PITCH EXTRACTION USING AN APPROXIMATELY IDEAL LOW-PASS FILTER

T. Matsuoka, A. Sugama, E. Onodera and Y. Ishida

Department of Electronics and Communication
Meiji University
1-1-1, Higashi-mita, Tama-ku
Kawasaki 214
Japan

ABSTRACT Although an ideal low-pass filter is not physically realizable, it can be approximated on the basis of time reversal techniques. In this paper, we describe a method to approximately implement the ideal low-pass filter and apply it to the pitch extraction system. Experimental results show that our method is effective to estimate the fundamental frequency of the speech signal.

1. INTRODUCTION

The fundamental frequency (pitch frequency) of the speech signal is one of necessary parameters in speech signal processing such as speaker recognition, speech recognition, speech coding and so on[1]. However, the accurate extraction method of the fundamental frequency has not been established yet.
In this paper, we propose a new method of estimating the fundamental frequency using an approximately ideal low-pass filter. The low-pass filter described here is approximately realized on the basis of*time reversal techniques*[2], [3]. This filter consists of causal and non causal ones. The filtered output is analyzed using the low-order LPC method and then the fundamental frequency is estimated by solving the roots of the polynomial obtained. The cutoff frequency of this filter is automatically tuned with neural networks so as to follow the fundamental frequency of the speaker.

2. DESIGN OF AN IDEAL LOW-PASS FILTER BASED ON TIME REVERSAL TECHNIQUES[2],[3]

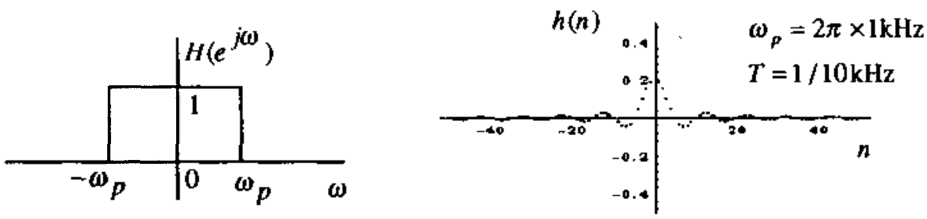
When the Fourier transform of an ideal low-pass filter is defined as

$$H(e^{j\omega}) = 1 \quad \text{for } |\omega| \le \omega_p$$
$$\qquad = 0 \quad \text{for } \omega_p < |\omega| \le \pi/T \tag{1}$$

its impulse response is equal to

$$h(n) = \frac{\sin n\omega_p T}{n\pi} \quad n = 0, \pm 1, \pm 2, \cdots \tag{2}$$

and is shown in Fig. 1(b).



(a)Frequency response (b)Impulse response

Fig. 1 An ideal low-pass filter.

Since $h(n) \neq 0$ for $n < 0$, an ideal low-pass filter is not causal and physically realizable. In order to approximately realize such a filter, we first divide the impulse response into two parts as follows:

$$h(n) = h_p(n) + h_m(n) \tag{3}$$

where

$$h_p(n) = 0 \qquad \text{for } n < 0$$
$$= \omega_p T / 2\pi \qquad \text{for } n = 0$$
$$= \frac{\sin n\omega_p T}{n\pi} \qquad \text{for } n > 0$$

and

$$h_m(n) = 0 \qquad \text{for } n > 0$$
$$= \omega_p T / 2\pi \qquad \text{for } n = 0$$

$$= \frac{\sin n\omega_p T}{n\pi} \qquad \text{for } n < 0$$

It is obvious that $h_p(n)$ is causal and $h_m(n)$ is not causal. The relation between the transfer functions of causal and non-causal filters is as follows:

$$H_p(z) = H_m(z^{-1}) \tag{4}$$

Therefore, the ideal low-pass filter can be block-diagrammed as in Fig. 2. The realizable IIR digital filter $H_p(z)$ can be simply designed using the Prony method[3],[4] and an eighth-order IIR digital filter (cutoff frequency=200Hz) is given by

$$H_p(z) = \frac{0.5z^8 - 2.033z^7 + 3.292z^6 - 1.96z^5 - 1.29z^4 + 3.125z^3 - 2.366z^2 + 0.8725z - 0.1327}{z^8 - 5.938z^7 + 16.18z^6 - 26.21z^5 + 27.5z^4 - 19.07z^3 + 8.523z^2 - 2.241z + 0.2653} \tag{5}$$

Fig. 3 shows the frequency response of the low-pass filter designed by the proposed method. On the other hand, from Eq. (4) and Fig. 2, it is clear that the low-pass filter is zero phase.

