T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	区分線形変換による尤度最大化雑音適応法の検討
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 2002年春季講演論文集, Vol. , No. 2-2-3, pp. 61-62
Citation(English)	, Vol. , No. 2-2-3, pp. 61-62
発行日 / Pub. date	2002, 3

Powered by T2R2 (Science Tokyo Research Repository)

区分線形変換による尤度最大化雑音適応法の検討*

◎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。これまでに我々は区分線形変換(PLT; piecewise linear transformation)による雑音適応法[1]を提案した。この論文では区分線形変換における分散適応の効果について考察する。

2.尤度最大化規準に基づく区分線形変換雑音適応

2.1尤度最大化規準による雑音適応法

一般に、音声に雑音が加算されたときに、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[2]やneural network[3]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。

尤度最大化規準が音声認識の各方面、例えばデコーディング、話者適応などによく用いられる。雑音適応に関しても、特に雑音が時間的に変化する場合に尤度最大化によるモデル適応が有効と考えられる。その一つの方法として、拡張HMM合成法[4]が提案された。しかしながら、この手法では大きな計算量を必要とする問題がある。計算量を減らすために、特徴パラメータ領域で、雑音の影響を尤度最大化規準に基づいて推定し、入力音声から除去する手法[5]が提案された。しかしながら非線形な効果に対して、入力信号からバイアスを引くだけでは不十分であるため、この手法には限界がある。そこで本論文では、非線形処理を区分線形変換で近似して,モデルの尤度最大化をはかる一つの方法を提案する。なお、すでに音素クラスタ別にMLLR変換を適用する雑音適応手法[6]が提案されているが、ここで提案する手法は、雑音の特性まで含めて区分化するところに特徴がある。

2.2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。分散をも考慮する場合[7]は変換後の共分散行列は以下のように計算する。

$$\hat{\Sigma} = LHL^T \qquad (1)$$

Hは変換行列で、Lは変換前の共分散行列のCholeski分解である。本研究では、平均だけと平均分散両方を適応する手法について検討した。

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,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[8]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

これまで我々はHMMモデル各分布の平均値だけ[1]を線形変換してきた。また、あらゆる音素を同じ変換行列によって変換した。今回は、HMMモデルの音素をクラスタに分けて線形変換を行った。さらに、HMMモデル各分布の平均と分散の両方を適応する実験を行った。図1,2,3,4に雑音クラスタ数16の場合のTest1雑音重畳音声に対し適応を行った効果を示す。雑音の種類によって分散の適応効果がある場合とない場合があることが分かる。効果がある場合でも、その度合いは小さい。音素クラスタの数を増やしても性能は上がらないことが分かる。

---

*Maximum likelihood-based piecewise linear transformation method for noise adaptation

By Zhipeng Zhang and Sadaoki Furui (Tokyo Institute of Technology)

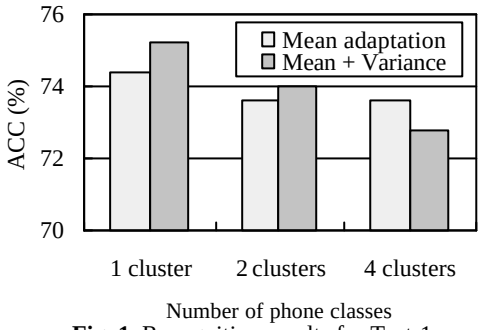

Fig. 1: Recognition results for Test-1
(crowd-noise-added speech, SNR: 10dB).

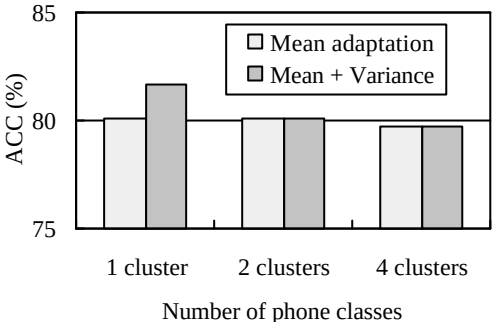

Fig. 2: Recognition results for Test-1
(crowd-noise-added speech, SNR:15dB).

