A Time Domain Estimation Algorithm for Speech Signal Pitch Period

Shuxing Wu¹*, Tiansong Li² and Xiuqin Zhang¹

¹School of Computer and Communication, Beijing Information Technology College, Beijing, 100015, China
²School of Information and Communication, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China
*Corresponding author’s e-mail: wushuxing@163.com

Abstract. In speech recognition and speech synthesis, accurate estimation of the pitch period is an important part of speech processing. The traditional direct peak estimation method and the autocorrelation function method are both effective time domain estimation algorithms. The autocorrelation method is a pitch period estimation algorithm suitable for low SNR. Both algorithms need to get accurate peak position estimation. In this paper, a multi-line cut method which is a method for judging the position of the peak point is proposed. The multi-line cut method is used to intercept the sampled data of the waveform by using multiple cut lines. The median value is calculated by the starting and ending points of the cut line position, and the peak position is indirectly evaluated. By minimizing the impact of interference on the peak estimate, the likelihood of falling into local extreme points is reduced, therefore a more accurate peak point estimate than the direct search for peak points can be obtained. The simulation results show that compared with the traditional direct peak estimation method, the performance of peak estimation by the multi-line cut method can be greatly improved, and the multi-line cut method can be used to estimate the peak value in the autocorrelation method, and also achieve a certain performance improvement. In addition, the number of cut lines is directly related to performance, and the more the number is, the better the performance is. The complexity of this method is not high and easy to implement.

1. Introduction

In speech signal processing technology, the estimation of the pitch period is a very important link [1-2]. Pitch detection is widely used in speech analysis, speech synthesis, speech compression coding, speech recognition and speech segmentation [3]. For many years, researchers have proposed various pitch detection algorithms, such as Autocorrelation Function method(ACF)[4], Average Magnitude Difference Function (AMDF) [5-6], wavelet transform method[7-8], Cepstrum method, etc. In general, pitch period extraction methods are mainly time domain estimation methods and transform domain estimation methods [9].

The time domain estimation method is to estimate the pitch period directly from the waveform of the speech signal, and it has been applied very early, and it is widely used because of its simple implementation and low computational complexity. The peak direct estimation method is one of the time domain estimation methods and is still widely used at present. The autocorrelation function method is also a time domain estimation method, which is suitable for the pitch period extraction in the case of low SNR [4]. The autocorrelation function method needs to estimate the peak position when performing...
the pitch extraction. When the peak position is inaccurate due to the local minimum value, the performance is affected.

In this paper, a peak point position estimation method will be described, which can make the judgment of the peak point position more accurate, and relatively accurately estimate the pitch period of the speech signal. The following is a description of the traditional peak direct estimation method and the short-term autocorrelation function estimation method to estimate the pitch period. Then, the multi-line cut method proposed in this paper is introduced, and then the four methods are verified and evaluated.

2. Two time domain pitch period estimation methods
In the time domain pitch period estimation method, the traditional direct peak estimation method and the autocorrelation function method are both effective algorithms. Among them, the autocorrelation method is a pitch period estimation algorithm suitable for low SNR. Both algorithms require an accurate estimate of the peak position. The following is a brief introduction to the two algorithms.

2.1. Traditional peak direct estimation
The peak direct estimation method is to directly find two adjacent peak points of the periodic signal, and calculate the interval time $T$ between the two peak points, that is, the period of the signal. However, due to the influence of noise and interference, this method may lead to inaccurate peak point estimation, which results in inaccurate period estimation. However, this method is simple and intuitive, and has low complexity. There are still many application scenarios.

2.2. Short-term autocorrelation function method
The autocorrelation function method belongs to the time domain estimation algorithm. Compared with other time domain algorithms, it has better anti-noise interference characteristics. The extracted pitch contour features are obvious, the accuracy is good, the implementation is simple, and it is also a widely used algorithm in the field of speech signal processing.

The principle of the algorithm is that the autocorrelation function value of the speech signal will peak at an integer multiple of the pitch period, and the pitch period can be extracted to estimate the pitch period, as shown in figure 1.

Autocorrelation calculation is performed for each frame by the calculation formula of the short-time autocorrelation function.

$$acf(\tau) = \sum_{i=0}^{n-1-\tau} s(i)s(i+\tau)$$

(1)

Then, by searching for the extreme point, the position of the peak is detected, as shown by the point P of figure 1, thereby estimating the pitch period.

The autocorrelation function method is used to detect the pitch period. In addition to the autocorrelation module, it is necessary to perform signal processing such as data acquisition, framing, windowing and filtering in advance to achieve better results.

