IRT filtering법에 의한 기본 주파수 추출

배 경 진　　안 수 길
서울대학교 대학원 전자공학과

Inverse Rate Type Filtering Method for the Fundamental Frequency Extraction

Kyungjin BAE　　Souguil ANN
Seoul National University

1. Introduction

Since vocoder theory was inverstigated by Dudley in 1939, many studies on the data compression of speech signals have been developed. One of them is the source coding method which proceeds the speech production model by dividing it into the excitation part and the filtering part.

This method can compress the data more than any other wave coding methods. To characterize the excitation part in the source coding method such as LPC vocoder, the pitch extraction is needed.

Available pitch extraction methods until now can be essentially divided into two types as follow:

*Direct Detection Methods : Parallel Processing Method, Average Magnitude Difference Function Area Comparison Method[4], etc.

*Post Detection Methods : Simplified Inverse Filter Tracking method[4], Center Clipping Autocorrelation Function Method, etc.

The latter is to find the pitch period after elimination of the interference of the resonance characteristics of the vocal tract to the fundamental frequency, and the former is to find the pitch period from the unmodified speech samples. In both methods we generally eliminate the high frequency formants by a preceding low pass filter stage. The formants adjacent to the fundamental frequency, however, are not eliminated. Thus, the direct detection methods in general take rather complex decision algorithm so that the halving or the doubling of the pitch which are caused by the existance of the major formants, could be avoided.

On the other hand, the post detection methods estimate continuously the major formants by an inverse filter, and remove them so that a formant free pitch wave could be obtained. It is accounted for the fact that the post detection methods take a long computation time and do not adapt to higher pitch speakers like children because of the limited number of filter oredr.

In this paper, we proposed an inverse rated filter whose gain is inversely proportional to the frequency, instead of the ordinary low pass filter with a sharp cutoff characteristic. The inverse rated filter is realized by associating filters with ramp function characteristics with different cutoff frequencies:The former could be approximated by a sinc squared function. The resulting impulse response is a triangle wave shape in time domain. The number of computation steps for the present speech sample by this method is ten additions and nine subtraction which is considerably short.

2. Delivation of the characteristics of the LPF

Considering a voiced speech in frequency domain,we observe that the components of the fundamental frequency is the origin of the fine structure in the resonance curve of the vocal tract. The lowest frequency of those peaks in the curve is called as the first formant, F1. Since the energy of the first formant is higher than the others, the effect of the first formant is obser-

ved clearly in the time domain, and since the first formant is most adjacent to the fundamental frequency, it can strongly influence the precision in extracting the fundamdntal frequency.

As the range of the fundamental frequency being generally 40-400Hz, most pitch extractors before the application of the algorithm, carry out the the pre-filtering, that is, low pass filtering in order to eliminate the effect of formants. However, the first formant being generally close to the fundamental frequency, can not be eliminated by the ordinary LPF.

It is necessary to estimate the first formant and substract it in order to reduce the interference from the original speech wave. For this purpose the pitch extractor uses a coefficient calcultor for the vocal tract parameter calculation and the obtained parameters are applied to the inverse filter for the elimination of the vocal tract characteristics, and then the periodic component (which is the pitch) is detected from the residual signal[6](published as the simplified inverse Filter Tracking, SIFT, method by Markel).

Since the computation time depends on the number of the filter order. We preprocess the signal by a LPF with the cutoff frequency of 900Hz in order to eliminate the higher frequency formants and be able to use a simple inverse filter, and by consequence simple calculation. But as this method is available upto no more than 250Hz pitch frequency. It is impossible to apply this method to high pitch speakers namely women and chileren. Also the calculation process is still too complicated for a microcomputer to finish the calculation in real time.

To resolve those difficulties it will be better to excute the calculations of the preprocessing filter and the inverse filter at the same time. The spectrum of a typical voiced speech show in fig. 1 demonstrates the known fact that the first formant is higher in frequency than the fundamental frequency and there is almost no correlation between them. we have every interest to suppress the first formant in comparison to the fundamental frequency. We have devised an inverse rated filter with the characteristic shown in fig. 2-a in which the gain is inversely proportional to the frequency.