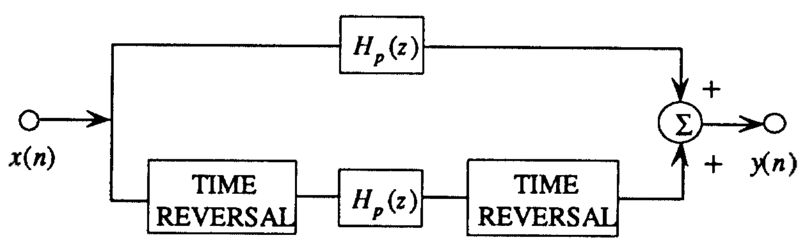


Fig. 2 Realization of the low-pass filter.

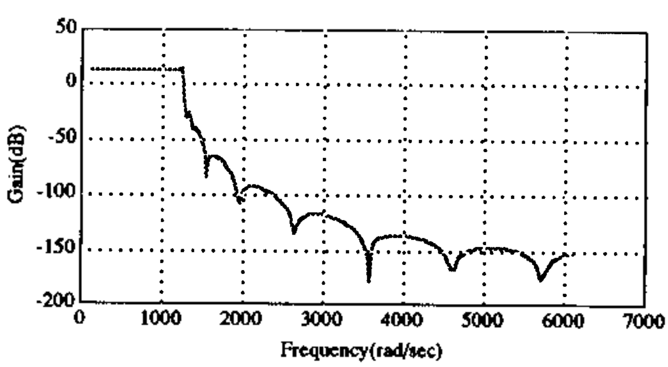


Fig. 3 Frequency response of the low-pass filter.

3. PITCH EXTRACTION BY AN APPROXIMATELY IDEAL LOW-PASS FILTER

Fig. 4 shows the pitch extraction system using an approximately ideal low-pass filter. In this system, the speech signal is sampled at 10kHz by using a 12bits A/D converter, and then the sampling rate is reduced to 2 kHz by a decimation process. The decimated output is then segmented in frames of 32 ms. Each frame is filtered by the low-pass filter and a fourth-order LPC analysis is performed on 64 data samples. The fundamental frequency can be estimated by solving the roots of the polynomial. However, this method may extract other poles rather that corresponding to the fundamental frequency of the speech signal. The fundamental frequency is determined by selecting the pole with a larger Q value from redundant analyzed poles.
On the other hand, the cutoff frequency of the low-pass filter is automatically tuned with neural networks to follow the fundamental frequency of the speaker. A roughly estimated value of the fundamental frequency calculated by the cepstral method is given to the input layer of the networks and then filter coefficients obtained in the output layers are transferred to the low-pass filter to control its cutoff frequency.

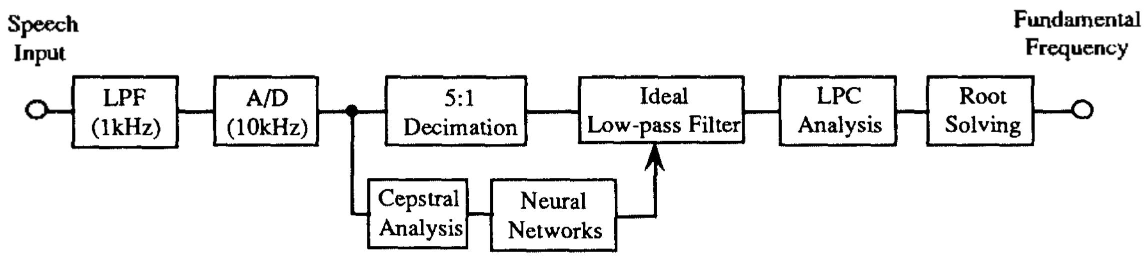


Fig. 4 Structure of the extraction system for the fundamental frequency.

4. NEURAL NETWORKS FOR CONTROLLING THE CUTOFF FREQUENCY OF THE LOW-PASS FILTER

Fig. 5 shows a part of neural networks for controlling the cutoff frequency of the low-pass filter. Networks are trained as follows:
Let the transfer function in Eq. (5) be described by

$$H_p(z) = 0.5\frac{\prod_{j=1}^{4}(z-z_j)(z-z_j^*)}{\prod_{i=1}^{4}(z-p_i)(z-p_i^*)} \tag{6}$$

where $\{p_i\}$ and $\{z_j\}$ are pole and zero, respectively. Each location of poles and zeros is non-linear for changes of the cutoff frequency. For this reason we use neural networks to control the cutoff frequency. As poles and zeros are complex values described by

$$\begin{aligned} p_i &= \mathrm{Re}\{p_i\} + j\,\mathrm{Im}\{p_i\} \\ z_j &= \mathrm{Re}\{z_j\} + j\,\mathrm{Im}\{z_j\} \end{aligned} \tag{7}$$

we give real and imaginary parts of poles and zeros pre-calculated by using the Prony method as target values of the networks. The networks are trained using the standard back-propagation algorithm[5]. Therefore, after training of the networks, we can get the filter coefficients corresponding to each cutoff frequency.

