

論文 / 著書情報
Article / Book Information

論題(和文)	区分線形変換による雑音適応の検討
Title(English)	
著者(和文)	張 志鵬, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 2001年秋季講演論文集, Vol. , No. 1-1-15, pp. 29-30
Citation(English)	, Vol. , No. 1-1-15, pp. 29-30
発行日 / Pub. date	2001, 10

区分線形変換による雑音適応の検討*

張 志鵬 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。一般に、音声に雑音を加算されたときに、そのケプストラム空間での効果は非線形なので、これまで、ケプストラム空間での確率分布を表すHMMに対して、雑音に適応するためにHMM合成[1]やneural network[2]などの種々の非線形処理が研究されてきた。しかし、非線形処理には、複雑な処理と大きな計算量を必要とする問題がある。特に非定常雑音に対して各入力文ごとに適応化する場合に、大きな問題となる。そこで、非線形処理を区分線形変換(piecewise linear transformation)で近似する一つの方法を提案する。

2. 区分線形変換による雑音適応手法

HMMパラメータ空間(雑音が重畳した音声のHMM空間)を区分化し、入力音声の条件に最も適合した部分空間を選ぶ。選ばれた空間で、尤度がさらに最大化するように線形変換(MLLR)を行う。

2.1 雑音クラスタリング

多数の雑音とSNRの条件に対し、HMMパラメータ空間を区分化(クラスタ化)し、各クラスタについて音響モデル(クラスタ雑音重畳HMM)を作成する。入力雑音重畳音声に一番近いクラスタ雑音重畳HMMを選択し、認識に用いる。

多数種類の雑音を直接区分化するのは困難なので、各雑音データから作成された雑音GMMにより区分化を行う。雑音GMMに対し尤度を計算し、雑音間の尤度行列に基づいて、SPLIT法[3]で用いられたクラスタリングアルゴリズムを適用した。

クラスタ雑音重畳HMMはクリーン音声モデルを初期モデルとして、クラスタに属する雑音を重畳した音声データを用いた連結学習により作成する。

2.2 線形変換による雑音適応手法

MLLR[4]変換手法は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。本研究では、各クラスタ雑音重

畳モデルを、MLLRを用いて適応化する。

3. 認識実験

3.1 音響モデル

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を用いる。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。学習用クリーン音声データは、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中の、男性53名による13,270発話である。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、間投詞を考慮した言語モデル[5]を用いた。

3.3 雑音データ

雑音データは電子協雑音データベースの28種類の雑音を用いた。Baum-Welchアルゴリズムを用いて64混合の各雑音GMMを学習した。

3.4 評価用データ

2種類の評価用データを用いた。まず、1996年7月に実際に放送された一人の男性話者による10文のクリーンなニュース音声に、3種類のSNR(SNR=0, 10, 15dB)で、学習に用いなかった2種類の雑音(人ごみと展示場)計6種類の組合せを重畳させたデータを用意した(test1)。また、1996年7月に実際に放送されたニュース音声から、背景に多種の雑音や音楽が乗っている発話や記者レポートなどの発話50文(test2、平均SNR=17dB)を使用した。

4. 認識実験結果

4.1 雑音クラスタリングによる効果

test1の6種類の雑音重畳したデータに対し、同じSNRのクラスタ雑音重畳HMMを用いて、クラスタ数を8, 16, 28とした時のクラスタ選択による適応実験を

*A piecewise linear transformation method for noise adaptation

行った。各SNRにおいて認識正解精度を図1,2に示す。全体として16クラスタ程度が最適であることが分かる。

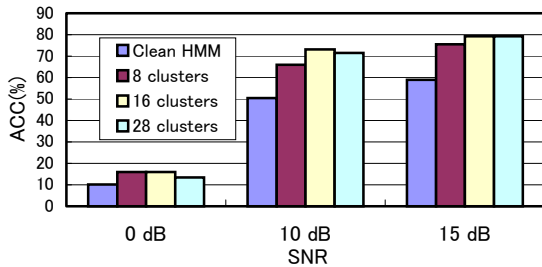


図1. 人ごみ雑音の認識結果

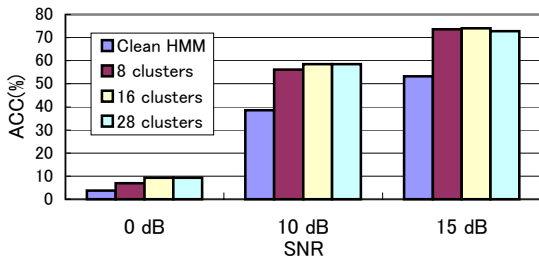


図2. 展示雑音の認識結果

またtest2に対し各SNRのクラスタ雑音重畳HMMを用いて適応実験を行った。各SNRにおいて認識正解精度を図3に示す。SNRによって結果が異なるが、8-16クラスタが一般的により結果を示している。

