

論文 / 著書情報
Article / Book Information

論題(和文)	雑音重畳音声に対する各種適応手法の検討
Title(English)	A Comparison of Various Methods for Noise Adaptation
著者(和文)	張志鵬, 伊藤大輔, 古井貞熙
Authors(English)	Daisuke Ito, SADAOKI FURUI
出典(和文)	日本音響学会 2001年春季講演論文集, Vol. , No. 2-3-1, pp. 55-56
Citation(English)	, Vol. , No. 2-3-1, pp. 55-56
発行日 / Pub. date	2001, 3

雑音重畳音声に対する各種適応手法の検討*

◎張 志鵬 伊藤大輔 古井 貞熙 (東工大)

1. はじめに

大語彙連続音声認識における問題の一つとして、背景に雑音や音楽を含む音声に対する認識性能の劣化が挙げられる。我々はニューラルネットワークを用いた雑音適応手法[1]を提案したが、他の適応手法との性能比較を行っていない。そこで本稿では雑音適応の各手法(HMM合成法、MLLR法)との性能比較について論じる。

2. 雑音適応手法

2.1 ニューラルネットワークを用いた手法

HMMの各状態の出力確率分布に対し、ニューラルネットワークを用いて非線形演算を行う。学習時に雑音HMM、音声HMM、及び雑音情報を含むHMM(目標HMM)が必要である。音声HMMは、クリーンな環境の下で学習された多数話者の音声から作成した不特定話者HMMを使用する。雑音HMMは、各雑音重畳音声データから、音声を含まない部分を雑音データとして抽出し、学習して作成する。

2.2 HMM合成法

HMM合成法[2]はよく用いられている手法で、音声のHMMと雑音のHMMの畳み込みをスペクトル空間で行って、雑音が重畳した音声のHMMを合成するものである。従って、HMMパラメータをスペクトル空間に変換する必要があるため、単純にはCMSを用いたHMMを合成法に利用することができない。

2.3 MLLRによる教師なし適応

MLLR[3]はHMMのガウス分布の平均値を尤度最大化の基準に基づいた線形変換により適応化する方法である。話者適応によく用いられる手法であるが、雑音の適応にも有効である。一般に音素に対しクラスターリングを行い、性質の近い音素を同じ変換行列に共有させる方が全音素を1クラスターとしたものより有効である。前述したニューラ

ルネットワーク法とHMM合成法には雑音HMMが必要になるが、MLLRの場合は雑音HMMが要らない。

3. 認識実験

3.1 音響モデル

今回の実験では、音声HMMとしてtree-based clusteringにより状態共有化を行った不特定話者文脈依存音素HMMを用いた。音響特徴量としては16次のLPCケプストラムと対数パワー、及びそれらの一次微分の計34次元を使用した。学習用音声データは、ATR音声データベースBセット、日本音響学会連続音声データベース、および同模擬対話データベースから、男性53名による13,270発話を用いて性別依存モデルを作成した。モデルの総状態数は2,106、各状態のガウス分布の混合数はすべて4である。

3.2 言語モデル

言語モデルの学習に用いたデータは放送ニュース原稿テキスト5年分、約50万文である。単語出現頻度上位2万語を認識語彙とし、間投詞と読みを考慮した言語モデル[4]を用いた。

3.3 評価用データ

1996年7月に実際に放送されたクリーンなニュース音声に、人工的に雑音を付加させる。一人の男性話者による14文を抽出し、うち4文をニューラルネットワーク法における目標HMMの作成用データ、残り10文はすべての手法において評価用データとして使用した。

4. 認識実験結果

4.1 ニューラルネットワークを用いた手法

4.1.1 雑音HMM及び目標HMMの作成

雑音HMMは各雑音データより Baum-Welch アルゴリズムを用いて作成した。また1状態1混合とし音声HMMと同様の音響特徴量を用いた。目標HMMはMLLR-MAP及びVFSの手法を併用し、音声HMMの平均ベクトルを各雑音重畳音声データに適応さ

*A comparison of various methods for noise adaptation By Zhipeng Zhang, Daisuke Ito and Sadaoki Furui (Tokyo Institute of Technology)

せることで作成した。よって音響特徴量，モデルの状態数は音声モデルと同一で，混合分布の重みを修正せずに各雑音に適応したHMMを得ることができる。なお，CMSを適用している。

4.1.2 学習

以上より得られる各種HMMを用いて，音声HMM及び雑音HMMの16次のLPCケプストラムとそれぞれの平均パワーを入力，目標HMMの16次元のLPCケプストラムを出力とするニューラルネットワークを学習した。

