

論文 / 著書情報
Article / Book Information

論題(和文)	雑音と歪みを含んだ音声へのHMM適応化手法の評価
Title(English)	
著者(和文)	南泰浩, 高木幸一, 古井貞熙
Authors(English)	Koichi Takagi, SADAOKI FURUI
出典(和文)	日本音響学会 1996年春季講演論文集, Vol. , No. 2-5-14, pp. 85-86
Citation(English)	, Vol. , No. 2-5-14, pp. 85-86
発行日 / Pub. date	1996, 3

○南 泰浩¹ 高木幸一² 古井貞照¹²(NTTヒューマンインタフェース研究所¹、東京工業大学²)

1.はじめに

実環境で音声認識を行う場合、雑音や歪みの影響を受けた音声認識しなければならず、認識率の低下をまねいていた。通常、このような環境の音声信号は、車の騒音やコンピュータ雑音などが加算され、それに加え回線歪みやマイク歪みなどが乗算されたような形の線形系で表現できる。しかし、ケプストラムと線形スペクトルの間には対数変換という非線形な関数があるため、加算性雑音と乗算性歪みとを同時に推定するのは難しかった。筆者らは、加算性の雑音だけを対象にしていたHMM合成法[1][2]を拡張し、合成されたHMMの適応音声に対する尤度が最大になるように乗算性の歪みを推定する方法を提案した(拡張HMM合成法)[3]。本論文ではこのアルゴリズムを用いて数字認識実験を行った結果について報告する。

2. 拡張HMM合成法による適応化

雑音が重畳し、かつ、ひずんだ音声の生成過程を図1のようにモデル化する。音声信号 S は音声HMMによって生成される。また、雑音信号 N は雑

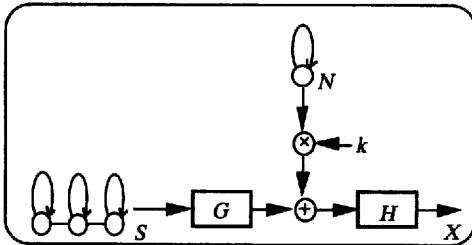
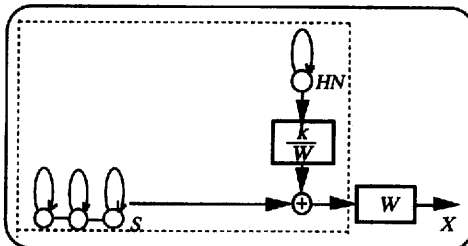


図1 雑音とひずみを含む音声のモデル化

図2 雑音とひずみを含む音声モデルの変形 ($W=HG$)

音HMMによって生成される。初めに S には、発声様式の変化やロンバート効果による発声の歪みなどに対応する乗算性の歪み G が掛けられる。その後、加算性の雑音 N が音声信号 GS に加えられる。このとき音声と雑音のパワー比に対応する係数 k がかけられる。最後に回線歪みやマイクの歪みなどに対応する乗算性歪み H が掛けられる。最終的な信号は $X (= HGS + kHN)$ となる。ここで $W = HG$ と置くと、 X は $X = WS + kHN$ となる。この式を用いて図1のモデルを図2のように書き直す。

S を生成するHMMとしては不特定話者のHMMを用いる。また HN を生成するHMMは環境雑音だけから作成する。図中の+の演算には、HMM合成法を使う。 k, W は固定値であり分散がないとする。雑音と歪みの推定はこの k, W を同時に推定することで実現できる。

合成によって生成される雑音の重畳したHMMのモデルパラメータセットを $M(k, W)$ とする。尤度 $P(O|M(k, W))$ を最大にすることによって k, W を求める。 $O = \{x_1, x_2, \dots, x_T\}$ は入力音声の時系列である。

このアルゴリズムを以下に示す。

- (1) W を初期設定する。
- (2) k を変化させてHMMを合成し、 $P(O|M(k, W))$ が最大になる k を選択する。
- (3) 次に点線の枠内の部分のHMMを、 $1/W$ を固定し、(2)の k を用いて合成する。
- (4) Sankarのケプストラムバイアス推定法を用いて点線枠外の W を推定する[4]。
- (5) $1/W$ を新しい W を用いて更新する。
- (6) (2)から(5)を繰り返す。

Sankarのケプストラムバイアス推定法は入力音声の特徴量 x_t と学習データの特徴量 y_t との間のバイアス b (線形スペクトラム上では乗算性歪み)をEMアルゴリズムにより推定する方法である。

ここでは W のlog変換の逆コサイン変換を b として推定する。

3. 教師なし適応法

本論文では教師なしで適応を行った。この様子を図3に示す。あらかじめ、学習用の数字音声データ全体を使って、1つの全音声HMMを作成する。また、雑音データから1つの雑音HMMを作成する。

*Evaluation of an HMM adaptation method for noisy and distorted speech.

