ISCA Archive
http://www.isca-speech.org/archive

International Symposium on Chinese Spoken Language Processing (ISCSLP 2002)
Taipei, Taiwan
August 23-24, 2002

A PITCH DETECTION ALGORITHM BASED ON SPECIAL POINTS AND AREA

Li WANG

Department of Computer Science and Engineering,
Harbin Institute of Technology, Harbin
wangli@mtlab.hit.edu.cn

Xin Lv

Department of Computer Science and Engineering,
Harbin Institute of Technology, Harbin
lxfancy@mtlab.hit.edu.cn

Tie-Jun ZHAO

Department of Computer Science and Engineering,
Harbin Institute of Technology, Harbin

tjzhao@mtlab.hit.edu.cn

Zhan-Yi LIU

Department of Computer Science and Engineering,
Harbin Institute of Technology, Harbin

zhanyi@mtlab.hit.edu.cn

ABSTRACT

Pitch detection and estimation is a very important problem in speech signal processing. Now some scholar has presented a simple and effective method in pitch detection. It lessens the computing burden, but still has some defects for practical application. Here we improve this simple algorithm effectively, and introduce a method based on positive-negative area into it for pitch detection. Its good performance has been shown in text-to-speech system and speaker recognition.

1. INTRODUCTION

Pitch is the periodicity caused by the vibration of vocal chords when people pronounce, and pitch period is the reciprocal of the vibration frequency [3]. The period variation over short-time (intrinsic period variation) carries phonemic, linguistic and speaker related information. So in speech analysis, synthesis and recognition applications the pitch period is an important parameter. Generally there are three methods for pitch detection [3]: first, waveform estimation, includes parallel processing, data reduction and so on; second, correlation processing, includes waveform autocorrelation, AMDF, etc.; third, transformation method, for example, Cepstrum method.

Now some scholar has presented a simple and effective method [4] in pitch detection. The central thought is special points matching. It lessens the computing burden, but still has some defects for practical application. Here we improve this simple algorithm effectively, and introduce a method of positive-negative area into it for pitch detection. Its good performance has been shown in TTS (text-to-speech system) and SR (speaker recognition). As to the methods based on area, someone has once proposed a method named "The Function of Area Difference" [5]. It can be regarded as an improved algorithm based on AMDF [1], and belongs to the category of correlation processing. The method based on positive-negative area in this paper belongs to the category of waveform estimation. So they are different though they have some relationship.

2. PITCH DETECTION

2.1 Method of Special Points Matching

The special points matching method believes that [4] in the two successive periods in a periodic signal, their corresponding points should be suited (Generally they are not suited absolutely and have difference for some reasons). In fact we only need to verify that the special points (for example, the local peak-point) but not all the points in the two successive sets of points are suited, and then we can find a suited period. Thus the method for pitch detection is made up of two steps: first, searching for the sets of points for matching (here we call the set of points "window".); second, matching the two windows. If they are suited, then we have found two periods, or we have to search for the windows for matching again. This method makes separating unvoiced sound from voiced sound and detecting pitch proceed at the same time. And we don't need to segment the speech signal into frames. So the period detected is not the average period in a frame but an independent and accurate period. In addition this method lessens the computing burden, and we find it feasible through analyzing the oscillogram. So the algorithm in this paper adopts the thought above.

2.2 The Improvement of Special Points Method

The advantages of the special points matching method have been analyzed above. But there is a problem in it: First, two windows are found for matching. Then if some points in the two windows are not suited, the current point will be discarded and the next peak-point will be regarded as the current point, and the algorithm will begin to search for the point suited to the current point. Actually from the oscillogram we can find that in some cases there are two or more points whose amplitudes are similar in one period (this case is illustrated in Figure 1). And they are easy to be regarded as suited points. In this way two windows are found for matching: if they are suited in the algorithm, the matching is wrong; if they are not suited, then there will be some useless points in the successive periods (they are successive periods in a same vocal signal), just like some air bubbles in a solution with even density. So these useless points are called “air bubble” and this phenomenon is “air bubble phenomenon”.