4.1.3 結果

二種類の雑音（人ごみと展示場）の雑音条件で適応実験を行った。各SN比における認識結果を図1,2に示す。MLLR法で教師なしのオンライン適応を行った結果も併せて示してある。ほとんどの場合ニューラルネットワーク法の方がMLLR法より性能が良いことが分かる。

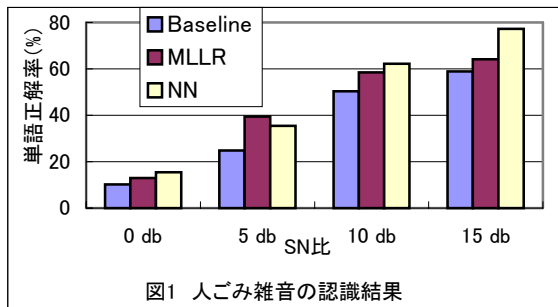


図1 人ごみ雑音の認識結果

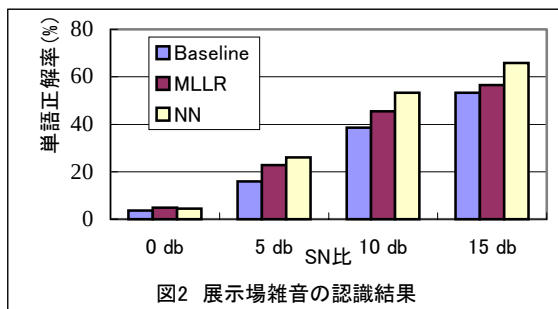


図2 展示場雑音の認識結果

4.2 HMM合成法による結果との比較

図3,4にHMM合成法を用いた各SN比における適応効果を示す。雑音HMMはニューラルネットワーク法と同じ構造である。HMM各状態のガウス分布の平均値だけを適応した。本手法ではCMSを用いていないことから、ベースラインの認識率がニューラルネットワーク法より劣っている。そのため、適応後の結果もニューラルネットワーク法の方が高い認識率を示した。

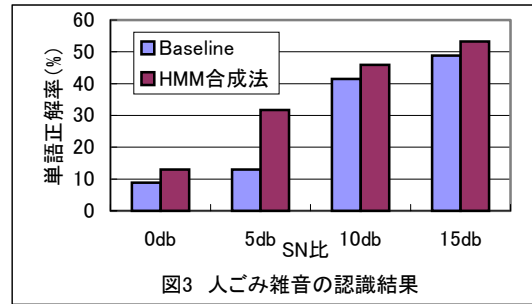


図3 人ごみ雑音の認識結果

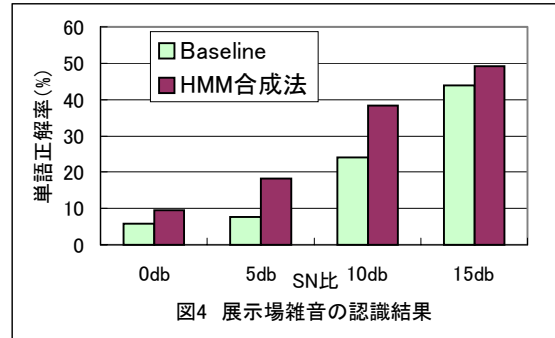


図4 展示場雑音の認識結果

5. まとめ

これまでに提案していたニューラルネットワークを用いた雑音適応の手法についてHMM合成法、MLLR法との性能比較を行った。実験結果からニューラルネットワーク法はほとんどの条件で優れた適応性能を有することがわかった。

なお，HMM合成法ではCMSを適用していないため、ベースラインの性能が劣化している。これに対処するために拡張HMM合成法[5]が提案されているが，計算量が大きくなる問題がある。今後はこのような手法も比較の対象として検討を進めていきたい。

謝辞

ニュース原稿及び音声データを提供して頂いたNHK放送技術研究所に感謝します。ご助言を頂いたNTTサイバースペース研究所の堀貴明、大附克年氏に感謝します。日頃討論頂く東工大の研究室の方々に感謝します。

参考文献

- [1] S.Furui et al., Proc of the ISCA ITRW ASR2000 pp160-167,2000.
- [2] F.Martin et al., 信学技報 SP92-96, 1992.
- [3] C.J.Leggetter et al., Computer Speech and Language, Vol.9, pp.171-185, 1995
- [4] 桜井 他, 春季音講論, pp.57-58, 1999.
- [5] 南 他, 信学論, pp.1179-1186, 1997.