# Precise pitch detection in noise environments

Tianwei Li, Zhengyou Li, Qian Huang, Jiao Guo
Department of Navigation
Dalian Naval Academy
Dalian, China
E-mail:TianweiLee@sina.com

***Abstract*—It is still challenging to detect speech pitch precisely in various noise environments at low SNR (signal to noise ratio). Voiced speech harmonics can be clearly found in noisy speech. Some of the harmonic frequencies with higher energy were estimated by spectral tracking method, a pectinate filter was designed according to those harmonic frequencies to filter the noisy speech, and the ponicaré pitch detection method was used to detect pitch from the filtered speech. The error is less than 2Hz and accuracy ratio is more than 90%, which exceeds parallel methods' by 20%, under the experimental condition of 5 noisy speech and -10dB SNR. Experiments demonstrate the effectiveness of the pitch detection method in various noise environments at low-SNR.**

***Keywords—pitch detection; spectral tracking; pectinate filter; ponicaré pitch map***

## I. Introduction

Pitch detection is very important in voice print, speech coding, and emotion recognition, etc, especially for Chinese speech, which is tonal. There are many pitch detection methods, such as harmonic peek detection [1], correlation function [2], Average Magnitude Difference Function (AMDF) [3, 4], etc. These methods work effectively in pure speech or high-SNR circumstance. But it is hard to detect pitch precisely at low-SNR and in noisy environment, because of the diverse kinds of noises and complex variety of speech signals.

The voiced speech signal is obviously harmonic, and the harmonic frequencies are integral multiple of pitch frequency. These harmonic frequencies mean bright lines in spectrum figure. These lines are still bright at low-SNR and in noisy environment. In this paper, a precise pitch detection algorithm was proposed in various noise environments at low-SNR. First we improve a spectral tracking algorithm, which has powerful anti-noise ability, to depict the harmonic characteristic by some spectral lines. Then a pectinate filter was designed according to those spectral lines to filter the noisy speech, and a periodic signal was obtained which contains pitch frequency and harmonic frequencies. Last the ponicaré map algorithm was advanced to detect pitch precisely.

## II. Harmonic Peek Calculation

Spectral tracking algorithm is to depict the bright harmonic spectral lines in spectrum figure by some spectral lines (as shown in Fig.1.(c)) [5, 6]. First the noisy speech signal was cut into overlap frames, and aggravated. Then the spectral peeks were detected in every frame, and the adjacent peeks which have close frequency and amplification were lined. Last the lines which are longer and have more energy were selected, and the bright harmonic spectral lines in spectrum figure were depicted.

