論文 / 著書情報
Article / Book Information

論題(和文)	雑音環境を考慮した自律型話者適応化
Title(English)	
著者(和文)	高木啓三郎, 篠田浩一, 服部浩明, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成8年度春季研究発表会講演論文集, Vol. 1-5-24, No. , pp.
Citation(English)	, Vol. 1-5-24, No. , pp.
発行日 / Pub. date	1996,

1—5—24 雑音環境を考慮した自律型話者適応化 *

⊙ 高木 啓三郎　篠田 浩一　服部 浩明　渡辺 隆夫
（ＮＥＣ　情報メディア研究所）

1. はじめに

本報告では、雑音環境下で発声された音声に対する教師なし逐次話者適応方式において、話者適応の際に事前に環境の要因を除去する方法が有効であるという考え [6][7] に基づき、高速環境適応 (REALISE 法)[1] -[3] を用いて１発声毎に環境の影響を除去した後に、なお存在する分布毎の特徴量の差異を、木構造化された確率分布を用いた自律型話者適応化手法 (ACT)[4][5] を用いて蓄積・適応化する方式について検討した。

2. 教師なし逐次適応化方式

雑音下で発声された音声に対する話者適応化を行なう場合、話者固有の変動と雑音環境による変動の２つの変動を別個に対処することが必要となる。話者固有の変動は比較的長期の間安定しているため、その変動を蓄積して対処する方式が有効と考えられるが、環境による変動は、事前に予測することが出来ず、しかも常に変化すると考えられる。このため、長期間蓄積して用いる話者固有の変動に環境の要因を擾乱として含んでいると期待した話者適応化性能が得られない。これに対処するためには、環境による変動は話者適応とは別に、例えば１発声毎に除去することが必要となる。

ここでは、REALISE 法を用いた単語発声毎の環境適応化を行い、環境に関する平均的な差異を除去した後になお存在する分布毎の特徴ベクトルの差異を話者固有のパラメータとして蓄積し、ACT を用いた逐次適応化を行なう。ここで対象とした特徴ベクトルは 10 次元のメルケプストラムである。

2.1. 高速環境適応方式 (REALISE 法)

REALISE 法は、１単語程度の短い発声から、標準パタンと入力との間の乗算性雑音、付加雑音のスペクトル上での異なりを推定し、適応化を行なう方式である。

適応化は、以下のように行なう。

$$\tilde{\mathbf{W}}(t) = \frac{\mathbf{S}_v - \mathbf{N}_v}{\mathbf{S}_w - \mathbf{N}_w}(\mathbf{W}(t) - \mathbf{N}_w) + \mathbf{N}_v \quad (1)$$

ここで、$\mathbf{W}(t)$ および $\tilde{\mathbf{W}}(t)$ は各々適応化前および適応化後の標準パタンのスペクトル、$\mathbf{S}_v$、$\mathbf{N}_v$ は各々入力音声の音声、雑音の平均スペクトル、$\mathbf{S}_w$、$\mathbf{N}_w$ は各々標準パタンの音声、雑音の平均スペクトルである。なお、本アルゴリズムについては文献 [3] に詳しい。

2.2. ACT を用いた話者適応化

ACT は、予め作成された木構造確率分布を用い、適応化データ量に応じて適応化の複雑さを自律的に決定する方式である。

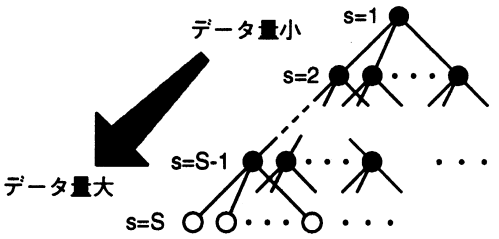

図 1: 木構造確率分布

図 1に木構造確率分布の例を示す (階層数 s=1 ～ S)。ここでは、REALISE 法を行なった後の標準パタンについて、分布毎のケプストラムの差異を ACT を用いて蓄積・適応化する。式 (1) で求めた不特定話者標準パタンからの環境適応化後の分布 k のスペクトル $\hat{\mathbf{W}}(k)$ に対応するメルケプストラム μ_k^R を用い、逐次適応における w 番目の発声が入力されたとした場合の w 番目の発声における分布 k の平均の差ベクトル $\delta_k^{(w)}$ を求める。

$$\begin{cases} \bar{\mathbf{x}}_k^{(w)} = \dfrac{1}{n_k^{(w)}} \displaystyle\sum_{t \leftrightarrow k} \mathbf{x}_t^{(w)} \\ L_k^{(w)} = n_k^{(w)} + L_k^{(w-1)} \\ \delta_k^{(w)} = \dfrac{L_k^{(w-1)}\delta_k^{(w-1)} + n_k^{(w)}(\bar{\mathbf{x}}_k^{(w)} - \mu_k^R)}{L_k^{(w)}} \end{cases} \quad (2)$$