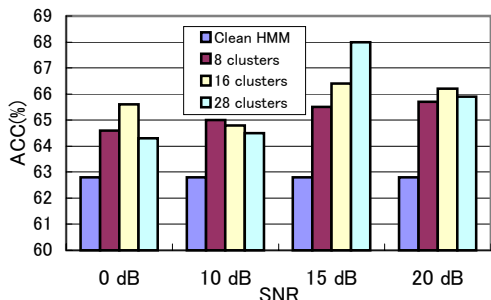


図3. test2雑音音声の認識結果

4.2 区分線形変換による適応結果

0, 10, 15, 20dBのSNRにおけるあらゆる雑音音声モデルの中から最適なクラスタ雑音重畳HMMを選択し、さらに線形変換手法を用いて適応を行う。多数のモデルから尤度最大のHMMを選択するのは計算量が膨大になるため、クラスタ雑音重畳HMMの代わりにクラスタ雑音重畳GMMを用いて選択する手法を用いた。クラスタ雑音重畳GMMはクラスタ雑音重畳HMMの学習に使われるデータと同じものを用いて作成した。

入力音声に対し、尤度最大を示すクラスタ雑音重畳GMMに対応するHMMを選択し、さらにモデルの線形変換を行った。図4,5にtest1の雑音重畳音声に対し適応を行った効果を示す。

またtest2に対し適応実験を行った。結果を図6

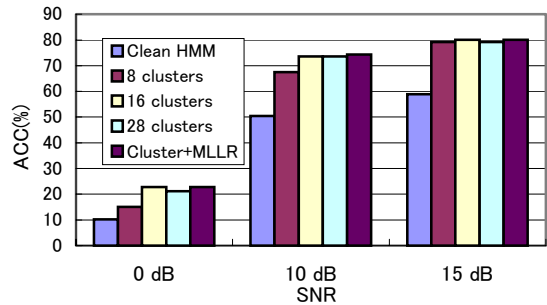


図4. 人ごみ雑音の区分線形法の結果

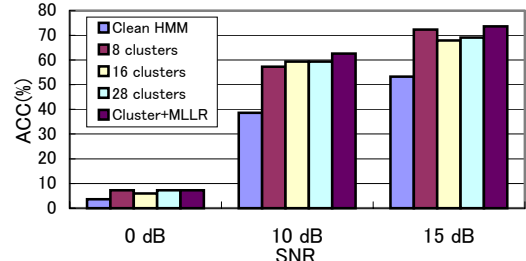


図5. 展示雑音の区分線形法の結果

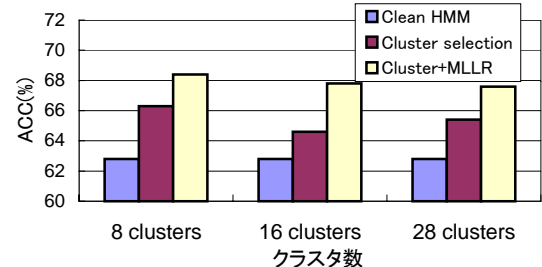


図6. test2雑音音声の区分線形法の認識結果

に示す。これらの結果から、音声に多種類の雑音があり、雑音が変動しているような環境の場合での本手法の有効性が確認された。

5. まとめ

雑音とSNRの組合せについて、クラスタ選択を行ってから、各クラスタごとに線形変換手法を用いる雑音適応法を提案した。クラスタの選択には計算量を考慮し、GMMを用いた。評価実験によって提案手法の有効性を確認した。ここで得られた結果は、HMM合成法やneural network法による結果よりも認識率の向上において優れていることが確認されている。

謝辞

この研究はNTTドコモ株式会社の研究委託を受けて行われました。ここに深く感謝いたします。

参考文献

- [1] F.Martin et al., 信学技報 SP92-96, 1992.
- [2] 張 他, 春季音講論, pp.55-56, 2001.
- [3] 菅村 他, 音声研究会資料, S82-64,1982.
- [4] C.J.Leggetter et al., Computer Speech and Language, Vol.9, pp.171-185, 1995.
- [5] 桜井 他, 春季音講論, pp.57-58, 1999.