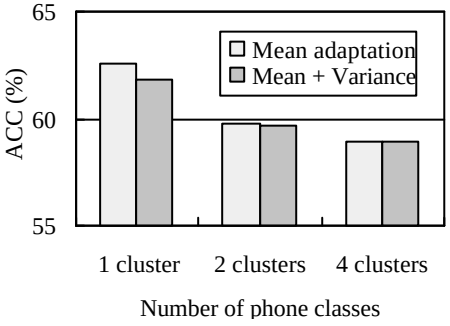

Fig. 3: Recognition results for Test-1
(exhibition hall noise-added speech, SNR: 10dB).

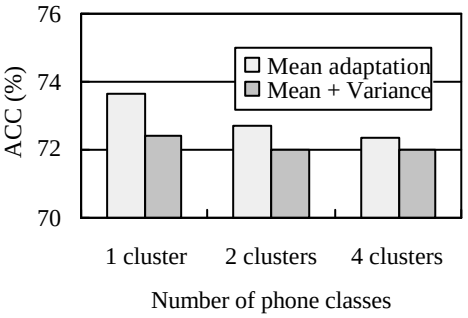

Fig.4: Recognition results for Test-1
(exhibition hall noise-added speech, SNR: 15dB)

次にTest2に対し適応実験を行った。雑音クラスタの数が28で各音素クラスタ数を変えた時、平均だけを適応する場合と、平均と分散両方を適応する場合の結果を図5に示す。1クラスタで平均と分散の両方を適応する場合に最も高い精度が得られた。

最後に3種類の雑音クラスタの条件で平均と分散両方を適応する実験を行った。結果は図6に示す。16

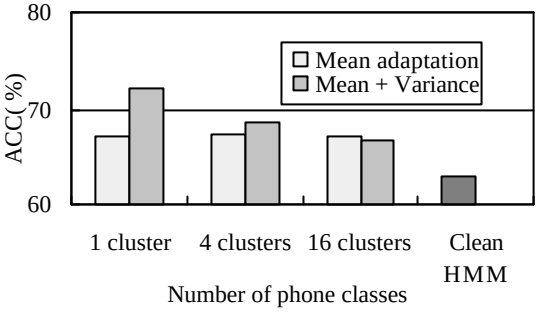

Fig. 5: Recognition results for Test-2
(number of noise clusters: 28).

クラスタの場合に最も高い精度が得られた。人工的に単一雑音を付加した音声より,音声に多種類の雑音が加わり、雑音が変動しているような実環境の場合に分散適応の効果が大きいことが確認された。ベースラインに比べ、単語正解精度は9.5%増加した。

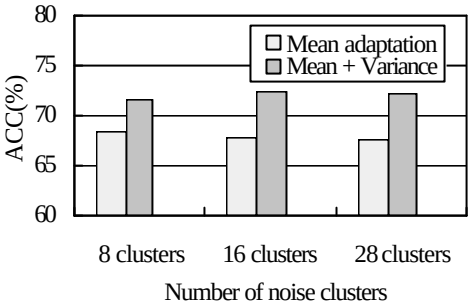

Fig. 6: Recognition results for Test-2
(number of phone clusters: 1).

5. まとめ

尤度最大化規準に基づく区分線形雑音適応法に関し、実環境での雑音重畳音声に対して、分散適応の効果を確認した。音素数を1クラスタにして、平均と分散両方を適応する方法が最も有効であることが分かった。分散適応の効果については、さらに検討を進める必要がある。

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] 張, 古井, 秋季音講論, pp.29-30, 2001.
[2] F.Martin et al., 信学技報 SP92-96, 1992.
[3] 張, 古井, 春季音講論, pp.55-56, 2001.
[4] Y. Minami and S. Furui, *Proc. ICASSP*, pp. 129-132, 1995.
[5] C. Lawrence, et al., *Computer Speech and Language*, vol.13, No.3, pp. 283-298, 1999.
[6] Y. Gong, et al., *Proc. ICASSP*, vol.1, pp. 297-300, 1999.
[7] C.J.Leggetter et al., Computer Speech and Language,Vol.9, pp.171-185, 1995.
[8] 桜井 他, 春季音講論, pp.57-58, 1999.