

TR-IT-0004

話者適応を用いた不特定話者音声認識
Speaker-Independent Speech Recognition using Speaker
Adaptation

宮沢 康永
Yasunaga Miyazawa

1993.7

概要

著者が 1991 年 7 月 21 日から 1993 年 7 月 20 日まで、ATR 自動翻訳電話研究所及び ATR 音声翻訳通信研究所で行なった、話者適応を用いた不特定話者音声認識に関する研究発表論文リストを載せた報告書である。研究発表の内容は、「移動ベクトル場平滑化話者適応における標準話者選択方式」及び「全音素エルゴディック HMM を用いた教師なし話者適応方式」である。本テクニカルレポートの最後に、「全音素エルゴディック HMnet を用いた教師なし話者適応」の方式と実験結果について載せる。

ATR Interpreting Telecommunications Research Labs.
ATR 音声翻訳通信研究所

© (株)ATR 音声翻訳通信研究所 1993

© 1993 ATR Interpreting Telecommunications Research Laboratories

目次

1	はじめに	1
2	発表論文リスト	1
3	付録	2
A	付録: 日本音響学会 平成 4 年度秋期	3
B	付録: 電子情報通信学会 投稿論文	5
C	付録: Abstract for ICASSP94	13

図目次

1 はじめに

本テクニカルレポートに、話者適応に関して著者が発表した論文のリストを載せる。著者は、話者適応を用いた不特定話者の連続音声認識の性能向上を目的として研究を行ってきた。話者適応における新しい標準話者選択方式、統計的な言語情報を利用した教師なし話者適応方式を提案し、その有効性を示した。

今までの主な研究の概要を以下に示す。

● 移動ベクトル場平滑化話者適応方式における標準話者選択方式の検討

「音声の個人性」の類似度の基準として HMM の出力尤度を用いた場合、話者適応を行なった後に最も入力話者音声に類似するモデル (HMM) とは、話者適応した後の HMM の出力尤度が最大になる標準話者モデルであると考えられる。そこで話者適応後の HMM の出力尤度を「話者正規化 HMM 尤度」と名付け、標準話者選択を行なうための評価基準として用いる方法を提案した。本方式の有効性を、標準話者 12 人、未知話者 10 人を用いた音素認識実験で確認した。

● 全音素エルゴディック HMM を用いた教師なし話者適応方式

従来の多くの教師なし話者適応方式では、言語的な知識が使われていなかった。本研究ではテキストデータより得られた言語的な統計情報を用いることにより、教師なし話者適応方式の性能向上を図った。構造は、全ての音素 HMM をエルゴディックに連結した「全音素エルゴディック HMM」とし、各音素間の遷移確率の初期値に音素 bigram 確率値を用いた。本方式が話者適応及び発話様式適応に対して有効であることを、音素認識、文節認識実験で確認した。更に音素モデルの初期モデルを不特定話者モデルにすることにより高性能が得られることも文節認識実験により確認した。

● 全音素エルゴディック HMnet を用いた教師なし話者適応

我々は統計的な言語情報として音素 bigram を用いる教師なし話者適応方式を提案し、その有効性を確認した。更に性能向上を目指し、より強い言語制約である音素環境依存音素 bigram と、より精度の高い音響モデルである隠れマルコフ網 (HMnet) を統合した「全音素エルゴディック HMnet」を用いる教師なし話者適応方式を提案する。実際に実験で用いた構造は HMnet の各状態を全結合し、その状態間遷移確率をテキストデータより得られた状態 bigram 確率値とした。本方式の有効性を文節認識実験で示す。

2 発表論文リスト

1. 宮沢 康永, 嵯峨山 茂樹: “移動ベクトル場平滑化話者適応方式における標準話者選択方式の検討,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-5-2, pp. 121-122 (1992.10).
2. 宮沢 康永, 嵯峨山 茂樹: “全音素エルゴディック HMM を用いた教師なし話者適応方式,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-5-3, pp. 123-124 (1992.10).
3. 宮沢 康永, 嵯峨山 茂樹: “全音素エルゴディック HMM を用いた教師なし話者適応における初期モデルの検討,” 日本音響学会平成 5 年度春季研究発表会講演論文集, 2-4-10, pp. 37-38 (1993.03).
4. 鷹見 淳一, 宮沢 康永, 永井 明人, 嵯峨山 茂樹, “話者適応型 SSS-LR 連続音声認識方式における標準話者予備選択の効果,” 音講論集, 2-Q-16 (1993.03).

5. 宮沢 康永, 大倉 計美, 嵯峨山 茂樹: “全音素エルゴディック HMM による教師なし話者適応,” 電子情報通信学会技術研究報告, SP92-75, pp. 15-20 (1992.10).
6. Yasunaga Miyazawa, Shigeki Sagayama: “Speaker-normalized HMM-likelihood for selecting a reference speaker in speaker-adaptive speech recognition,” Proc. of Fourth Australian International Conference on Speech Science and Technology pp. 431-436, (1992.12).
7. Y. Miyazawa, (S. Sagayama): “An All-Phoneme Ergodic HMM for Unsupervised Speaker Adaptation,” Proc. ICASSP93.
8. 宮沢 康永, 大倉 計美, 嵯峨山 茂樹: “全音素エルゴディック HMM を用いた教師なし話者適応,” 電子情報通信学会論文誌 D-II, 投稿中.
9. 宮沢 康永, 鷹見 淳一, 松永 昭一, 嵯峨山 茂樹: “全音素エルゴディック HMnet を用いた教師なし話者適応,” 日本音響学会平成5年度秋季研究発表会発表予定 (1993.10).

3 付録

これらの方法、実験結果を付録に示す。