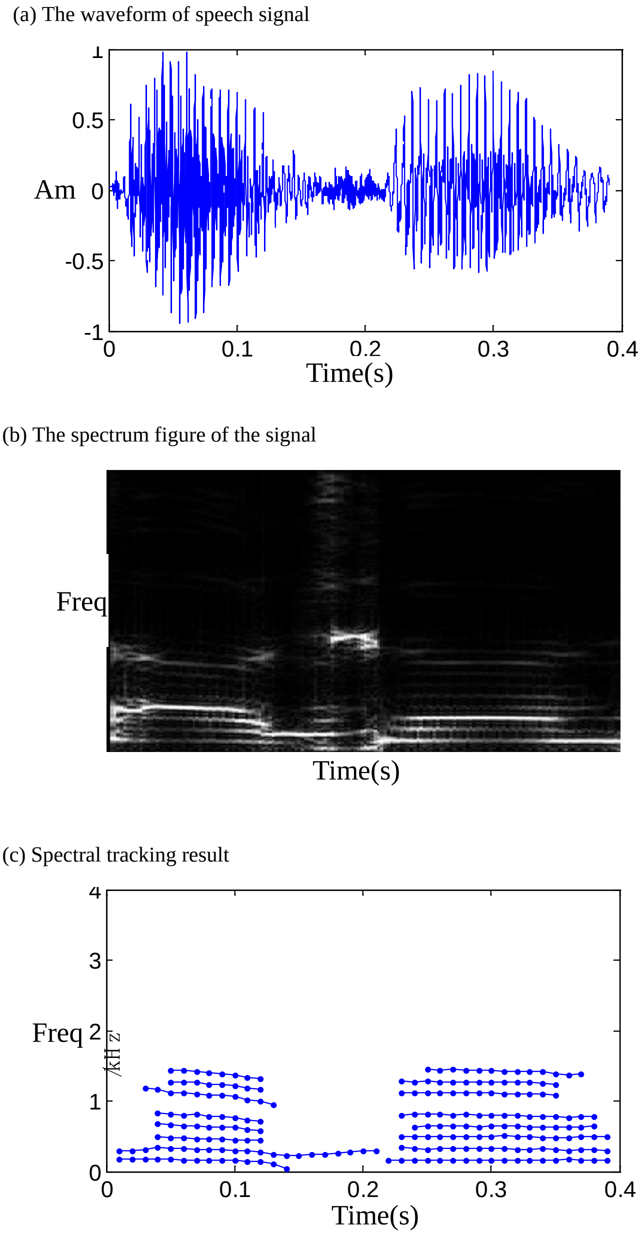



Fig.1. The process of spectral tracking

In noise environment, although most of the spectral tracking lines are harmonic, there are still some non-harmonic lines in spectral tracking result, which lead to wrong pitch detection. The non-harmonic lines can be checked out by the character that harmonic frequency intervals are equal. Because there are errors when calculate the frequency peeks, the frequency intervals are actually not firmly equal. The harmonic peek frequencies were calculated as follows:

- Get the frequency of all the spectral line peeks in every frame, and sort them by frequency.
- Calculate the frequency difference $\Delta f_i$ of two adjacent peeks. For each frequency difference $\Delta f_i$ , if $60 < \Delta f_i < 450$ and $\left|k \cdot \Delta f_i - \Delta f_j\right| < \Delta F$ ( $\Delta F$ is an experimental constant and means the maximal allowed error, $k = 1, \cdots, 5$ , and $\Delta f_j$ are the other all frequency differences), then the two adjacent peeks of $\Delta f_j$ are spare harmonic peeks related to $\Delta f_i$ . Last calculate the number $m_i$ of the spare harmonic peeks related to $\Delta f_i$ .
- Set $m_\gamma = \max_{i \in I}(m_i)$ , then all the spare harmonic peeks related to $\Delta f_\gamma$ are gross estimated harmonic peeks.
- The gross estimated harmonic peeks are not always harmonic peeks because the frequency of non-harmonic peeks is random. If there are three-quarter gross estimated harmonic peeks in a spectral line, we think that it is a harmonic spectral line. All the peeks in harmonic spectral lines are harmonic peeks.

Because the harmonic frequencies are integral multiple of pitch frequency, we may estimate the pitch by the harmonic peeks. But the Fourier transform was used when spectral tracking, and there are barrier effect and leak effect in Fourier transform, which lead to frequency estimate error. Estimate the pitch frequency by the harmonic peeks directly will have larger errors.

## III. Design of the Pectinate Filter

The pectinate filter is actually a group of band-pass filter. In each frame, each frequency of harmonic peek is the central frequency of a band-pass filter. The pectinate filter was obtained by adding the whole band-pass filter. The noise in speech signal was suppressed by filtering it. And the harmonic part, which has more energy, was kept in filtered speech.

We use the Gabor filter as the band-pass filter, because it performs better both in time domain and frequency domain. The discrete Gabor filter [6] is

$$h(n) = \exp(-b^2 n^2)\cos(\Omega_c n) \qquad (1)$$

$$b = \sqrt{2\pi}\ f_b / f_s \qquad (2)$$

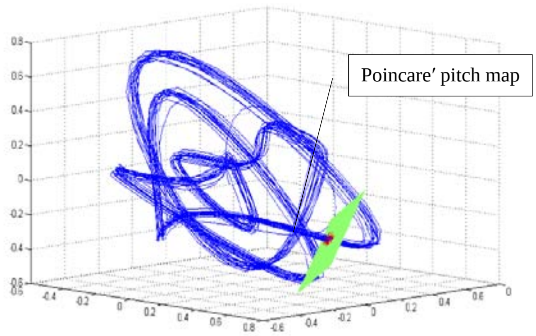
The pectinate filter was obtained as follows:

- Calculate $b$ by the signal sampling frequency $f_s$ and the band width of Gabor filter $f_c$ .
- Calculate the filter length $N$ by $N = \left\lfloor \frac{\sqrt{5\ln 10}}{b} \right\rfloor$.
- For each harmonic peek frequency $f_c^k$ , calculate the filter angle frequency $\Omega_c^k$ by $\Omega_c^k = 2\pi f_c^k / f_s$ .
- Calculate every band-pass Gabor filter $h_d^k(n)$ , and the pectinate filter $h(n) = \sum_{k=1}^{K} h_d^k(n)$ .