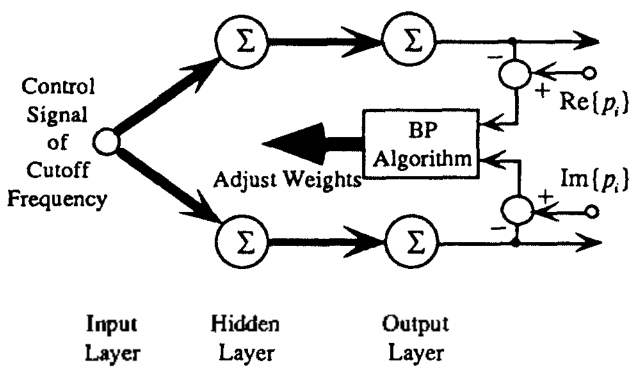


Fig. 5 Neural networks.

5. EXPERIMENTAL RESULTS

Fig. 6 shows an extraction result of the fundamental frequency component for a vowel /a/. We can see that high-frequency components are completely reduced. Fig. 7 illustrates an example of pitch extraction for the utterance /o-ha-yoo/ (this means "Good morning" in English). It is shown that our method is superior to the conventional cepstral method.

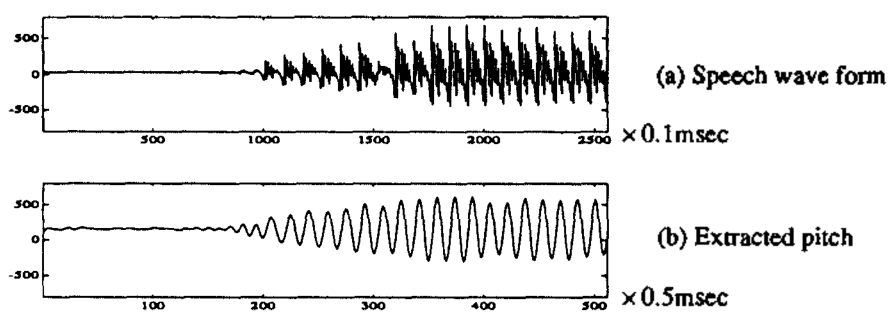


Fig. 6 Extraction of the fundamental frequency for a vowel /a/.

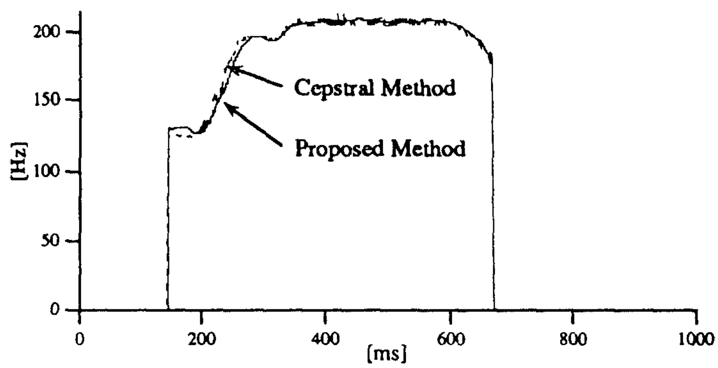


Fig. 7 Example of fundamental frequency extraction.

6. CONCLUSIONS

We have presented a new method of estimating the fundamental frequency of the real speech. The method proposed in this paper uses an approximately ideal low-pass filter which can be realized by using time reversal techniques. From experiments, it is shown that the our method is useful to the pitch extraction.

ACKNOWLEDGMENT

This work is partially supported by SMC Co., Ltd. in Japan.

REFERENCES

1. L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978 .
2. L. R. Rabiner and B. Gold, Theory and Application of Digital Signal Processing, Prentice-Hall, 1975.
3. A. Hiroi and Y. Ishida, Linear Phase IIR Hilbert Transformers Using Time Reversal Techniques, Trans. IEICE, 1994 (in press).
4. T. W. Parks and C. S. Burrus, Digital Filter Design, John Wiley and Sons, 1987.
5. D. E. Rumelhart, G. E. Hinton and R. J. Williams, Parallel Distributed Processing, The MIT Press, 1986.