As the first formant frequencies generally are higher than 250Hz, we have adjusted the centre point of the inverse rated filter around 230Hz, And as the frequency range of the pitch being between 40Hz and 400Hz, the cutoff frequency of the filter is fixed to 460Hz. To avoid the boosting of DC component, the gain under 40Hz is kept constant.

The inverse rated filter, we propose, is the combination of three ramp filters of which the characteristics are shown in the fig 2-a. These ramp characteristics can be approximated by $\sin^2$ (.) functions as shown in fig 2-b. To utilize the linear part of the $\sin^2$ (.), the $3/4\pi$ -point of the variable must be adjusted to the cutoff

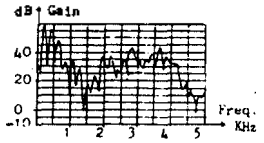

Fig. 1 Spectrum of the voiced speech "i".

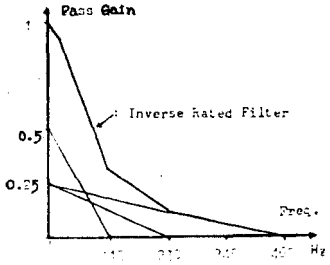

Fig.2-a Modelling LPF for the pitch Extraction

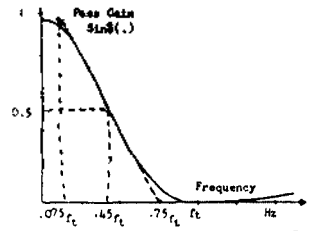

Fig. 2-b Equivalent Function to the Ramp Function : $\sin^2(.)$

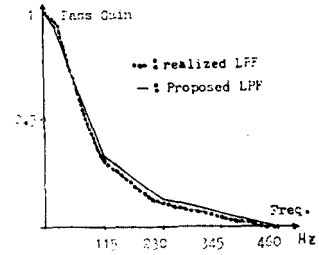

Fig. 2-c Comparison the proposed LPF with the realized LPF

frequency. The combined characteristics of three filters with each different $\prod$-point is shown in the fig. 2-c with the ideal inverse rated filter which is in the fig. 2-a.

3. LPF Algorithm

The Sinc^2 (.) Function characteristics shown in the fig. 2-b has a triangular wave response in the time domain as indicated in the fig. 3. The unit in the time domain is equivalent to 1/fa=8KHz. Assuming that the $\prod$-point adjusted to 615Hz, the zero gain point is N=13.

The area under the triangular wave is calculated as follow:

$$
\begin{aligned}
&H(n-N-1)\\
&=\sum_{k=0}^{N-1}S(n-1-k)\\
&=\sum_{k=0}^{N-1}S(n-1-k)-\sum_{k=0}^{N-1}S(n-k-1)+\sum_{k=0}^{N-1}S(n-k-N-1)\\
&=H(n-N)-\{S(n-1)+S(n-N)\}+\{S(n-N)-S(n-2N)\}\\
&=H(n-N)-\{A(n)-S(N)+S(n-N)\}+\{A(n-N)-S(n-N)+S(n-2N)\}\\
&=H(n-N)-A(n-1)+A(n-N-1)\quad\cdots\cdots\cdots\cdots\ 1
\end{aligned}
$$

Here the function H(n-N) is the area of the signal in the foregoing sample time, and A(n-N) and A(n-N-1) are the contents of the temporary addition buffer. If we add -S(n) and S(n-N) to A(n), it will give A(n-1); -S(n-N) and S(n-2N) to A(n-N), A(n-N-1).

Thus, to calculate the area for the present sample value, we calculate A(n-1) and A(n-N-1). Those values are subtracted 2nd added respectively from and to the triangle area which was calculated one sample time earlier. The result is that we need only three additions or substractions for each sample to obtain the value for the next step.

To realize the inverse rated filter of fig. 2-a we have composed three parallel branches as shown in fig. 4 with N-value 13, 26, and 52 each in expression 1. The $\prod$-point in the sinc^2(.)function which is the cutoff frequency and the gain will vary in accordance with the N-value.

