

## 6 D-3

## HMM-LR 音声認識の大語彙への適用

北 研二 江原 暉将 森元 遼

ATR 自動翻訳電話研究所

## 1 はじめに

ATR では、HMM 音韻認識と拡張 LR 構文解析法を統合した音声認識の手法 HMM-LR 法を提案し、日本語の文節認識実験によって高い認識率が得られることを示した。

これまでの、約 1000 語程度の規模の語彙に対して認識実験を行ってきたが、今回、語彙を数千語に増やした場合について認識実験を行ったので、その結果を報告する。

## 2 HMM-LR 音声認識システム [1][2]

図 1 に HMM-LR 音声認識システムの構成図を示す。HMM-LR 音声認識システムは大きく分けて、HMM 音韻照合部と予測 LR パーザから成る。HMM 音韻照合部は、予め学習された音韻の HMM を用いて音韻照合を行い、指定された音声区間に対する音韻の照合スコアを計算する。予測 LR パーザは、LR 構文解析表と呼ばれる表を参照して、文法上で次にくる音韻を予測し、それらの音韻に対して HMM 音韻照合部を駆動する。照合に成功したすべての音韻に対して、並行して音韻連鎖の枝をのばしていく。実際には、音韻連鎖の枝をのばす過程において解析する候補の数が増加してくるので、照合スコアによるビームサーチ

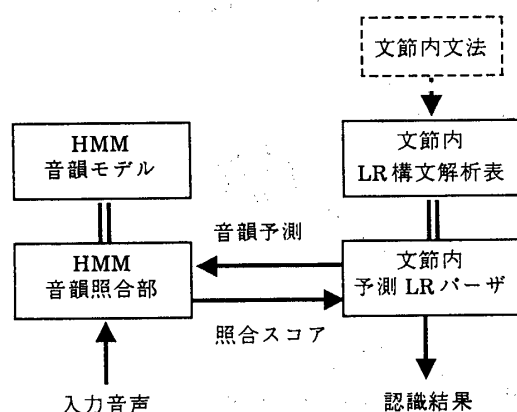


図 1. HMM-LR 音声認識システム

を行いながら認識を進める。照合がすべて終わった段階で照合スコアの高い候補を認識結果として出力する。

## 3 大語彙認識実験

## 3-1 タスクおよび文法

現在のところ、音声認識では文節発声を仮定しており、このため、文節内部の文法的制約を記述した文法を HMM-LR 音声認識システムでは用いている。文節文法は、我々の認識対象であるタスク「国際会議への参加問い合わせ」に対する文節内部の構文規則および語彙規則から構成されている。音声認識の単位は音韻であるため、語彙規則には各単語の音韻列が書かれている。

## 3-2 実験概要

今回の実験では、語彙規則中の単語数を増やして認識実験を行った。このために、ATR 言語データベース [3] 中の頻出語と新明解国語辞典の重要語(固有名詞、数詞、機能語、間投詞は除く)に対して単語の音韻列を作成した。表 1 にそれぞれの単語数、異なり単語数を示す。なお、表 1 で基礎文節文法と示されているものは、日本語の基本的な機能語(助詞、助動詞等)に加え、我々のタスクに現れる基本語彙を含んだ文節文法の語彙数である。

認識実験では、言語データベースおよび新明解重要語(計 7190 語)から任意に語彙を抽出し、それを基礎文節文法に付け加えたものを用いた。このようにして、1000 語ずつ語彙を増やし(最大 7934

表 1. 単語数、異なり単語数

|          | 単語数   | 異なり単語数 |
|----------|-------|--------|
| 基礎文節文法   | 744   | 497    |
| 言語データベース | 2,050 | 1,747  |
| 新明解重要語   | 5,140 | 4,445  |

Very Large-Vocabulary Speech Recognition Using HMM-LR

Kenji Kita, Terumasa Ehara, Tsuyoshi Morimoto

ATR Interpreting Telephony Research Laboratories

語まで)、認識実験を行った。ただし、音声認識率は文法に含まれている語彙によって変わるので、各語数に対して、ランダムな語彙抽出と認識実験を3回行い、その平均値をその語数に対する認識率とした(実験のパラメータは[2]と同じ)。