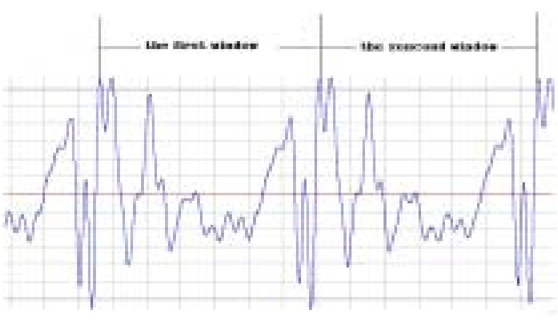

Figure 1: An amplificatory waveform fragment of “啊”

In order to eliminate these air bubbles we use automatic adjusting windows. That is to say, if the two windows are not suited, we are not eager to change the current point but adjust them according to the length of the windows. If the first window is shorter, from its end point we search for the suited point for its start point again. Because the start point of the second window is changed, the whole second window will be changed too; if the second window is shorter, then from its end point we search for the suited point for its start point again and make the first window unchanged. In this way the “air bubble phenomenon” is eliminated greatly.

In the subsequent step for matching-windows, the method of special points is very accurate, but sometimes the successive periods in a same vocal signal still have great local difference between each other. So if the matching rule is wider, when searching for the windows to match the algorithm is inclined to get wrong windows; if the accurate rate of searching for the windows was kept, then in the step of matching some correct periods would not be found. In order to solve this problem, we present the method of positive-negative area to improve it in this paper. That is in the step of matching-windows, if the method of special points fails, then the matching will be executed again using the method of positive-negative area. It has achieved fair results for practical application.

2.3 Method of Positive-Negative Area

After two windows have been found, the algorithm of special points begins to match the corresponding points. But because vocal signal has the approximate-periodicity, its periods are not suited absolutely and sometimes have great local difference between each other. For example, several corresponding points of them are not suited, but the two windows are suited on the whole. This case inspired us with a thought that we can match the windows on the whole. So in this paper, the method of positive-negative area is proposed. Its central thought is that the positive area and the negative area of the two periods are compared apart, and if they are all suited we can believe that the two periods are suited. Here the area is the closed area, which is embraced by the waveform curve and X-axis on the oscillogram. The formulas for computing the areas is following:

$$A_p = \sum_{k=1}^{p} | P_1(k) - P_2(k) | \qquad (1)$$

$$A_n = \sum_{k=1}^{n} | N_1(k) - N_2(k) | \qquad (2)$$

In the formulas, $P_1(k)$, $P_2(k)$, A_p are the kth positive area of the first period, the kth positive area of the second period, and the remainder of all the positive areas in the two periods; $N_1(k)$, $N_2(k)$, A_n are the kth negative area of the first period, the kth negative area of the second period, and the remainder of all the negative areas in the two periods; p and n are the number of the positive area and the number of the negative area in a period.

3. EXPERIMENTAL RESULTS

3.1 Matching rules

Not only the matching of the corresponding points but also the area matching, its matching rule (the allowable range of their remainder) is an experiential value. So it needs to be determined through experiments and is different in the different practical conditions. In this algorithm, the matching rule for two points is that the remainder of their amplitudes is less than 15% of the bigger one, and that for two areas is that the remainder of their values is less than 2% of the bigger one.

3.2 Success rate of pitch detection

The sampling rate of speech in this experiment is 32kHz. The author has implemented the algorithm of special points and the algorithm presented in this paper distinguishingly using C++ programming language. The experiments show that the algorithm here is better than the special points matching method by detecting the pitch of 2039 monosyllables and many dissyllable in our speech library. The success rate of pitch detection for monosyllables is over 92%, and that for dissyllable is over 80%.

Success rate of pitch detection = the number of periods detected / the number of periods in the syllable

Figure 2: Waveform of “啊”

In Table 1, the output results of the speech “啊” (shown in Figure 2) based on the algorithm of special points and the algorithm in this paper are shown. There are three columns of data each algorithm: the first is the flag by which we can judge the corresponding wave segment is a period or not, it is a period when the flag is “1” and it is surd or the useless points when the flag is “0”. Thus the “0”s appearing in the successive “1”s seem air bubbles, and the “air bubble phenomenon” is exhibited well. The second and the third column are the start and the end point of a wave segment. It is obvious that the algorithm presented in this paper has outstanding result, especially for eliminating the “air bubble phenomenon”.