By Yasuhiro Minami¹, Kouichi Takagi² and Sadaoki Furui^{1,2} (NTT Human Interface Laboratories¹, Tokyo Institute of Technology²)

音声が入力されると、全音声モデルと雑音モデルから拡張HMM合成法を使って、発声された音声に対する k, W を推定する。この k, W からHMM合成法によって雑音や歪みを考慮した数字HMMが作られる。この数字HMMを使って、入力された音声を認識して、数字列を出力する。

4. 認識実験

4.1 実験条件

本論文では、4桁数字の認識実験を行った。表1に実験に用いたデータを示す。全音声HMMとして1状態16混合分布のHMMを用いた。数字HMMの状態数は数字によって異なるが、状態毎の分布数は4である。雑音HMMとして1状態1混合分布のHMMを用いた。乗算性の歪みとして、バウンダリマイクロフォンを用いた。加算性雑音としては電子協データベース中の2つの雑音を用いた(表1)。雑音は音声とは別にバウンダリマイクロフォンを使って収録し、計算機上で所望のS/Nになるように音声データに加えた。このとき、S/Nを音声データ全体に対して求め、音声データ全体に対して同じ大きさの雑音を付加した。

表1 実験に用いたデータ

不特定話者数字HMM 学習用データ	電子協 4桁数字9702発声(70人)
評価用データ	4桁数字1785発声(51人)
雑音データ	電子協雑音データベース 電子協雑音データベース 走行自動車内雑音

4.2 実験結果

比較実験に使った方法は以下の通りである。

- (1) 適応なし
- (2) CMN

Cepstrum mean normalization[5]を行った。

- (3) HMM合成

HMM合成だけを行った。合成の際、合成されたHMMの適応用音声に対する尤度が最大になるように k を設定した。

- (4) 拡張HMM合成法

バウンダリマイクロフォンで収録した、雑音を含

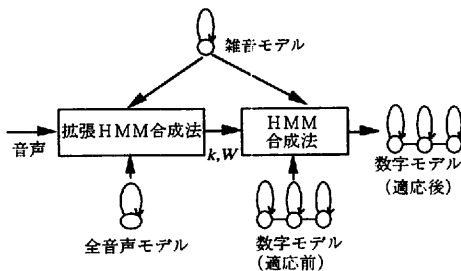


図3 教師なし雑音、歪み適応法

んでいない音声で認識実験を行った場合、数字列認識率は93%であった。このことから、バウンダリマイクロフォンによる認識率の劣化は最大でも7%であることがわかる。数字列認識率の実験結果を表2、表3に示す。CMNは乗算性雑音だけを考慮した方法であるにも関わらず、総ての実験で認識率が7%以上改善しており、加算性雑音の補償にも役立っていることがわかる。HMM合成法を使うと、加算性雑音による認識率の劣化が大きく改善されることがわかる。さらに、拡張HMM合成法はHMM合成法より1~4%程度の認識率の向上を示している。このとき、乗算性雑音だけを考慮しているCMNに比べても高い認識率を示しており、拡張HMM合成法が、加算性雑音、乗算性雑音のどちらにも適応できることがわかる。

5. まとめ

本報告では、加算性雑音と乗算性歪みに適応できる拡張HMM合成法を数字認識実験において評価した。従来法と比較した認識実験結果から、拡張HMM合成法が、加算性雑音と乗算性歪みを含んだ音声に対して有効であることが確認された。

謝辞

日頃、有意義な議論をしてくださる古井特別研究室の方々に感謝致します。本実験では、電子協のデータベースを利用した。

文献

- [1] F. Martin, K. Shikano, and Y. Minami, "Recognition of noisy speech by composition of hidden Markov models", *EUROSPEECH* 93, pp. 1031-1034, (1993 9).
- [2] M. J. F. Gales and S. J. Young, "An improved approach to the hidden Markov model decomposition of speech and noise", *ICASSP* 92, pp. 233-236 (1992 3).
- [3] 南 泰浩, 古井貞照, "最尤推定法を用いたHMM適応化", 音論集, 1-2-1, (1995-9).
- [4] A. Sankar, C.-H. Lee, "Robust speech recognition based on stochastic matching", *ICASSP* 95., pp. 121-124 (1995 5).
- [5] B. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification", *Journal of the Acoustic Society of America*, vol. 55, pp. 1304-1312, (1974, 9).

S/N	適応なし	CMN	HMM合成	拡張HMM合成
12dB	25.6%	33.6%	59.2%	60.3%
18dB	56.9%	69.9%	82.5%	84.6%

表2 計算機雑音中の4桁数字認識実験結果

S/N	適応なし	CMN	HMM合成	拡張HMM合成
12dB	68.4%	82.6%	87.2%	89.0%
18dB	80.8%	94.0%	91.8%	96.0%

表3 走行自動車内雑音中の4桁数字認識実験結果