

論文 / 著書情報
Article / Book Information

論題(和文)	電話音声認識のための音響モデルの回線特性への適応化
Title(English)	
著者(和文)	松岡 達雄, Pierre-Emmanuel Gros, 南 泰浩, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1994年秋季講演論文集, , , pp. 103-104
Citation(English)	, , , pp. 103-104
発行日 / Pub. date	1994, 10

○松岡達雄、Pierre-Emmanuel Gros[†]、南泰浩、古井貞照(NTT ヒューマンインタフェース研究所、[†]Eurecom Institute)

1. まえがき

本報告では防音室等でマイクにより収録された比較的雑音やひずみの少ない音声を用いて学習したHMM音素モデルを電話音声に適応化する2つの教師なし適応化法を提案する。一つは、「全音素HMM」によりTied-mixture HMMのコードブックを適応化する方法、もう一つは、コードブックを適応化する際にコードブックをパワーでクラスタリングし、クラスごとに適応化ベクトルを求める方法である。前者では、コードブックの適応化によりパラメータ空間の変換が行なえること、後者では音声パワーの大小に応じて適応化ベクトルを求めることで間接的にSN比に応じた適応化を行なえることが期待できる。提案法を、実際の公衆回線を介して収録した電話音声に適用し、ケプストラム平均値正規化法と比較、評価を行ない、有効性を確認した。

2. 回線特性適応化

防音室などでマイク収録した音声に対して高い認識性能を示す認識システムも、電話音声に適用すると大幅に認識精度が低下することが知られており、これまでも、これを改善しようとする試みが数多く報告されている⁽¹⁾⁻⁽³⁾。認識精度低下の原因としては、背景雑音、回線雑音など付加的な雑音によるものと、収録系の周波数特性の違いなどフィルタ的なひずみによるものが考えられる(図1)。付加的な雑音に対しては、雑音を含まない音声で学習された音素HMMと実環境での雑音のHMMを作成し、この2つのHMMを合成することにより雑音の付加された音声に対する認識精度を改善する方法が提案されている⁽⁴⁾。本報告ではおもにフィルタ的なひずみに対処する方法について検討する。

2.1 ケプストラム平均値正規化

フィルタ的なひずみに対して簡単で効果のある方法としてケプストラム平均値正規化が知られて

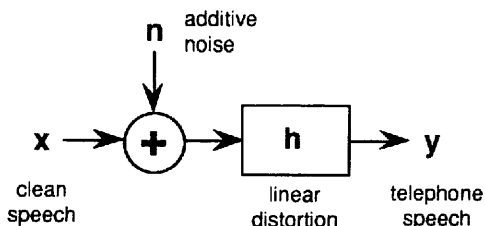


図1 電話音声の品質劣化のモデル

いる⁽¹⁾⁻⁽³⁾。スペクトラム領域でのフィルタは、ケプストラム領域ではケプストラム係数の加減算により実現できるため、ある程度長い音声区間にわたってケプストラム係数の平均値を差し引けばスペクトルを平坦化することができる。学習音声、認識対象音声の両方のスペクトルを平坦化することによりフィルタ的なひずみの影響は除去できると考えられる。しかし、単純にある区間のケプストラム係数の平均値を差し引くだけであるため、学習時と認識時のSN比の違いの影響などが考慮されない。

2.2 全音素HMMによるコードブック適応化

本方法は、Tied-mixture HMMのコードブックをマイク収録された音声のパラメータ空間から電話音声のパラメータ空間へ変換するものである。電話音声を用いてコードブックを再推定することによりパラメータ空間を変換する。

まず、音素の区別なく音声そのものに対して尤度を最大化したHMM（全音素HMMと呼ぶ）をマイク音声を用いて作成する。全音素HMMは1状態のTied-mixture HMMで、コードブックは既学習の音素HMMのものを用い、その平均値、分散は変更しない。次に、この全音素HMMのパラメータを電話音声を用いて再推定する。この時、コードブックの平均値、分散も更新することによりコードブックを電話音声へ適応化する。そして、このコードブックを、マイク音声で既学習の音素HMMのコードブックとして用いれば、音素HMMも電話音声に適応化されると考えられる。このようにして、教師なしで、すなわち発声内容を知らずにコードブックをマイク音声から電話音声へ適応化することができる。

2.3 パワークラスタリング適応化

本方法は、コードブックをパワーでクラスタリングし、各クラスタのパワーに応じてコードワードの修正量を決定する方法である。前述の全音素モデルを用いたコードブックの適応化で求められた適応化ベクトルを各クラスタに属するコードワードについて平均化し、そのクラスタの適応化ベクトルとする。音声パワーの大小によりSN比が異なると考えられるので、パワーに応じて修正量を決定することで間接的にSN比に応じた適応化を行なえると考えられる。

3. 実験

日本音響学会連続音声データベースで学習した音素HMMを、(1)ケプストラム平均値正規化、(2)全音素HMMによるコードブックの適応化、(3)パワー

* Channel adaptation of acoustic models for telephone speech recognition, Tatsuo Matsuoka, Pierre-Emmanuel Gros, Yasuhiro Minami and Sadaoki Furui (NTT Human Interface Laboratories, Eurecom Institute)