## IV. Poincare′ Pitch Detection

Voiced speech signal is a group of quasi-periodic curves in delayed coordinate reconstruct state space (as shown in Fig.2.(a)). If we choose a point in the curve and make a vertical plane at the point, the time interval of the curve drilling through the plane twice is the pitch period. Poincare′ pitch detection algorithm [7, 8] can locate the period point directly at the time domain waveform (as shown in Fig.2.(b)).

(a) Speech signal curve in state space and poincare′ pitch map




(b) Time domain waveform and period points

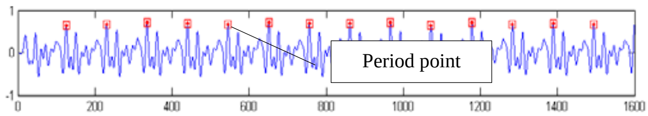
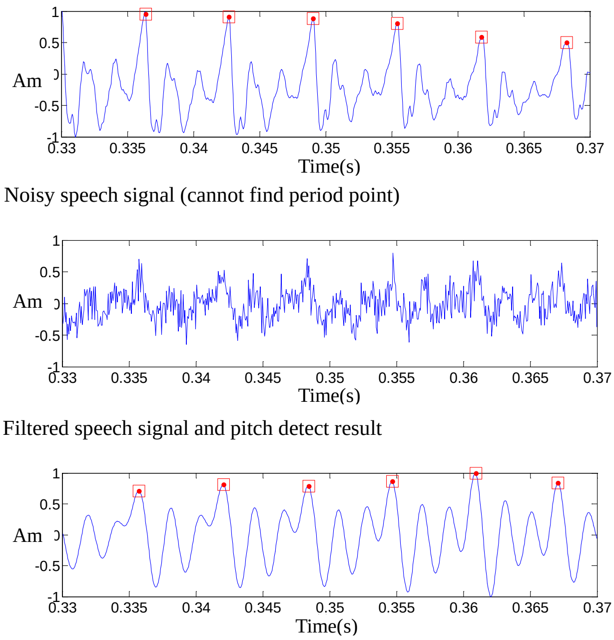



Fig.2. Poincare′ pitch detection

Poincare′ pitch detection algorithm performances well in the period signal, but not satisfying when there are noises in speech signal. The pure speech signal can be detected by poincare′ pitch detection algorithm accurately (as shown in Fig.3.(a)). When the signal was add to 0dB white noise (as shown in Fig.3.(b)), the poincare′ pitch detection algorithm can not detect the pitch. We filtered the noisy speech signal using the pectinate filter designed in this paper, then the poincare′ pitch detection algorithm can detect performance good (as shown in Fig.3.(c)).




(a) Pure speech signal and pitch detect result

(b) Noisy speech signal (cannot find period point)

(c) Filtered speech signal and pitch detect result

Fig.3. Example of pitch detection

## V. EXPERIMENTS

The method was evaluated on Gross Pitch Error (GPE) and Fine Pitch Error (FPE) proposed by Rabiner in [9]. GPE is the percent of frames which frequency error doesn't exceed 10Hz. FPE is the mean error frequency when error doesn't exceed 10Hz. GPE reveals the accuracy of the algorithm, and FPE shows the precision of the algorithm.

We obtain noise speech by adding pure speech and various kinds of and SNR noises. The pure speech was first detected by poincare′ pitch detection algorithm, and the period points were marked on waveform curves. The results of pitch detection were datum pitch frequency after the period points were affirmed artificially.

The added noises are white noises and noises in building site, steer house, dockyard and market. White noises were adopted in many pitch detection evaluation. Noises in building site burst out continually. Noises in steer house have many kinds of noises. Noises in dockyard appear continually and alterably. Noises in market contain many human voices. White noises were produced by software, other noises were recorded in certain environments, and pure speech signals are four person's speech.

The proposed pitch detection method (STPM) was tested in 5 kinds of noises at 3 kinds of SNR, and was compared to the Average Magnitude Difference Function (AMDF) [3, 4] and Neural Network (NN) [10] based pitch detection method. The results is shown in Table1 and Table2. The precision of pitch detection isn’t meaningful when the accuracy is low. So FPE is calculated only when GPE is less than 30%.