また、音声認識では言語の統計的な情報を用いることにより、認識精度の向上が期待できるので、文節文法に加え、日本語の音節の3つ組確率(trigram)を用いた場合についても実験を行った。

HMM-LR 音声認識システムでは、言語の構文情報は LR 構文解析表の中に埋め込まれている。大語彙化した場合、LR 構文解析表の大きさがどうなるかも、システムにとって重要な要因である。単語数の増加にともなう、LR 構文解析表の大きさについても調べてみた。

### 3-3 実験結果

図2は、単語数と文節認識率の関係を示している。単語数の増加にともない、認識率も低下しているが、第5位までの累積認識率では、いずれも94~95%以上の認識率を達成している。正しい文節が5位程度までに認識されていれば、後の文単位での言語処理の段階で正しく認識されることが期待できるので、HMM-LR 音声認識システムは数千語規模の大語彙化に対しても十分な認識性能を持っていると考えられる。

また、日本語の音節の3つ組確率を用いることにより、単語数の増加にともなう文節認識率の低下は小さくなっており、言語の統計情報が有効に働くことを示している。

図3は、単語数と LR 構文解析表の大きさ(状態数)の関係を示したものである。単語数の増加にともなう LR 構文解析表の状態数の増えかたは、ほぼ線形である。これは、HMM-LR 音声認識システムが単語数の増加に対して耐性があることを示している。

図4は、単語数と文法の音韻当たりの複雑さ(phone perplexity)[4]の関係を示したものである。perplexityは単語数が増えても、ゆるやかに増えるだけである。日本語の文節構造は、基礎文節文法で一応網羅しているので、8000語以上に単語数を増やしても十分な認識性能が得られる可能性を示している。

#### 謝辞

研究の機会を与えて頂いた ATR 自動翻訳電話研究所 樽松明 社長に深謝致します。また熱心に討論して頂いた同研究所 データ処理研究室の皆様にも感謝致します。

#### 参考文献

- [1] 北,川端,斎藤:「HMM 音韻認識と拡張 LR 構文解析法を用いた連続音声認識」,情報処理学会論文誌 Vol. 31, No. 3 (1990).
- [2] 花沢,北,中村,川端,鹿野:「HMM-LR 音声認識システムの性能評価」,日本音響学会論文誌 Vol. 46, No. 10 (1990).
- [3] 江原,小倉,森元:「対話データベースの構築」,40 回情全大 6F-1 (1990).
- [4] 川端,鹿野,北:「音韻パープレキシティの提案」,音講論集 3-6-11 (1989).

図2. 単語数と認識率の関係

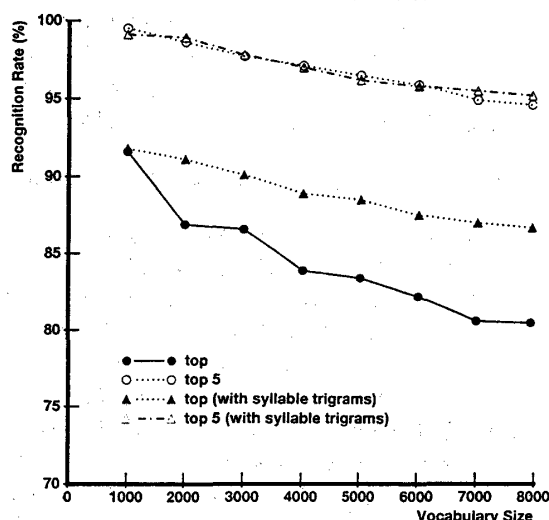


図3. 単語数と LR テーブルの大きさの関係

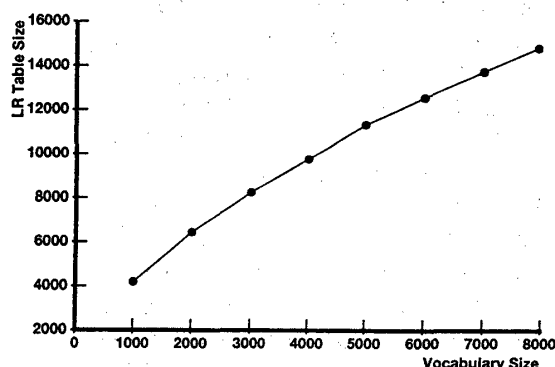


図4. 単語数と phone perplexity の関係