ここで、$n_k^{(w)}$ は、w 番目の発声の分布 k に対応付けられたデータのフレーム数、$L_k^{(w)}$ は w

* Environmentally Robust Speaker Adatation with Autonomous Control, by Keizaburo Takagi, Koichi Shinoda, Hiroaki Hattori, and Takao Watanabe, NEC Corporation

番目までの発声のうち、分布 k に対応付けられたデータの累積フレーム数を示す。この結果を用いて全ての階層のノード (s,l) における適応化ベクトル $\Delta(s,l)$ を求め、各リーフ (最下位) ノードから上位へ順次辿り、予め定めたフレーム数の閾値を最初に越えるノード (s',l') の適応化ベクトル $\Delta(s',l')$ を用いて平均ベクトルの更新を行なう。なお、本アルゴリズムについては、文献 [5] に詳しい。

3. 評価実験

評価は、半音節 HMM による不特定話者音声認識 [8] をベースとした 100 単語認識で行なった。不特定話者半音節 HMM は、85 名の話者が発声した音素バランス 250 単語学習したものを用いた。

ここでは、環境が変化する場合を想定し、適応化音声と認識音声の環境が異なる以下の場合について評価した。6 名 (男性 3、女性 3) がアイドリング、50km/h 走行、100km/h 走行の 3 種類の条件で自動車内助手席で発声した電子協地名 No.1 ～ 100 の各 100 単語のうち、100km/h 走行の音声を適応化に用い、アイドリングおよび 50km/h 走行の 200 単語を認識に用いた。

分析はサンプリング周波数 11KHz、帯域制限 0.1 ～ 5KHz、フレーム周期 10ms で行なった。特徴量は 21 次元 (パワー変化量 1、メルケプストラム 10、メルケプストラム変化量 10) である。なお、ここでは評価に用いた音声に対して 1 入力によるスペクトルサブトラクション [9] を前処理として用いている。

比較のため環境適応手法として REALISE 法の代わりにケプストラム平均値整合 (CME 法) [10] [11] を ACT と併用した場合についても評価を行なった。また、木構造のトポロジーおよびフレーム数の閾値については、各々の条件について予備実験で最適化したものを用いた。結果を図 2 に示す。

図 2において、REALISE+ACT が最も高く有効であることが明らかとなった。また CME+ACT が ACT のみよりも高かったことから、適応化データと評価データとの環境が異なっている場合には、いったん環境の除去を行なう方式が有効であることを示している。CME 法は伝送歪み成分のみ考慮した方式であり、自動車内音声のように付加雑音の存在が無視出来ない場合にはその効果が低下し、CME+ACT は REALISE+ACT に比べてやや低い性能となっている。

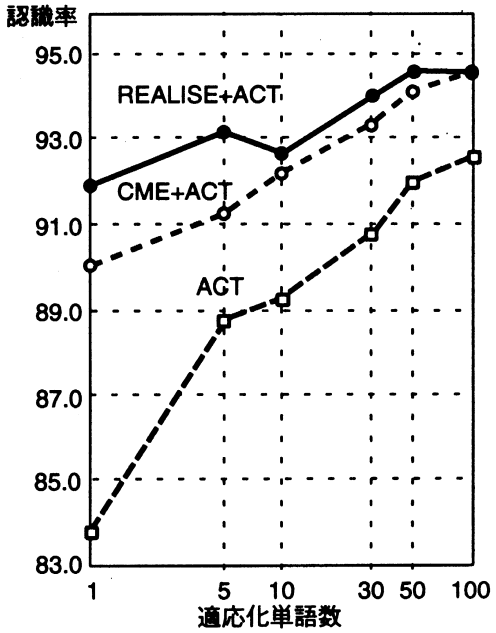

図 2: 適応化性能の比較

4. おわりに

REALISE 法を用いて環境の影響を除去した後に ACT を用いて特徴ベクトルの分布毎の差異を蓄積・適応化を行なう、教師なし逐次話者適応方式の提案を行なった。評価実験から、自動車内発声のように環境が変化する場合においても安定した性能向上が見られ有効性を確認した。

謝辞

日頃から熱心に御討論いただく音声言語研の諸氏に感謝致します。

参考文献

[1] 高木、服部、渡辺、音響講論, 2-P-8(1994.3).

[2] 高木、服部、渡辺、信学技報、SP94-19(1994.6).

[3] K. Takagi, et al., 音響論 (E)16, 5(1995.9).

[4] 篠田、渡辺、音響講論, 2-5-10(1995.3).

[5] K. Shinoda, et al., EUROSPEECH95, pp. 1143-1146(1995.9).

[6] M. Feng, ICASSP95, pp. 704-707(1995.5).

[7] Y. Zhao, ICASSP95, pp. 712-715(1995.5).

[8] 渡辺他、信学論文誌、J72-D-II、8(1989.8).

[9] 高木、吉田、渡辺、音響講論、2-5-3(1991.3).

[10] S. Lerner, et al., ICASSP92, pp. I-261-264(1992.5).

[11] 高木、服部、渡辺、音響講論, 2-5-14(1995.3).