

210 音声波形からの帰納的学習を用いた音声翻訳手法

村上浩司 広重真人 荒木健治 堀内香次

北海道大学大学院工学研究科

1 はじめに

現在、音声翻訳分野では、様々な手法やアルゴリズムを利用し研究が進められているにもかかわらず、多くの問題が残されている。特に、有用な音声翻訳を実現するためにはシステムの即時性が最も必要とされているが、まだ確立されていない。

しかし、原言語の音声をもとに目的言語に翻訳するためには大きく分けて音声認識、形態素解析、構文解析、機械翻訳の処理があり、それぞれの処理もまた細分化される。さらに、これらの処理もまた複雑なためにそれぞれを精度良く実現させるのは難しく、たとえこの四処理の精度がそれぞれ90%であっても、全体の精度は約66%にしかない。この点が音声翻訳の難しさであり、全体の精度を上げることは困難とされている。

そこで本稿では多くの複雑な処理を省略し、音声波形に着目した簡素なシステムを用いる。対応する二言語の文を直接的に変換することにより、及方向の対話的な実時間の翻訳処理を可能とし、言語に依存することなく汎用的な手法について提案、検討する。

2 帰納的学習を用いた音声翻訳手法

2.1 従来の手法との比較

音声翻訳は、音声の入力に対して音声認識を行い、その結果を形態素解析で品詞列に変換した後に、構文解析でその文の構造を解析し、その後目的言語の構文構造に変換し、文生成をした後、最後に音声を合成することで実現される。

これに対して本手法では、そのシステムの構造をより簡素化し、原言語音声と目的言語音声の音声的特徴を、直接結び付け処理することでスピードアップを図り、かつ構造の複雑さによる正解率の減少を防ぐ構造となっている。両手法の処理過程を図1に示す。

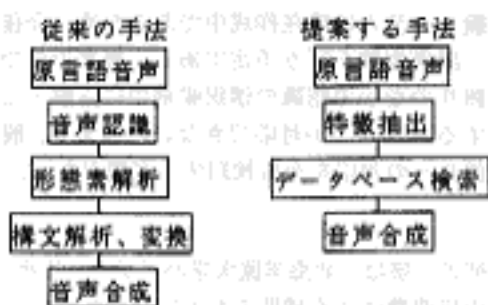


図1 処理過程の比較

2.2 手法の特長と帰納的学習

本手法では、音声の特徴を直接利用するだけでなく、音声の中の単語や節には着目せず、音響的な共通部分と差異部分の関係を利用することが特長である。これにより、単語、音節等の切り出し処理が不要となる。着目したそれぞれの部分の声道情報と時間的な特徴変化が、帰納的学習によりルール中に表現される。帰納的学習においては、それらのルールをデータベース化し、それを繰り返すことで汎用性を高める。

3 処理構造

本手法の処理構造を図2に示す。

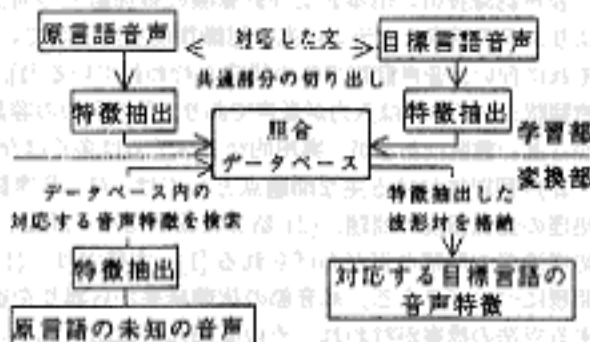


図2 処理構造

4 音声処理

4.1 データ

音声波形には様々な情報の他に雑音なども含まれている。そのため、音声波形から声道情報を抽出し、データとして用いる。本手法では、DATにより48kHzで録音した音声を10kHzにダウンサンプリングする。また、時間的な音声の特徴変化を捉えるため短時間スペクトルを用い、特徴抽出はLPCケブストラムとする。

4.2 DP マッチング

共通部分の抽出にはDP マッチングを用いることで、対応するベクトル間の話速の違いを吸収することができる。

5 適応される例

図3に適応される例を示す。ここでは、近い意味をもつ英語、日本語の文をそれぞれ二文ずつ比較し、共通部分を抽出していく。また、便宜上波形で示す。

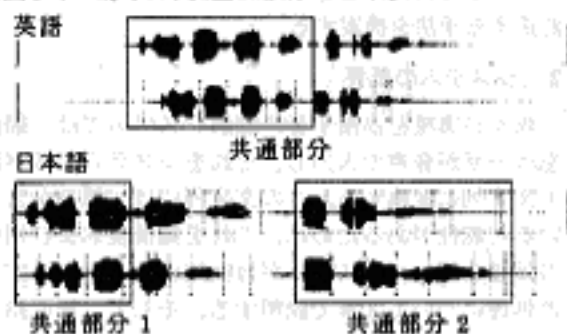


図3 英語文と日本語文の比較

6 おわりに

言語に依存しない汎用的な音声翻訳手法を提案、検討した。今後は本手法の有効性を、実験により示す予定である。

参考文献

- [1] K.MURAKAMI, M.HIROSHIGE, Y.MIYANAGA, K.TOCHINAI: "A Prototype System for Continuous Speech Recognition using Group Training based on Neural Network", ITC-CSCC, pp1013-1016, 1997
- [2] 森元, 田代, 竹澤, 他: "音声翻訳実験システム (ASURA) のシステム構成と評価", 情報処理学会論文誌, Vol.37, No9, pp.1726-1733(1996)