Analysis and Improvement of Auto-Correlation Pitch Extraction Algorithm based on Candidate Set

YongJin So, Jia Jia and LianHong Cai

Abstract The auto-correlation pitch extraction algorithm based on candidate set is widely adopted due to low complexity, strong noise immunity and high accuracy. However, there are still some discontinuous or unsmooth places in the pitch contours got by this algorithm. This paper analyses types and causes of error pitches in the results of such algorithm, and proposes detection methods and amendment strategies for each type of error. Using the pitch contour smoothing algorithm based on segmenting, 91.07 percent of error pitches in the result are eliminated.

Keywords pitch extraction · auto-correlation algorithm · pitch contour smoothing

1. Introduction

Pitch extraction is always a hot topic in the field of speech signal processing. A lot of pitch extraction methods have been proposed by many investigators [3][4][5]. But there are still some discontinuous or unsmooth places in the pitch contours got by this algorithm. The pitch extraction algorithm based on auto-correlation function is widely adopted in the field of speech technology, because its simplicity of implementation and robust noise immunity [1].Further, this algorithm is adopted in speech signal processing tool – Pratt, and through the continuous improvement and update, it can now achieve a much better pitch extraction result [2].

To resolve this problem, this paper first identified the different types of error pitches in the results of the auto-correlation pitch extraction algorithm based on candidate set, and analyzed the causes for each type of error. On this basis, the pitch contour smoothing algorithm based on segmenting is proposed. The experimental result shows that the proposed algorithm can effectively improve the accuracy rate of pitch extraction algorithm.

YongJin So, Jia Jia, LianHong Cai
Key Laboratory of Pervasive Computing, Ministry of Education,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science & Technology, Tsinghua University, Beijing, China

2. Auto-correlation pitch extraction algorithm based on candidate set

2.1 Introduction to the algorithm

This algorithm is derived from the classical auto-correlation method with the following improvements [1][2]:

1） F0-candidate set is constructed for each frame instead of using only the maximum point

2） A cost function is utilized to select the pitch period from F0-candidate set

In the classical autocorrelation method, the maximum point is selected from the autocorrelation coefficients for a frame, but in this algorithm, all of the peaks in the autocorrelation coefficients of a frame are gather to form a F0-candidate set, and a cost function is used to select the pitch period for each frame. The cost function is expressed with autocorrelation and F0-contour continuity to solve the problem of discontinuity.

2.2 Result Analysis of the algorithm

Next this paper analyzed the different types of errors and their causes and distribution in the results of auto-correlation algorithm based on candidate set.

1) error type and causes

This paper divides the errors in results of the algorithm into three categories.

(1) pitch in unvoiced segment

There are no pitches in unvoiced part of speech, but the algorithm sometimes extracts error pitches in unvoiced area. Usually, these error pitch segments are much higher than normal pitch contour and the durations are very short.

(2) singular points in voiced segment

The singular points can appear in anywhere of a continuous pitch contour. It has two causes. The first cause is the octave errors from the auto-correlation method. Though with candidate set most of the octave errors can be eliminated, there are still some left in the final result. This kind of singular point can appear in anywhere of a syllable.

The other cause is transition period errors. When a continuous pitch contour is composed of two or more voiced areas, transition pitch period may appear in boundary between voiced areas. These errors must appear in middle of a syllable.

(3) discontinuation of pitch contour inside a syllable

In theory, pitch contour of a syllable is continuous and smooth in Mandarin. A syllable in Mandarin is composed of zero or one unvoiced segment followed by one voiced segment. There would be no separated voiced segments inside a syllable.

But sometimes in the results of the auto-correlation pitch extraction algorithm based on candidate set, discontinuous pitch contours can be observed in a syllable.

2) error distribution

This paper selected 59 sentences for analysis. Among them, 12 sentences are recorded by one male speaker, and the other 47 sentences are from one female

speaker. One hundred and twelve errors were observed in the pitch extraction results.

The most obvious feature of first category errors and second category errors is that its duration is very short. The details of duration of first category and second category are shown as Table 1. It can be seen from Table 1 that the duration of error segments is distributed in from 0.01ms to 0.04ms. Another characteristic for the first category and second category is that the difference between error pitches and the real pitch contour is very large as shown in Table 2. It can be seen from Table 2 that the differences with error pitches and real pitch contour are distributed in from 50Hz to 500Hz.

