

論文 / 著書情報
Article / Book Information

論題(和文)	雑音の種類とSNRを同時に考慮した木構造クラスタリングによる雑音適応法
Title(English)	
著者(和文)	張 志鵬, 杉村 利明, 古井 貞熙
Authors(English)	Toshiaki Sugimura, SADAOKI FURUI
出典(和文)	日本音響学会 2003年秋季講演論文集, Vol. , No. 1-6-10, pp. 19-20
Citation(English)	, Vol. , No. 1-6-10, pp. 19-20
発行日 / Pub. date	2003, 9

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)による雑音適応法における木構造クラスタリング法[1]を提案した。手法[1]では、SNRごとに木構造HMMモデルを作成し、入力音声に対し、最適なモデルを選択した。この論文では雑音の種類及びSNRを統合した木構造クラスタリング手法を提案する。

2. 区分線形変換雑音適応法における木構造クラスタリング

2.1 区分線形変換による雑音適応手法

一般に、音声に雑音が重畳されたときに、雑音音声信号 \hat{s} は次のようにモデル化される。

$$\hat{s} = F(s, n, SNR)$$

s, n, SNR はそれぞれクリーンな音声信号、重畳する雑音、信号対雑音比を表す。 F はケプストラム空間では一般に非線形変換になる。この問題に対応するために、ケプストラム空間での確率分布を表すHMMに対して、HMM合成法[2]やneural network法[3]などの種々の非線形処理が研究されてきた。しかし、これらの手法には二つの欠点がある。①複雑な処理と大きな計算量を必要とする。②雑音 n 及び信号対雑音比SNRが常に変動するような雑音音声に対応するのが難しい。これに対し、我々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方法を提案した。さらにHMMパラメータ空間(雑音が重畳した音声のHMM空間)を雑音の性質とSNRによって木構造に区別化し、入力音声の条件に最も適合した部分空間を選ぶ方法[1]を提案した。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[4])を行う。

2.2 雑音重畳音声GMMによる木構造クラスタリング手法

以前の手法[1]ではSNRごとに雑音のクラスタリングを行い、木構造HMMモデルを作成し、また、入力音声を認識する時、SNRと雑音の種類を二段階の探索を行うことが必要であった。即ち、まず各SNR条件で尤度最大のモデルを選択し、次にあらゆるSNR条件の最適なモデルを比較して、その中の一番良いモデルを全体の最適なモデルとしていた。このためモデル選択の計算量が多くなる。そこで本報告では雑音の種類とSNRを同時に考慮した木構造クラスタリングによる雑音適

応手法を提案する。この方法ではすべてのSNR条件のあらゆる雑音重畳音声をクラスタリングすることによって、雑音の種類とSNRの変動を一つの木にまとめ、入力音声を認識する時に一段階の探索で最適なモデルを選択する。

2.3 クラスタリング

雑音データは電子協雑音データベース[5]の30種類の雑音を用いた。そのうちの28種類を学習に、ほかの2種類(駅のコンコース"eki", 百貨店のエレベータ"hyakka")をテストに用いた。クリーンな音声に種々のSNRで各雑音を重畳させた後、CMSを適用した。Baum-Welchアルゴリズムを用いて64混合の各雑音重畳音声GMMを学習した。雑音重畳音声GMM間の距離を計算し、3種類のSNR(SNR=5, 10, 15dB)の距離行列(84行84列)に基づいてクラスタリングを行った。

3. 認識実験

3.1 音響モデル

実験で利用する音響モデルは、「話し言葉工学」プロジェクト[7]で作成された音声コーパス中の男性話者による338講演、約59時間分の音声データを用いて作成した2,000状態16混合の状態共有型(triphone) HMMである[8]。音声特徴量は、25次元のMFCCベクトル系列を用いた。

3.2 言語モデル

この実験は対話システムでの音声入力による飲食店舗検索とする。ユーザーは場所(最寄り駅)、料理の種類などの検索条件を発声することにより、希望の店舗の詳細情報を得る。発話内容を受理する言語モデルは、それぞれの発話内容用のテキストコーパスから作成される。クラス言語モデルを用い、クラスに属する単語は全て等確率で生起すると仮定する[9]。

3.3 評価用データ

2種類の評価用データを用いた。まず、10名の話者の50発話の対話音声に対し3種類のSNR(SNR=5, 10, 15dB)で、学習に用いなかった2種類の雑音("eki", "hyakka")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、三日間にわたって二箇所(横浜駅"yokohama", オフィス"office")の実際の雑音環境で収録した12名の話者の540文(平均SNR=10, 12dB)を使用した(Test2)。

4. 認識実験結果

まずTest1の2種類の雑音に対し、モデル選択による認識実験を行った。尤度最大基準に基づいて、木構造のルートから下にたどり最適なモデルを選択し、認

識に用いる。クリーンHMMの結果(“Baseline”), SNRごとに木構造モデルを構築した場合(“従来法”), SNRと雑音の種類を統合した木構造モデルを用いた場合(“提案法”)の単語正解精度(ACC%)を図1.2に示す。“提案法”により“従来法”と同等或いはより高い認識率が得られた。最適モデルの選択にかかる計算時間は、“提案法”により、“従来法”に比べて一文当たり平均で約1/3に削減された。

