A Study of Pitch Contour Restoration Based on Superpositional Prosodic Model

Kumata Akihiro

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

1997 February 14

Keywords: pitch contour,superpositional model,restoration,filter,beam search.

1 Introduction

A pitch contour is one of the prosodic information. For speech recognition,a pitch contour is useful. But a pitch extraction method is not adequate. Even if we assume that its extraction is reliable, we get only partial data because a pitch contour is not defined at unvoiced segments. So a pitch contour is often modified smoothly. We have modified a pitch contour by using numerical analysis such as linear interpolation or spline interpolation and so on. Those methods are likely that a pitch contour shall lose prosodic information. Thus this paper presents a method for pitch contour restoration based on superpositional prosodic model.

2 Superpositional prosodic model

A pitch contour is modeled as a sequence of parametered commands by using the superpositional prosodic model. The model consists 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.

We use two kinds of inverse filters derived by superpositional prosodic model. Those filters detect an amplitude of respective commands from a pitch contour. But filters detect too mach commands from a real pitch contour extracted from the speech wave. Thus it is necessary to select only proper commands.

3 Pitch contour restoration

The algorithm operate to minimize the distortion of a command sequence. The distortion of a command sequence is defined as the weighted square difference between the observed pitch and the modeled pitch where the weights are given by the pitch extraction reliability measure. We adopt beam search to select proper commands.

3.1 Estimation of parametered command

Phrase command and accent command is detected by the filters. We attempt two kinds of input to a filter.

method 1 The pitch contour extracted from the speech wave directly is inputed to filters. But by using this method filters detect an amplitude with an error by commands occured in the past.

method 2 The difference between the observed pitch and the estimated pitch is inputed to filters. This method can detect an amplitude without an error.

But by using these filters the pitch extraction must be adequate at the time which a command occured. Then We propose another filter.

a method using unit component of command The pitch contours generated by respective commands divided by the pitch contour generated by command with unit amplitude make amplitudes of commands.

3.2 Time sequential beam search

The input pitch is processed in a Left-Right manner. A beam width is defined the number of command sequences taken thought at each processing time. And a beam width is limited to M. N-best value is defined the number of command sequences assumed command to occur at processing time. And N-best value is limited to N. At each time it select M command sequences with smaller distortion.

3.3 Pitch contour restoration

Finally the pitch contour is restored from the command sequence with the smallest distortion.

4 An evaluation by a distortion

We calculated the distortion between restored pitch contours and ideal pitch contours. And our methods compared with usual methods.

5 accent phrase segmentation using restored pitch contours

We experimented with accent phrase segmentation using restored pitch contours. And our methods compared with usual methods.

6 Results

The restored pitch contour is approximate. We examined a distortion with beam width M and N-best value N. Beam width is larger,the distortion is smaller. And with fixed beam width,one N-best value has minimum distortion. The method using unit component of command is the smallest distortion in our methods. These method improved the rate which accent phrase segmentation is detected. But all methods has a tendency that commands occure at time which is not adequate as prosodic command because the definition of distorton is based on mean squared error.