Table 1 Duration distribution of first and second category

Duration range	Percent
0.01-0.1s	100%
0.01-0.06s	97%
0.01-0.05s	93%
0.01-0.04s	91%
0.01-0.03s	79%

Table 2 Distribution of difference between error pitches and real pitch contour

Pitch difference range	Percent
0-500Hz	100%
50-500Hz	98.15%
100-500Hz	75.93%

3. Pitch Contour Smoothing Algorithm based on Segmenting

As the improved auto-correlation method with candidate set improved the continuity of extracted pitch contour, the error points would also appear in short segments, which can also be seen in Table 1. Based on this pattern of error occurrence, a segment-based pitch contour smoothing algorithm is proposed. The basic flow of the smoothing algorithm is shown Fig. 1.

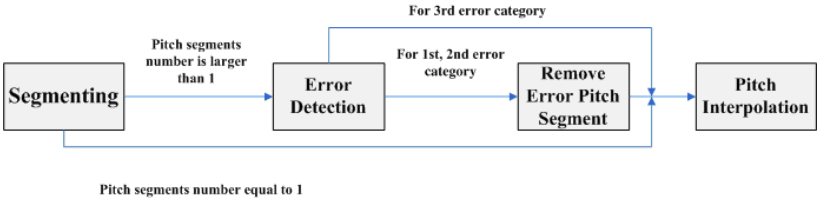

Fig. 1. Flow chart of segment-based pitch contour smoothing algorithm

3.1 Segmenting

As there exists great differences between error pitches and real pitches (as shown in Table 2), segment boundaries can be identified by great difference between adjacent pitch points.

More specifically, suppose the pitch sequence of a syllable is $P=\{P_1,P_2,P_3,\cdots P_N\}$. When the difference of two adjacent pitch points is above a certain threshold, that is, $|P_{i+1}-P_i|>f0\_TH$, the pitch sequence is divided into two segments : $\{P_1,\cdots P_i\}$ and $\{P_{i+1},\cdots P_N\}$. This process then goes on from P_{i+1}. The result of Table 3 shows that the differences between error pitches and real pitch contour are mainly distributed in 50-500Hz, so the threshold is defined as $f0\_TH$ =50Hz.

Use the above segmenting standard, the pitch contour of a syllable is divided into *k* pitch segments. If *k*=1, the error in the pitch contour is not in the first category or the second category.

3.2 Error detection and removal

1) Detection of first category error

This kind of error must be in the front of a pitch contour of a syllable, as only the front part of a Mandarin syllable can be unvoiced. Thus, the first pitch segment of every syllable will be checked to see where it is in voiced or not. If it is unvoiced, then it will be considered to be a first category error pitch segment.

The voiced ness of the 1st segment is determined using duration, energy and zero-crossing rate. Suppose the duration, average energy and zero-crossing rate of every pitch segments are $\{t_1,t_2,\cdots t_K\}$, $\{e_1,e_2,\cdots e_K\}$ and $\{c_1,c_2,\cdots c_K\}$. The 1st segment would be determined as voiceless if the following conditions are met:

$$t_1<UV\_M \qquad e_1<\frac{e_i}{VU\_B},\forall i,2\le i\le k \qquad c_1>c_i\times VU\_B,\forall i,2\le i\le k \tag{1}$$

where UV_M is duration threshold of first category error pitch segment, from the result of Table 2, it is defined as: $UV\_M=0.1s$.

VU_B is ratio threshold for energy and zero-crossing rate. By means of experiments, it is defined as: $VU\_B=1.2$. If the first pitch segment is voiceless, then it should be removed from the pitch sequence.

2) Detection of second category error

As the second category error occurs in the middle of a syllable, it can be identified by checking the difference between the average pitch of a pitch segment with the two adjacent pitch segments. A large enough difference would be treated as a singular point.

If octave error occurs at the front or the end of a syllable, it can be identified by that the average pitch of one pitch segment would be double of the adjacent pitch segment. As the duration of error pitch segment is very short, a shorter pitch segment in the two adjacent pitch segments is considered to be error pitch segment.

