

Title	韻律情報を利用した雑音重畳音声の認識に関する研究
Author(s)	川崎, 真護
Citation	
Issue Date	1998-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1158
Rights	
Description	Supervisor: 下平 博, 情報科学研究科, 修士

A Study of Noisy Word Speech Recognition Using Prosodic Information

Kawasaki Shingo

School of Information Science,
Japan Advanced Institute of Science and Technology

February 13, 1998

Keywords: white noise, word speech recognition, pitch pattern, F0 generation model, prosodic likelihood.

1 Introduction

A frequency-weighted HMM and a spectral subtraction is one of the acoustic information. Generally, for noisy word speech recognition, these parameters is useful. But acoustic information is weak for noise. On the other hand a pitch pattern is robust noise because pitch pattern is very simple information. So we thought a pitch pattern is effective information for noisy speech recognition. This paper describes a new technique of noisy word speech recognition using prosodic information and acoustic information.

2 Word speech recognition system using prosodic information

The system consists of a phonetic recognition part using HMM and a prosodic recognition part using pitch pattern. Phonetic recognition part has phonetic HMM and phoneme sequences. Prosodic recognition part has pitch pattern templates and prosody dictionary. When speech input this system. Acoustic likelihood S^{ph} is calculated on Phonetic recognition part. The other side prosodic likelihood S^{pr} is calculated prosodic recognition part. Both results (S^{ph} and S^{pr}) is add up S using α . S is word recognition result.

$$S = (1.0 - \alpha)S^{ph} + \alpha S^{pr}$$

3 Study of prosodic recognition part

The pitch pattern extracted from the all words speech using lag window. But these pitch pattern contains error. So these pitch pattern modify smoothly using F_0 generation model. This model is consist of the addition to a constant minimum frequency of only two kinds of functions corresponding to two kinds of commands. One is a phrase command that initiate a steep rise followed by a gradual decay to zero. The other is an accent command that is a stepwise function, which rapidly reaches its target level, and fall back to zero after a delay.

These all pitch pattern are classify J categories using LBG clustering. The pitch pattern templates are these categories. The prosody dictionary write that probability of pitch pattern appearance in categories.

Prosodic likelihood S_i^{pr} is computation by next expression.

$$S_i^{pr} = - \sum_{j=1}^J p_{ij} d_j$$

p_{ij} is word i probability of appearance class j based on prosody dictionary. d_i is pitch pattern distance between pitch pattern of the input speech and template j .

4 Experiment

Recognition experiments use 520 words by 20 male speakers (m101 ~ m120). 15 mail speakers data (m101 ~ m115) are use training data (training HMM or make templates and so on). The rest 5 mail speakers data (m116 ~ m120) are use testing data. These testing data add on white noise, giving SNR gains of 45dB, 25dB, 5dB.

Recognition experiments using testing data were carried out using prosodic information.

5 Results

When the prosodic information was considered, error rate was reduced for all SNR words speech compared with discrete HMM word recognition results. When α is 0.90. The error rate was most reduced by 8% compared with discrete HMM word recognition results. If α is variable for all words. The error rate was most reduced by 50% compared with discrete HMM word recognition-results. These results indicated that prosodic information is effective when noisy word speech recognition.