クラスタリング適応化、により実際の公衆回線を介した電話音声に適応化し、連続音声の音素認識実験により評価した。電話音声の発声内容は、日本音響学会連続音声データベースと同様の503文章からのサブセットである。男性110名の音声を用いて評価した。各話者15文章づつ発声しており、このうち5文章を適応化に用いた。音素認識実験では、電話音声の音素ラベリングされていないため、発声内容の書き下しを与えて音声区間に対してViterbiアライメントを取り、それを正解の音素区間と仮定し、その音素区間で全ての音素モデルのうち最大尤度を示すものを認識結果とした。音素HMMは音素環境独立の43種類である。音声はマイク収録の連続音声データベース、電話音声とも8kHzでサンプリングした。したがって、信号帯域の違いによる影響はない。

3.1 全音素HMMによるコードブック適応化

図2に全音素HMMによりコードブックを適応化した場合の電話音声に対する音素認識結果を示す。音響的特徴量はケプストラムと Δ ケプストラム各12次元である。図中、CMNはケプストラム平均値正規化、meanは各コードワードの平均値だけを適応化したもの、mean and var.は平均値と分散を同時に適応化したもの、mean+var.は平均値だけを適応化した後に、分散だけを適応化したものである。CMNに提案法を組み合わせた平均値だけを適応化した場合でもCMNに対して改善が見られ、CMNと平均値を適応化してから分散を適応化する方法を組み合わせると55.4%まで認識率が向上した。平均値と分散を同時に適応化する方法は、再推定を一回しか行っていないので分散の推定に適応化前の平均値が用いられてしまうため分散の適応化の効果が上がらないと考えられる。

3.2 パワークラスタリング適応化

パワーでコードブックをクラスタリングし、クラスごとに適応化ベクトルを求める方法を用いた場合の認識結果を図3に示す。音響的特徴量と

して、ケプストラム、 Δ ケプストラムに加え、正規化対数パワーとその一次微分(Δ パワー)を用いた。クラスタリングは正規化対数パワーにより行なった。クラス数は実験的に最適値を求め、5とした。特徴量が増えたことにより先の実験より全体的に認識率が向上しているが、パワーでクラスタリングした場合はCMNやmean(全音素HMMでコードブックの平均値を適応化した場合)より高い認識率を示している。本方法もCMNと組み合わせ、分散も適応化することによりさらに認識率の向上が可能と考えられる。

4. まとめ

音響モデルを回線特性の異なる収録系へ教師なしで適応化する方法として、(1)「全音素HMM」によりTied-mixture HMMのコードブックを適応化する方法、(2)コードブックをパワーに応じてクラスタリングしクラスごとに適応化ベクトルを求める方法を提案した。公衆回線を介して収録した電話音声により評価し有効性を確認した。

パワークラスタリング適応化では、クラスタの適応化ベクトルと、コードワード自体の適応化ベクトルの荷重平均を新たに適応化ベクトルとし、その重みを音声パワーに応じて調整することで、どちらかをより重視するという詳細な適応化も可能である。重みの調整法について検討中である。

今後は、提案法をHMMの合成により付加的な雑音に対応する方法⁽⁶⁾と組み合わせることにより、さらに認識精度の向上を図る予定である。

参考文献

1. F.-H.Liu, R.M.Stern, A.Acero, and P.J.Moreno, Proc. ICASSP-94, pp.II-61-64
2. M. Weintraub and L. Neumeyer, Proc. ICASSP-94, pp. I-85-88
3. P.J.Moreno and R.M.Stern, Proc. ICASSP-94, pp.I-109-112
4. L. Neumeyer and M. Weintraub, Proc. ICASSP-94, pp.I-417-420
5. M.G.Rahim and B.H.Junag, Proc. ICASSP-94, pp.I-445-448
6. F.Martin, K.Shikano, and Y.Minami, Proc.Eurospeech-93, pp.1031-1034
7. T.Matsui and S.Furui, Proc. ICASSP-94, pp.I-125-129

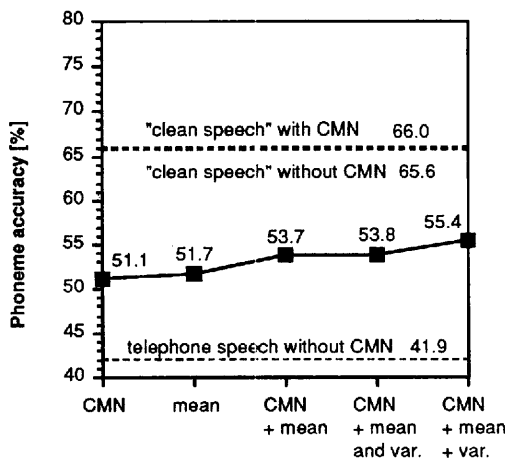


図2 電話音声に対する音素認識結果
(全音素HMMによるコードブック適応化)

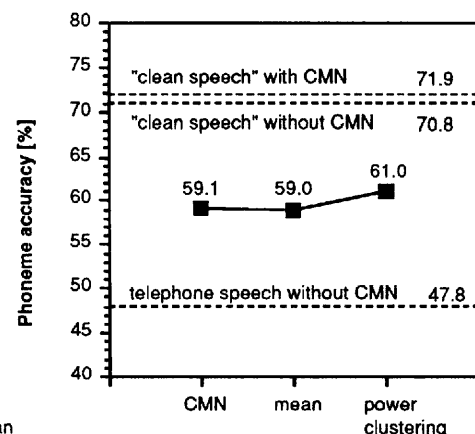


図3 電話音声に対する音素認識結果
(パワークラスタリングによる適応化)
正規化対数パワー、 Δ 正規化対数パワーを含む