We havd used an eight bit A-D converter. When we have taken 52 as N-value the gain became 2704 ($=N\langle 2^{12}$). By consequence to obtain the sinc^2(.) characteristics, the operation of 20(8+12)bits will suffice. And it enables the calculation by Finite word Length(FWL) calculation, thus, with an ordinary microcomputer. And before to combine the 20 bit values which are calculated, we divide them by 2^7=128 which will enable us to pursue the LPF calculation with 16-bit arithmetics.

Finally, we conclude that for the processing of the human speech signal it is sufficient with ten additions and nine subtractions with the proposed method.

4. Experimentation and results.

By processing the speech signal with the proposed LPF, the glottal part is emphasized. Inspecting it visually we see that the first positive pulse in a pitch period is intensified.

As the area under the positively launched glottal wave increases the contrast of the area of the first positive curve and those of the followings. The area, A(n2) under the first positive curve in the time domain is calculated as follows:

$$A(n2)=\sum_{i=N+1}^{2N}H(i-N-1)*S(i-N-1)\quad\cdots\cdots\ 2$$

Here the positive area A(n2) exists in the dom-

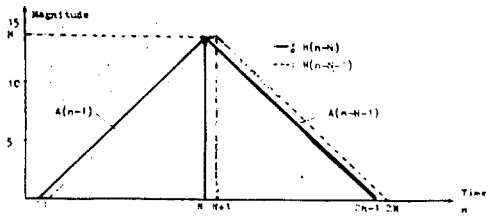

Fig.3 Impulse Response for the Sinc^2(.)Spectrum.

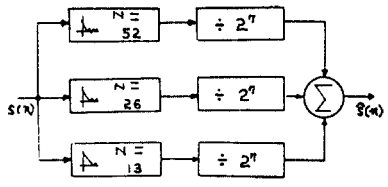

Fig.4 Realization of the proposed LPF.

ain $n1 \leq i \leq n2$, and $\{H(i-N-1)*S(i-N-1)\}$ is the processed value for i'th sample.

To find the true positive peak which is the beginning of pitch period, the threshold level, Ath(i-1) must be fixed as follows;

$$Ath(i-1)=3/4\ A(i) \quad \ldots\ldots\ldots\ 3$$

A(i) is the positive area value calculated in the true positive peak.

In the fig. 5 the processed signal and extraced pitch of a masculine voice "May I Help You?" are shown.

5. Conclusion

We have proposed an inverse rated type LPF which has the characteristics of inversely proportional gain to the frequency and which we can use in stead of ordinary sharp cutoff LPF when processing a voiced speech signal. Thus, the noxious major formants proximate to the pitch are attenuated to have almost the effect of the post pitch detection method without the complexity.

As this method is implemented, in the time domain with three triangular wave shapes, a calcu-