Suppose the average pitches of every segment are $\{f_1, f_2, \cdots f_K\}$. When the following conditions are met, the *i*-th pitch segment is considered to be error pitches.

when $i = 1$,

$$t_i < t_{i+1} \quad \& \quad (f_{i+1} \times MIN\_TH < f_i < f_{i+1} \times MAX\_TH \quad or \quad \frac{f_{i+1}}{MIN\_TH} > f_i > \frac{f_{i+1}}{MAX\_TH}) \tag{2}$$

when $i = K$,

$$t_i < t_{i-1} \quad \& \quad (f_{i-1} \times MIN\_TH < f_i < f_{i-1} \times MAX\_TH \quad or \quad \frac{f_{i-1}}{MIN\_TH} > f_i > \frac{f_{i-1}}{MAX\_TH}) \tag{3}$$

when $1 < i < K$,

$$(t_{i-1} < t_i < t_{i+1} \quad \& \quad f_{i-1} < f_i < f_{i-1}) \quad or \quad (t_{i-1} > t_i > t_{i+1} \quad \& \quad f_{i-1} > f_i > f_{i-1}) \tag{4}$$

where t_i is the duration of the *i*-th pitch segment, f_i is the average pitch of the i-th pitch segment, MIN_TH and MAX_TH are octave error threshold. By means of experiments, they are defined as: $MIN\_TH = 1.5$, $MAX\_TH = 2.25$ All pitch points in the detected error pitch segment should be removed. The pitch values of such segments would be interpolated using values in adjacent segments, which will be discussed in details in 3.3.

3) Detection of third category error

One syllable in Chinese has only one voiced segment, so when no-pitch frame appears between two voiced segments, it is the third category error. Pitches for no-pitch frames would be interpolated as described in section 3.3.

3.3 Pitch Interpolation

While first category error and second category errors are detected, these pitch segments would immediately be removed from pitch contour, so after error detection and removal, the pitch contour only has third category errors. The sinc interpolation method is used in this paper.

4. Experiment and Result analysis

To test the usability of the proposed method, it is applied on the result of the auto-correlation method with candidate set. The 57 sentences introduced in 2.2.2 are

used as tests. Errors before and after pitch contour smoothing for the 57 sentences are shown in Table 3, and the percent of error removed is 91.07%.

Table 3 Error numbers of after smoothing and before smoothing

Error type	Error reason	Before smoothing	After smoothing	Percent of error removing
First category	pitches in unvoiced segment	66	6	90.9%
Second category	singular points in voiced segment	34	4	88.24%
Third category	discontinuation of pitch contour	12	0	100%

The result shows that the pitch contour smoothing algorithm based on segmenting is effective, the smoothing algorithm resolves diverse problems of the auto-correlation pitch extraction algorithm based on candidate set, partly improves the accuracy of pitch extraction.

5. Conclusions

This paper introduced the auto-correlation pitch extraction algorithm based on candidate set, analyzed the three types of errors, and their causes and distributions. Based on the analysis, the pitch contour smoothing algorithm based on segmenting is proposed. The detection methods and the amendment strategies for each type of errors are described in this algorithm. Finally, objective experiments verified the effectiveness of the smoothing algorithm. The experimental result shows that the segmenting-based pitch contour smoothing algorithm is suitable for the auto-correlation pitch extraction algorithm based on candidate set, 91.07 percent of error pitch points in original result are eliminated.

References

1. Boersma P: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Institute of Phonetic Sciences, University of Amsterdam, Proceedings, 17, 97-110 (1993)
2. Boersma P., Weenink D: Praat: Doing phonetics by computer (Version 4.6.09) [Computer program]. Retrieved January 4, from www.praat.org (2009)
3. Y. Hu, N. Chen, X. Xu: Pitch Detection Using a Improved Algorithm Based on ACF. Electronic Science and Technology, 2, pp.25-28 [In Chinese] (2007)
4. K. Abdullah-Al-Mamun etc: A High Resolution Pitch Detection Algorithm Based on AMDF and ACF. Journal of Scientific Research, 1 (3) : 508-515 (2009)
5. S. L. Liang: ACF-CEF Pitch Detection Algorithm based on De-noising. China Science and Technology Information, Jun (12) [In Chinese] (2008)