In signal processing, windowing usually uses a rectangular window and a Hamming window [10]. When selecting the window length, you need to consider the pitch period of the voice signal. Generally, 1-7 pitch periods should be included in one frame. At a sampling rate of 16 kHz, a window length of 10-20 ms is generally selected. After selecting the window function and window length, you also need to set the frame shift parameter to make a smooth transition between frames and frames to maintain continuity.

3. Multi-line cut method
Whether using the traditional peak estimation method or the autocorrelation function method to estimate the pitch period of the speech signal, the most important step is to accurately detect the position of the peak. The detection of the peak position is generally implemented directly by searching for the maximum point. The biggest advantage of this method is that it is simple and intuitive, but it is easy to fall into the local extreme point, thus affecting the final effect. Here, a method named the multi-line cut
A method is proposed for peak position detection. The basic idea is to evaluate the peak position by calculating the median value by performing multiple interceptions on most positions of the waveform according to the symmetry of the waveform, as shown in figure 2. The starting position A1 and the ending position B1 obtained by intersecting the uppermost line with the waveform, the median value of A1 and B1 is the estimation of the peak position P, and the A2 and B2 obtained by the intercept line in the middle, the A3 and B3 obtained by the lowest line can still obtain the estimation of the peak position P by taking the median value. The average value of the obtained median is obtained, and an accurate estimation of the peak position can be obtained. Due to the filtering effect of the multi-section line, this method can avoid the mistake of using the local extremum point as the peak point. Compared with the method of directly searching for the peak point, a more accurate peak point position can be obtained, thereby accurately estimating the pitch period of speech signal.

If the multi-line cut method proposed here is used for peak position detection, it is necessary to first make a rough estimate of the period corresponding to the current peak point to be estimated, and assume that the estimate is T. Generally, the estimated value is used to obtain the current peak position. Before the step the several periods have been obtained, by which T will be predicted, but the T value is inaccurate, and the exact position of the peak point cannot be derived from it. The previous peak point has been obtained by the decision method introduced here, which is known. Although we do not use the period T to estimate the peak point directly, we need to use it to assist in the peak point evaluation process. In this process, the period estimate T is used to determine the starting and ending point of the evaluation process, which is a vital value. Here the T mentioned is the T in the "Peak Point Evaluation Process" below.

To correctly determine the peak point position, we first determine the starting point of the peak point evaluation process. The usual operation is to use the end point of the peak point evaluation process as the starting point of the next peak point evaluation process. Whether the value reaches a pre-set value m, if it is reached, the search process stops, otherwise it is judged whether 5/4T sample points have been reached from the previous peak point to the currently sample point, if 5/4T samples have been reached, the search process stops, and the new peak point is evaluated. If the current sampling point has exceeded the new peak point by 3/8T, the current sampling point is used as the starting point of the next peak point evaluation process, otherwise it pushes forward to exceed the new peak point 3/8T position, which is used as the starting point of the next peak point evaluation process, which is also the end point of the current peak point evaluation process. As shown in figure 3, there are 5/4T sample points between point A and point E in the figure, and 3/8T sample points between the new peak point C and F. If m=4 is set, then search process will end at point D. If m=5 or m is greater than 5, the search process will end at
point E. After the search process is over, if point F is to the right of D and E (as shown in figure 3), then the point F is the starting point for the next peak point evaluation process.

After the starting point of the evaluation process is determined, it needs to be implemented by the following “peak point evaluation process”:

Step (1): the sampling value of the starting point is set to the threshold A (1), and the intersection value A (1) with the vertical axis, and the sectional line 1 parallel to the horizontal axis, that is, y = A (1), intersects the waveform at M (1) Point and N(1) point, M(1) point is also the starting point, and the position is recorded as ps(1). And the cut line 1 corresponds to the counter c(1), and the value of c(1) is 1, and the threshold A(2) corresponding to the next cut line 2 is set, and A(2) = A(1) + Δ(1). As shown in figure 4.

Step (2): if the cut-off completion counter value does not reach the number m and the interval from the previous peak point to the current sampling point does not reach 5/4T sampling points, then step (3) is performed, otherwise the search process stops and jumps to the step (5).

Step (3): assuming that the Nth cut line has been completed, the threshold corresponding to the N+1th cut line is A(N+1), A(N+1)=A(N)+ Δ (N). We can get the current sample point value. If the sample point value is greater than or equal to A(N+1), the intersection value with the vertical axis is A(N+1), and the line N+1 parallel to the horizontal axis is intersected with the waveform at M ( N+1) point and N(N+1) point, the position is recorded as ps(N+1), and the intercept line N+1 corresponds to the counter c(N+1), and the c(N+1) value is set to 0. And then update the threshold A(N+2) corresponding to the next intercept line N+2, and A(N+2)= A(N+1)+Δ(N+1), as shown in figure 5. The sample point value is compared with all existing cut lines. If the sample point is above the cut line, the corresponding line counter value is increased by 1, as shown in figure 5, if the sampling point is below the cut line and the corresponding cut line counter value is greater than 0, the position is recorded as the end point of the cut line, and the position is recorded as pe(N+1), the cut-off completion counter value is incremented by one and updated.
Step (4): repeat step (2) and step (3).