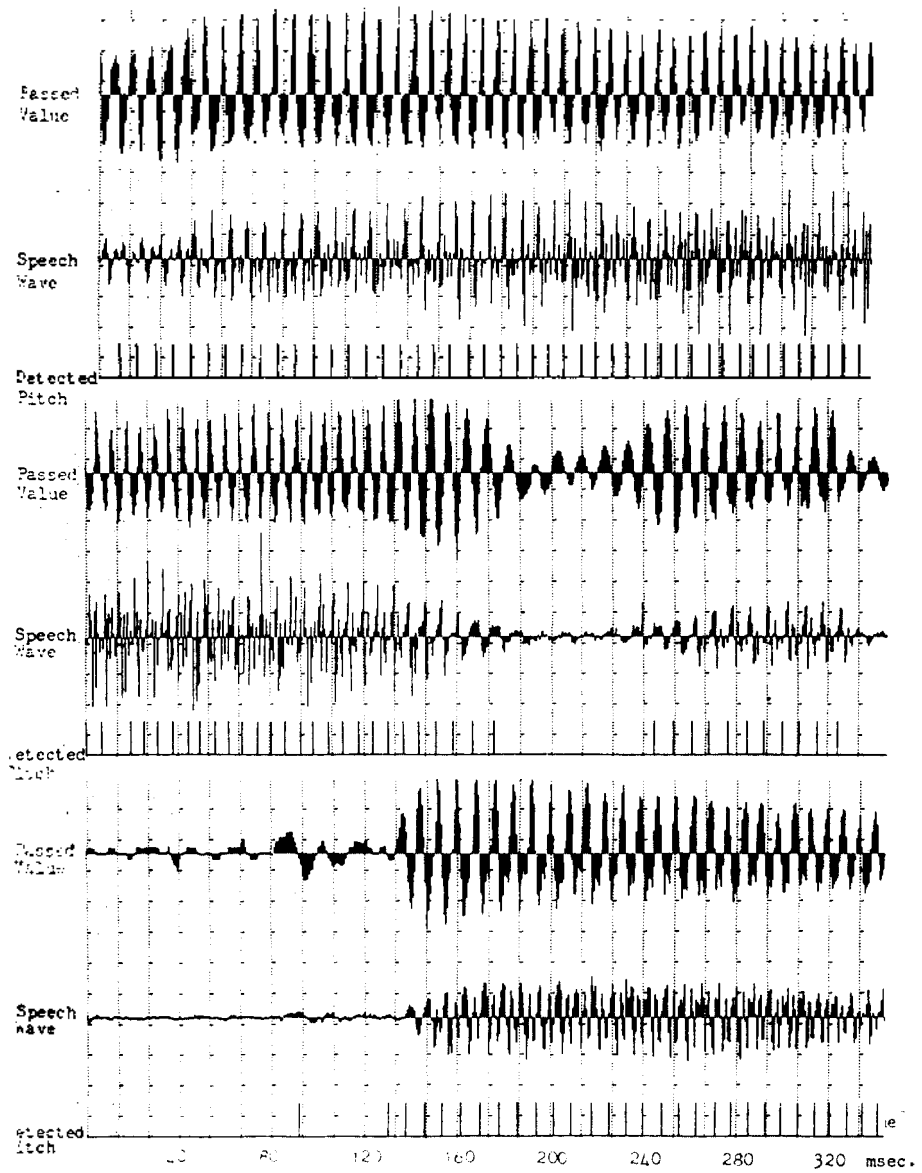

Fig. 5 Result for speech "May I Help You".

lation of ten additions and nine subetractions to have the desired response for the sampled speech signal. This process can be treated with a finite word length integer arithmetics and requires only addition and subtraction which enables a real time process with an ordinary microprocessors.

The modification of this method to adapt for the case of high pitch speakers or other special case is facilitated by adjusting the cutoff frequencies. This extends the applicable range of the process. As this method does the double process of LPF, filtering and major formants reduction, required process time is considerably short and the period calculation is simpler as the periodicity is ameriorated by the filtering; Both permit the real time pitch calculation with a microprocessor.

6. References

1) L.R.Rabiner and R.W,Schafer, "Digital Processing of Speech Signals",Prentice-Hall, Elglewood Cliffs, New Jersey, 1978.

2) J. D. Markel and A.H.Gray, "Linear Prediction of Speech", Springer-Verlag,Berlin,1976.

3) Myungjin BAE,"A Study on the Fundamental frequency Extraction of Speech Signals using Second Order Rundown Method", Seoul National University, MA Paper, Jan. 1983.

4) Myungjin BAE and Souguil ANN, "The High Speed pitch Extraction of Speech Signals using the Area Comparison Method", KIEE, Vol. 22, No.2,pp. 101-105, Feb. 1985.

5) Myungjin BAE and Souguil ANN,"The Voiced-Unvoiced-Silence Classification by Emphasized Spectrum of Speech Signals", JASK, Vol. 4, No.1, pp.9-16, June 1985.

6) J,D.Markel,"The SIFT Algorithm for Fundamental Frequency Estimation", IEEE Trans. on Audio and Electroacoustics, Vol. Au-20, No. 5, PP367-377, December. 1972.