

1. はじめに

近年、カーナビゲーションシステム等の車載情報機器において、使いやすさや安全性向上のために音声認識を用いたインタフェースが採用されるようになってきた。このようなシステムでは認識率がタスク達成時間に大きく影響を与えるため、低認識率語を開発段階で発見し対策を行うことはシステムの品質保証上、重要である。しかしながら、カーナビゲーションシステムでは数十万語規模の地名を認識しなければならないため、実際の発声を用いての認識性能評価は困難である。そこで、実発声を用いずに低認識率語を発見する手法を開発した。

2. 認識率の予測手法

2.1 単語対認識率と単語認識率

N 単語からなる語彙セット W を考える。各々の単語と他のすべての単語における単語対認識率の最小値（最小単語対認識率）、すなわち任意の単語 w_i と w_j からなる単語対における単語 w_i の単語対認識率を $r_{w_i, j}$ とすると、 $\min(r_{w_i, 1}, r_{w_i, 2}, \dots, r_{w_i, N})$ が w_i の単語認識率の上限と考えられる。よって、本手法では単語の認識困難性を最小単語対認識率から予測する。

2.2 対数尤度差分布による単語対認識率の予測

一般に、音声に対する単語の適合度（対数尤度）は各音素区間の対数尤度の和で計算できる。単語対の認識では正解の語の対数尤度から不正解の語の対数尤度を減じた値が正であれば、正しく認識できたことを表す。したがって対数尤度差の分布を求め、その正の領域の面積比から単語対認識率を予測できる。本手法では以下の方法で単語対数尤度差分布を予測する (Fig. 1)。

(1) あらかじめ全ての音素の組み合わせに対して音素対数尤度差分布を求める。

(2) 単語を音素に分解し、単語中の相異なる音素の対数尤度差分布の和によって単語対数尤度差分布を求める。

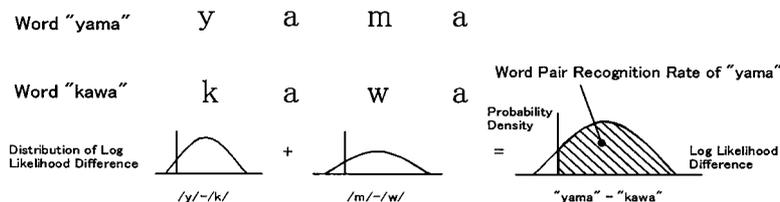


Fig. 1 Prediction of word pair recognition rate.

3. 単語対認識実験と結果

Table 1の実験条件から求めた音素対数尤度差分布を用いて単語対認識率の予測実験を行った。NTT単語データベースから任意に抽出した単語対350における実験結果をFig. 2に示す。予測値と実測値の相関係数は0.87であり、本手法によって単語対認識率を予測できることがわかった。

4. おわりに

対数尤度差分布に基づいて単語対の認識性能を予測する手法を検討し、低認識率語の発見に利用できることが確認できた。今後は、予測精度をより向上させる手法の検討を行う予定である。

参考文献

- 1) Tanaka, K., et al.: "Estimation of a Degree of Speech Recognition Difficulty for Word Sets", J. Acoust. Soc. Jpn(E), 19-5(1998), 339 (2001年2月1日原稿受付)

Table 1 Experimental condition.

Decoder	HVite(HTK)
Acoustic Model	monophone, 4mixture/state(IPA'97)
Utterance	NTT Japanese Words Database (Command set and Japanese city names)
Acoustic Analysis	Sampled at 16kHz with 16bit 12_m order MFCC + Δ MFCC + Δ Log Pow

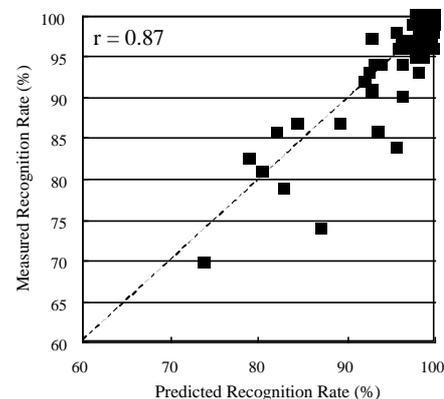


Fig. 2 Prediction result.