者HMMの入力環境への適応化を行なう方式を提案する。

2. 入力環境適応化方式

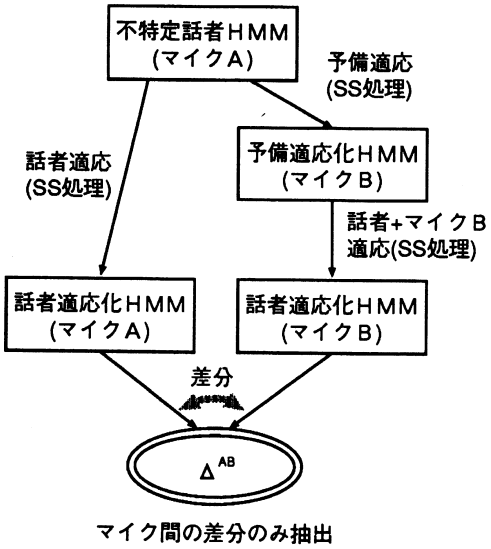

図１マイク間の差ベクトルの抽出

提案する入力環境適応化方式は、異種のマイク間の差ベクトル抽出と、抽出した差ベクトルを用いた標準パタンの適応化の２つの処理から構成される。

図１にマイク間の差ベクトル抽出方式を示す。図において、元となる不特定話者HMMは多数の話者がマイクＡを用いて発声した音声から学習されており、このHMMに対して少数話者が発声したマイクＡの音声およびマイクＢの音声を用いてそれぞれ別個に話者適応化HMMを作成する。

まず、マイクＡの発声に対しては、１入力SS法を用いて雑音除去を行なった後にスペクトル内挿写像法[4]を用いて話者適応を行ない、マイクＡによる話者適応化HMMを作成する。ここで用いたスペクトル内挿写像法は、語彙が既知であるような少数の発声から、連続HMMの各分布毎に平均ベクトルμの適応化ベクトルを抽出し高精度に話者適応化を行なう方式である。一方、マイクＢの発声に対する話者適応を行う場合には、マイクＡの場合とは異なり、直接話者適応を行なわず、一旦マイクＢにある程度適応化した不特定話者HMM（予備適応化HMMとする）を作成し、この予備適応化HMMを介してマイクＢによる話者適応化HMMを作成する。この理由は、元にしたマイクＡによる不特定話者HMMは、マイクＡの音声に比べてマイクＢの音声に対しては認識精度があまり高くないため話者適応時の時間対応付けの精度が低下する場合があり、このことにより差ベクトル抽出の精度が低下するのを防ぐためである。予備適応化HMMの作成は、まず、２つのマイクで同時に収録した発声を用いて、両チャンネルの同じ時間位置の特徴ベクトル間の差ベクトルの平均値を求め、その差ベクトルの平均値を用いて対象となるHMMの各分布の平均ベクトルを共通に適応化することで行なっている。この、予備適応化HMMの作成のための音声および話者適応化のための音声に対してもマイクＡと同様、

* Input Environment Adaptation for Speech Recognition, by Keizaburo Takagi, Koichi Shinoda, Takao Watanabe, NEC Corporation

ＳＳ法により付加的な雑音は取り除いている。 マイク間の差ベクトルΔ^{AB}は、上述のようにして作成した2つのマイク発声の話者適応化HMM間の差分を各分布の平均ベクトル毎に求めることにより行なっている。

標準パタンの適応化は、先に求めたマイク間の差ベクトルΔ^{AB}を元となる不特定話者HMMの各分布の平均ベクトル毎に加算することにより行なっている。

3. 評価実験

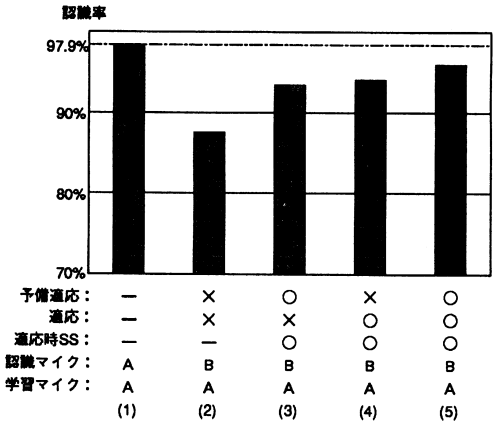

図 2 認識結果

実験は、半音節HMMによる不特定話者音声認識[2][3]において、既に85名の発声から学習済みの不特定話者HMMを、2名の話者が学習とは異なるマイクを用いて発声した音声を用いて入力環境の適応化を行ない、適応時とは異なる話者の発声により評価した。

マイクＡとしてダイナミック型ボーカルマイク、マイクＢとして机上設置タイプコンデンサ型バウンダリーマイクを用い、男女各1名が発声した音素バランス250単語を2つのマイクで同時に収録した音声を用いて適応化を行なった。評価は、適応時とは異なる6名(男4、女2)の話者が発声した別の音素バランス250単語を用いて行なった。なお、評価用発声に対しても比較実験のため2種のマイクで同時に収録しておいた。特徴量は、サンプリング周波数16KHz、帯域制限0.1～7.2KHz、分析窓幅512点、フレーム周期16msで分析を行ない、Δパワー1、メルケプストラム10、Δメルケプストラム10の合計21次元に変換したものを用いた。認識結果を図2に示す。図2(1)は、元にしたマイクＡによる不特定話者HMMによるマイクＡの発声の場合であり97.9%の認識率が得られたが、(2)に示すように同じHMMを用いてマイクＢの発声を認識した場合、87.5%まで低下した。図2(3)は、マイクＢによる予備適応直後のHMMを用いてマイクＢの発声を認識した場合であり、93.3%まで改善され予備適応単独でも効果が見られた。

図2(4)は、予備適応を介さず直接マイクＢへの話者適応を行なった場合、(5)は提案する方式であり、ぞれぞれ93.9%、95.3%となり提案手法が最も高い認識率が得られた。

4. おわりに

スペクトルサブトラクションを用いて雑音の除去を行い、タスク適応方式をもとに少数の話者の発声を用いて不特定話者標準パタンの適応化を行なうことにより入力環境の適応化を行なう方式について提案した。

評価実験から、提案する方式の有効性が示された。同一マイクによる学習時の性能よりはまだ若干低いが、今後検討、改善を行なってゆく予定である。

謝辞

日頃から熱心に御討論いただくメディア研の諸氏に感謝致します。

参考文献

[1] 篠田、渡辺、音響講論、1-P-15(1992.3).

[2] 渡辺他、信学論文誌、J72-D-II、8(1989.8).

[3] 磯谷他、音響講論、3-5-13(1991.3).

[4] K.Shinoda et al, ICASSP91, S13.7(1991).

[5] 高木、吉田、渡辺, 音響講論, 2-5-3(1991.3).