Step (5): evaluate the new peak point. If the current sampling point has exceeded the new peak point by 3/8T, then the current sampling point is used as the starting point of the next peak point evaluation process, otherwise it will advance to exceed the new peak point 3/8T position. The position is the starting point of the next peak point evaluation process and is the end point of this “peak point evaluation process”.

In step (5), the new peak point is estimated according to the recorded cut line position. The calculation method is that the starting point position and the ending point position of all the completed cut lines are summed and averaged, the result of which is the new peak point position, such as in figure 6, ps(4), ps(5), ps(6), ps(7), pe(4), pe(5), pe(6), pe(7), the eight values are obtained. And summing and taking the average and taking the integer is the new peak point position, which can also be obtained by other weighting methods.

In the specific implementation process, we can limit the maximum number of cut lines, and recycle the cut lines to save memory resources, but the algorithm complexity will be improved.

The above method of peak estimation by the multi-line cut method can be used for the improvement of the traditional direct peak estimation method, and can also be used as an improved algorithm for the autocorrelation peak estimation, and the complexity is not high.

Of course, low-pass filtering is required before the operation here, and after the voiced sound is determined, the pitch period estimation is performed by the algorithm.

4. Algorithm performance verification

In general, the speech signal emitted by humans is only within a certain range. The audio frequency range of adult males is between 70Hz and 200Hz. The audio frequency range of adult females is between 140Hz and 400Hz, and the frequency of infants can reach 600Hz. In order to verify the performance of the multi-line cut method proposed in the previous section, a test signal is generated here by simulating a change in the audio frequency of a woman. Using a signal similar to FM, the frequency range is changed from 140H to 400Hz in 0.15 seconds, and then is changed from 400Hz to 140Hz, and the amplitude is modulated. The sampling rate is 22050 samples per second, 320 sampling points per frame. The frame shift is set to 160 sample points, where 812 frames of speech data are generated for performance verification. The speech signal is recorded as s(t), and figure 7 plots the frequency value as a function of different frames.
The generated $s(t)$ signal is added to Gaussian white noise to form four test signals with signal to noise ratios of 5 dB, 10 dB, 15 dB and 20 dB, respectively. Add the signal after Gaussian white noise, and then perform frame division, windowing and filtering. Windowing uses the aforementioned rectangular window or Hamming window. The filter is implemented using an elliptical bandpass filter with a passband of 140 Hz to 400 Hz.

The filtered signal is used as the verification signal of the pitch period estimation algorithm, and the result of the algorithm's evaluation of the pitch period is obtained. Figure 8 shows the results of pitch estimation using a traditional direct peak method for a signal with a signal-to-noise ratio of 5 dB. From the figure, we see that the estimated result is significantly different from the pitch period of the verification signal. Figure 9 shows the results of peak estimation of a signal with a signal-to-noise ratio of 5 dB using the multi-line cut method. From the figure, we can see that the estimated result is better than the traditional direct peak method. Figure 10 is a result of estimating the pitch period by directly estimating the peak point to the signal with a signal-to-noise ratio of 5 dB using the autocorrelation method. From the figure, we can see that the estimated result is very close to the pitch period of the verification signal. Figure 11 shows the results of the pitch period estimation of the signal with a signal-to-noise ratio of 5 dB using autocorrelation + multi-line cut method. From the figure, we can see that the estimated result is very close to the pitch period of the verification signal, and compared with the autocorrelation method there is no significant difference.
Figure 11. Evaluation results of autocorrelation + multi-line cut method (SNR= 5dB)

It can be seen from the given figures 8-11 that the autocorrelation method and the autocorrelation method + multi-line cut method are significantly better than the traditional peak method and the multi-line cut method. However, the autocorrelation method and the autocorrelation method + multi-line cut method cannot directly judge the performance difference from the figures. And the traditional peak method and the multi-line cut method cannot directly see the performance advantages and disadvantages from the figures. In order to judge whether the multi-line cut method can bring more excellent performance, we use the mean square error to calculate the difference between the evaluation results obtained by various algorithms and the verification signal. The smaller the mean square error is, the better the performance is, and the closer the result is to its verification signal. Table 1 shows the mean square error of the results of the traditional peak method and the multi-line cut method with the verification signals of SNRs of the 5dB, 10dB, 15dB and 20dB. From the table, we can see the multi-line cut method is significantly better than the traditional direct peak method in various signal-to-noise ratios. Table 2 shows the mean square error of the results of the autocorrelation method and autocorrelation method + multi-line cut method with the verification signals of 5dB, 10dB, 15dB and 20dB, respectively. From the table, we can see that the autocorrelation method + multi-line cut method is slightly better than the separate autocorrelation method at different signal-to-noise ratios.