TABLE I GPE(%)

| | NN | | | AMDF | | | STPM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10dB | 0dB | -10dB | 10dB | 0dB | -10dB | 10dB | 0dB | -10dB |
| white noises | 3.38 | 3.21 | 59.1 | 2.12 | 4.95 | 47.4 | 1.46 | 3.18 | 11.5 |
| noises in building site | 5.33 | 8.28 | 35.7 | 3.57 | 9.84 | 33.6 | 1.56 | 3.22 | 8.63 |
| noises in steer house | 5.36 | 17.4 | 39.2 | 4.73 | 18.2 | 37.8 | 1.93 | 3.15 | 9.52 |
| noises in dockyard | 4.84 | 22.2 | 42.8 | 5.37 | 19.9 | 36.1 | 3.25 | 4.63 | 7.84 |
| noises in market | 13.8 | 52.2 | 82.3 | 16.3 | 47.4 | 88.6 | 6.36 | 45.7 | 89.1 |

TABLE II FPE (Hz)

| | NN | | | AMDF | | | STPM | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10dB | 0dB | -10dB | 10dB | 0dB | -10dB | 10dB | 0dB | -10dB |
| white noises | 0.34 | 1.47 | - | 0.69 | 1.82 | - | 0.05 | 0.81 | 1.45 |
| noises in building site | 0.36 | 2.54 | - | 0.76 | 3.56 | - | 0.28 | 0.53 | 1.15 |
| noises in steer house | 0.92 | 3.25 | - | 0.84 | 4.72 | - | 0.24 | 0.45 | 0.96 |
| noises in dockyard | 0.54 | 4.21 | - | 0.67 | 4.53 | - | 0.31 | 0.56 | 0.88 |
| noises in market | 0.87 | - | - | 0.83 | - | - | 0.52 | - | - |

Table I shows that the GPE of the three methods are all small when the SNR is 10dB except to the noise speech in market, which means the accuracy of three methods are all satisfying. The proposed method perform better than the other two methods when SNR decreasing. The GPE of the proposed method is less than 10% when the SNR is -10dB, however the GPE of the other two methods exceeds 30%. The three methods perform unsatisfying in market noises.

Table II shows that the FPE of the three methods are all small when the SNR is 10dB. The proposed method perform better than the other two methods when SNR decreasing. The FPE of proposed method is less than 2Hz even when the SNR is -10dB.

The proposed method is better than the other two methods in four environments, except for market. The market noises mainly contain other speech signals, and the three cannot distinguish just one speech from other speeches. So the three methods all failed to detect pitch.

## VI. CONCLUSION

The proposed method can detect pitch precisely in various noise environments at low SNR by combining the spectral tracking algorithm (which can detect pitch in various noise environments at low SNR) and the poincare′ pitch detection algorithm (which can detect pitch precisely). Although it cannot detect certain person's pitch in signals containing many people’s speech, the proposed method is still better than traditional methods in accuracy and precision.

## References

[1] C. Deshmukh, A. Wilson, T. Salomon, “Use of temporal information: Detection of periodicity and pitch in speech,” IEEE Trans. on Speech and Audio Processing, vol. 13, 2013, pp. 776-786.

[2] X. F. Zhao, D. Shaughnessy, M. Nguyen, “A processing method for pitch smoothing based on autocorrelation and cepstral F0 detection approaches,” International Symposium on Signals, Systems and Electronics, Montreal, 2007, pp. 59-62.

[3] W. Y. Zhang, G. Xu, Y. G. Wang, “Circular AMDF and pitch estimation based on it,” Acta Electronica Sinica, vol. 31, 2003, pp. 886-890.

[4] K. Bojan, H. Harald, K. Zdravko, “Noise robust F0 determination and epoch-marking algorithms,” Signal Processing, vol. 89, 2013, pp. 2555-2569.

[5] Z. Y. Li, T. W. Li, Q. Huang, “Voiced Chinese speech detection in noise environments,” Acta Acoustics, vol. 39, 2014, pp. 517-522.

[6] T. Taniguchi, T. Mikio, S. Katsuhiko, “Detection of speech and music based on spectral tracking,” Speech Communication,vol. 50, 2008, pp. 547-563.

[7] M. Hagmuller, G. Kubin, “Poincare′ pitch marks,” Speech Communication, vol. 48, 2006, pp. 1650-1665.

[8] Z. Y. Li, T. W. Li, Q. Huang, “Research on improved Poincare′ pitch detection method,” Applied Acoustics, vol. 31, 2012, pp. 379-386.

[9] L. Rabiner, M. Cheng, E. Rosenberg, “A comparative performance study of several pitch detection algorithms,” Proceedings of the first ICASSP, Philadelphia: IEEE Press, 1976, pp. 399-417.

[10] K. Han, D. L. Wang, “Neural Network Based Pitch Tracking in Very Noisy Speech,” IEEE/ACM Trans. On Audio, Speech, and Signal Processing, vol. 22, 2014, pp. 2158-2167.