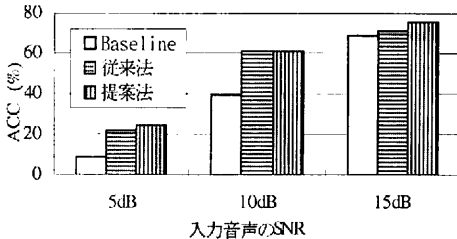


Fig. 1: Comparison of model selection for multi-trees (previous) and one-tree clustering (new) methods (eki noise)

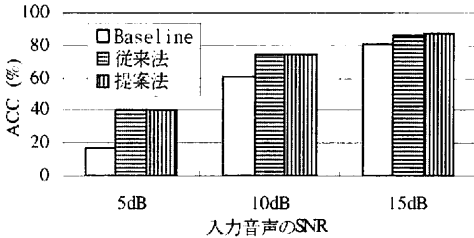


Fig. 2: Comparison of model selection for multi-trees (previous) and one-tree clustering (new) methods (hyakka noise)

次に、Test1に対し、最適なノードを選択し、尤度がさらに最大化するように線形変換をする実験を行った。クリーンHMMの結果(“Baseline”), SNRごとに木構

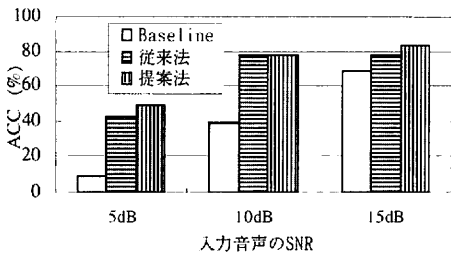


Fig. 3: Comparison of PLT for multi-trees (previous) and one-tree clustering (new) methods (eki_1 noise)

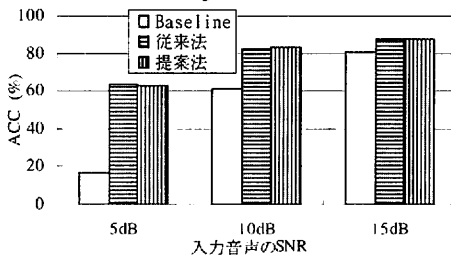


Fig. 4: Comparison of PLT for multi-trees (previous) and One-tree clustering (new) methods (hyakka noise)

造モデルを構築した場合(“従来法”), SNRと雑音の種類を統合した木構造モデルを用いた場合(“提案法”)の単語正解精度(ACC%)を図3.4に示す。ベースラインに比べ、提案法は6種類のデータの平均で単語誤り率は49.8%低下した。

次に、Test2に対し、モデル選択だけの実験及びモデルを選択してからさらに線形変換をする実験を行った。クリーンHMMの結果(“Baseline”), モデル選択だけの場合(“モデル選択”), 区分線形変換手法(“PLT”)の単語正解精度(ACC%)を図5.6に示す。ベースラインに比べ、単語誤り率は平均35.0%低下した。

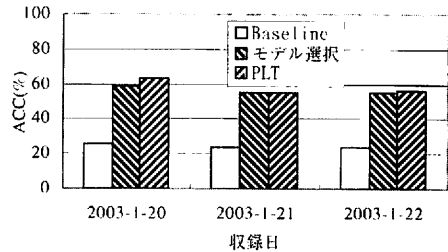


Fig. 5: Recognition results for Test2 (yokohama noise)

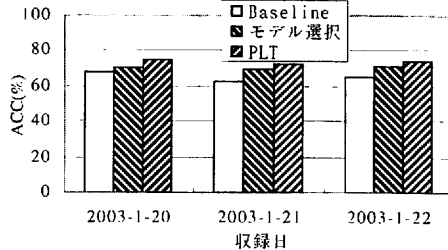


Fig. 6: Recognition results for Test2 (office noise)

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おいて、雑音の種類とSNRを同時に考慮した木構造クラスタリング手法を提案した。二種類の雑音に対する認識実験で、それぞれ49.8%、35.0%誤り率を削減し、提案手法の効果を確認した。従来法に比べて処理時間の削減だけでなく、性能の向上が見られたのは、木構造中のHMMモデルが、複数のSNR条件下の雑音重畳音声から学習されるため、一文の中にSNRの変動がある音声に有効であるためと考えられる。今後の課題には、学習雑音数の増加、雑音区間の自動切り出しなどがある。

参考文献

- [1] 張, 大江, 古井, 春季音講論, pp. 37-38, 2003.
- [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] C.J. Leggetter et al., Computer Speech and Language, Vol. 9, pp. 171-185, 1995.
- [5] http://www.nslab.is.tsukuba.ac.jp/corpus/noise_db.html
- [6] 菅, <<多変量解析の実践>>, 現代数学社, 1993.
- [7] 古井, 前川, 井佐原, 日本音響学会誌, vol. 56, no. 11, pp. 752-755, 2000.
- [8] 篠崎, 古井, 秋季音講論, pp. 87-88, 2002
- [9] 田熊, 森山, 岩野, 古井, 秋季音講論, pp. 79-80, 2002.