| Table 1. Mean square error values of direct peak method and multi-line cut method at different signal-to-noise ratios. |
|---------------------------------------------------------------|
|                  | 20dB | 15dB | 10dB | 5dB  |
| Direct peak method | 182.3| 247.5| 340.6| 545.6|
| Multi-line cut method | 103.4| 108.4| 157.6| 314.0|

| Table 2. Mean square error values of autocorrelation method and autocorrelation method + multi-line cut method at different signal-to-noise ratios. |
|---------------------------------------------------------------|
|                  | 20dB | 15dB | 10dB | 5dB  |
| Autocorrelation method | 40.6 | 41.9 | 43.1 | 78.8 |
| Autocorrelation method + multi-line cut method | 34.5 | 35.1 | 35.7 | 61.6 |

From the above evaluation results, we can see that the multi-line cut method has obvious advantages compared with the traditional peak detection. The autocorrelation method + multi-line cut method also has some improvement compared with the separate autocorrelation method. From the point of view of complexity, the multi-line cut method only performs corresponding search and calculation, so the complexity is not high and it is suitable for real-time processing. The autocorrelation method + multi-line cut method can be used in situations where hardware conditions are not limited.

In addition, we also carried out simulation evaluation on the respective performances of the number of different cut lines with the SNR of 8dB under the multi-line cut method and the autocorrelation
method + multi-line cut method. The results show that the performance is related to the number of cut lines, and the more the number of cut lines, the better the performance is, as shown in Table 3.

|                | 1   | 2   | 3   | 4   |
|----------------|-----|-----|-----|-----|
| Multi-line cut method | 40.6| 41.9| 43.1| 78.8|
| Autocorrelation method + multi-line cut method | 34.5| 35.1| 35.7| 61.6|

5. Conclusion
In this paper, a speech signal peak position detection algorithm called multi-line cut method is proposed, which can reduce the influence of interference on the speech signal in a certain extent, so as to obtain a more accurate estimation of the peak point position of speech signal. The algorithm can not only estimate the pitch period of speech signal, but also help the algorithms relying on the peak detection to obtain better performance, and the implementation complexity is not high, suitable for real-time environment applications.

Acknowledgments
This work was supported by the General Program of Science and Technology Development Project of Beijing Municipal Education Commission of China(KM201410857001)

References
[1] Swee, T.T., Salleh, S.H.S., Jamaludin, M.R.(2010) Speech pitch detection using short-time energy. In:International Conference on Computer and Communication Engineering(ICCCE). Kuala Lumpur. pp. 11-13.
[2] Rabiner, L., Cheng, M., Rosenberg, A., et al.(1976) A comparative performance study of several pitch detection algorithms. IEEE Transactions on Acoustics, Speech and Signal Processing., 24(5): 399-418.
[3] Zeng, L., Chen, L., Xiao, Q.(2010) Pitch period estimation base on voiced degree weighted sub-frame octave region dynamic programming. In: International Conference on Wireless Communications and Signal Processing (WCSP). Suzhou. pp. 1-5.
[4] Krubsack, D.A., Niederjohn, R.J.(1991) An autocorrelation pitch detector and voicing decision with confidence measures developed for noise-corrupted speech. IEEE Transactions on Signal Processing., 39(2):319-329.
[5] Ross, M., Shaffer, H., Cohen, A., et al.(1974) Average magnitude difference function pitch extractor. IEEE Transactions on Acoustics, Speech and Signal Processing., 22(5):353-362.
[6] Yang, Y., Zhang, H., Guo, X.(2011) A pitch tracking method mixing ACF & AMDF algorithms based on correlations. In: International Conference on Image Analysis and Signal Processing (IASP). Wuhan. pp. 553-556.
[7] Cai, R., Zhu, Y., Shi, S.(2010) A modified pitch detection method based on wavelet transform. In: Second International Conference on Multimedia and Information Technology (MMIT). Kaifeng. pp. 246-249.
[8] Kadambe, S., Boudreaux-Bartels, G.F.(1992) Application of the wavelet transform for pitch detection of speech signals. IEEE Transactions on Information Theory., 38(2):917-924.
[9] Hu, J., Xu, S., Chen, J.(2001) A modified pitch detection algorithm. IEEE Communications Letters., 5(2): 64-66.
[10] Zhao, X., O'Shaughnessy, D., Minh-Quang, N.(2007) A processing method for pitch smoothing based on autocorrelation and cepstral F0 detection approaches. In: International Symposium on Signals, Systems and Electronics(ISSSE). Montreal. pp. 59-62.