

216 自由会話音声の話速研究を指向した HMMによる音素境界自動割付ツール構築について

松村 真言 広重 真人 荒木 健治 柄内 香次

北海道大学大学院工学研究科

1 はじめに

自由会話音声を対象として話速などの韻律情報の研究を行なう際に、研究対象となる大量の自由会話音声に対する音素ラベリングなどの付加情報の作成は必要不可欠なものである。特に、通常用いられるくだけた自由会話文では調音結合が激しく、既存の市販品などの自動認識システムでの認識は困難が予想され、ラベリングに認識の結果を用いることは難しい。しかし筆者らの局所的話速研究 [1] においては、発話を構成する音素などの時間境界のデータが重要であり、必ずしも音素の内容まで自動認識する必要はない。発話内容について、文単位での書き起こしテキストの作成は、音素ラベルの付与に比べれば負担が軽いと考えられる。

そこで本研究では、音素の認識はせずに、発話内容を書き起こしテキストの形で与え、音素境界のみを自動的に決定するシステムの可能性を検討する。音素境界決定には HMM の強制割付の機能を利用し、具体的には一般的な HMM ツールである HTK を用いる。本稿では、自由会話音声の音素境界自動割付ツール構築の第一歩として、一般的な朗読音声とその書き起こしテキストを用いた音素境界割付ツールの作成について報告する。

2 HMM における強制割付

本稿では、音声認識用の HMM ツールを音素ラベル割付のみに機能限定して用いることになる。この場合 HMM は書き起こしテキストで示される順番を守って学習済の音素 HMM を配列し、各音素 HMM の境界のみを最速に調整する。これは、HTK において「強制割付 (forced alignment)」と呼ばれる機能に相当する。

書き起こしテキストが与えられているので、HTK において通常用いる認識可能な複数の文を表す文法記述は必要ない。また、音素境界を求めることが目的であるため、各音素が独立である monophone モデルを用いる。

3 実験

英文を対象とした HTK マニュアル [2] に沿って、まず英文の音素ラベル割付実験を行った。学習文・評価文共に当研究室で録音した。文は日本語を母国語とする話者 1 名によって発話された英文 50 文を用いた。割付実験は学習に用いたデータと同一のデータを用いる closed なものとした。割付結果の例を図 1 に示す。

DIAL ONE ONE OH

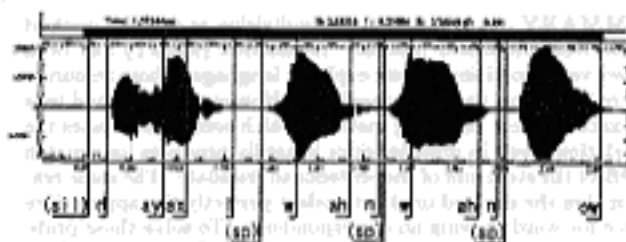


Fig. 1 ラベリング割付実験結果の例

4 考察

図 1 によると、closed な実験ではあるが、概ね良好な結果が得られているといえる。現在、日本語文への適応を進めており、その際、学習済の HMM パラメータとして「標準的な日本語音素 HMM」[3] を利用し、割付対象データとしては ATR 連続音声データベースを利用して、システムの調整を行っている。

また、学習話者と割付時の話者が異なることの影響や、話速研究に適したセグメンテーション単位 (モーラや文節など) などの検討が必要であると考えられる。

5 まとめ

本稿では、HMM を用いて、音声に音素ラベルを付加するシステムの構築経過を報告した。

今後の計画として、日本語のラベル付けシステムの完成後、当研究室で独自に録音した自由会話音声への適用を考えている。

参考文献

- [1] 高丸他: "各種の自由会話音声に含まれる局所的話速変化の特徴に関する検討", 音講論, 2-P-2, pp.271-272(1999-03)
- [2] Steve Young 他: 『The HTK Book ver. 2.1』 Entropic Cambridge Research Laboratory
- [3] 名古屋大板倉研究室, <http://www.itakura.nuee.nagoya-u.ac.jp/index-j.html>
- [4] 夏目貴史他: "言語的な情報を利用した韻律データベース作成の自動化", 音講論, 2-P-7, pp.281-282(1999-03)