Table 2: Success rate of pitch detection

	Special points method	Algorithm in this paper
啊 (a1)	80.0%	98.0%
夺 (duo2)	73.5%	92.5%
处 (chu3)	78.5%	95.0%
更 (geng4)	75.0%	96.0%
呼吸 (hu1xi1)	78.0%	85.2%
程序 (cheng2xu4)	74.0%	82.0%

Table 2 shows the success rate of pitch detection for some syllables based on the two algorithms. We can see that the success rate of “duo2” is lower in the monosyllables, and the primal reason is the periodicity is disturbed in the transitional period of ‘u’ and ‘o’. So it is very difficult to detect, and this is also an aspect needed to improve in the future.

4. CONCLUSIONS

In this paper, the method of special points is improved effectively by analyzing its advantages and disadvantages, and a method of positive-negative area is introduced into it for pitch detection. From the output results above we can find that the algorithm presented in this paper eliminates the “air bubble phenomenon” which appears in the algorithm of special points greatly, and its good performance has been shown in TTS and SR. But because of the approximate-periodicity of the speech signal, it is obvious that “none of them work perfectly for every voice, application and environmental condition” [2], and we still have vast space to research.

Table 1: The results of the two algorithms

Algorithm of special points	Algorithm in this paper
0,0,2493	0,0,2124
1,2494,2623	1,2125,2254
1,2623,2752	1,2254,2382
1,2752,2881	1,2382,2511
1,2881,3003	1,2511,2640
0,3004,3137	1,2640,2769
1,3138,3266	1,2769,2898
1,3266,3395	1,2898,3026
0,3396,3517	1,3026,3155
1,3518,3646	1,3155,3284
1,3646,3775	1,3284,3413
1,3775,3904	1,3413,3541
1,3904,4033	1,3541,3670
1,4033,4161	1,3670,3799
1,4161,4291	1,3799,3928
1,4291,4419	1,3928,4057
1,4419,4548	1,4057,4186
1,4548,4676	1,4186,4314
1,4676,4805	1,4314,4443
1,4805,4934	1,4443,4572
1,4934,5063	1,4572,4700
1,5063,5192	1,4700,4829
1,5192,5321	1,4829,4958
1,5321,5450	1,4958,5087
1,5450,5578	1,5087,5216
1,5578,5707	1,5216,5345
1,5707,5836	1,5345,5474
0,5837,5971	1,5474,5602
1,5972,6101	1,5602,5731
1,6101,6223	1,5731,5860
0,6224,6246	1,5860,5989
1,6247,6376	1,5989,6118
1,6376,6506	1,6118,6247
0,6507,6738	1,6247,6376
1,6739,6869	1,6376,6506
1,6869,6999	1,6506,6636
0,7000,7812	1,6636,6766
1,7813,7942	1,6766,6896
1,7942,8071	1,6896,7027
1,8071,8199	1,7027,7157
1,8199,8328	1,7157,7288
1,8328,8456	1,7288,7418
0,8457,10837	1,7418,7548
	1,7548,7678
	0,7679,7683
	1,7684,7813
	1,7813,7942
	1,7942,8071
	1,8071,8199
	0,8200,10837

5. REFERENCES

[1] L.R.Rabiner, R.W.Schafer. *Digital Processing of Speech Signals*, Prentice-Hall, Inc., 1978.

[2] W.Hess, *Pitch Determination of Speech Signal*, Springer-Verlag, Berlin, 1983.
[3] Hang HU. *Speech Signal Processing*, Harbin Institute of Technology Publishing House, 2000.
[4] Wei-Jun CHEN, Jian-Min LI, Fu-Zong LIN, Bo ZHANG. *A Chinese Text-to-Speech System*, Computer Engineering and Application, 2000.9
[5] Xue-Qing Xu, Tie-Cheng Yu. *The New Way Exploring in Pitch Detection*, Proc. NCMMSC5, 1998