PITCH ESTIMATION USING DISCRETE ANALYTIC SIGNALS

T. Matsuoka, N. Hayakawa, Y. Yashiba, Y. Ishida, T. Honda and Y. Ogawa

Department of Electronics and Communication
Meiji University

ABSTRACT - This paper proposes a new method for estimating the fundamental frequency of speech signal which uses the low-pass filter and Hilbert transformer with approximately ideal frequency responses.

INTRODUCTION

As it is well known, the discrete Hilbert transform can be used for calculating discrete analytic signals. However, the Hilbert transformer with ideal frequency responses is not physically realizable because it is non-causal. In order to approximately implement such a Hilbert transformer, Schüssler[1] and Ishida[2] proposed the design method using time reversal techniques. This method first divides the impulse response of the ideal Hilbert transformer into two parts, i.e., causal and non-causal parts and then approximately realizes it by the cascade connection of the causal and non-causal filters using time reversal techniques.

In this paper, we propose a new method of estimating the fundamental frequency using the Hilbert transformer and the low-pass filter based on the method described above. Experimental results show that our method is effective to estimate the fundamental frequency of speech signal.

DESIGN OF THE HILBERT TRANSFORMER BASED ON TIME REVERSAL TECHNIQUES[1],[2]

A Hilbert transformer is a linear time-invariant system whose ideal frequency response is defined as

$$H(e^{j\omega}) = \begin{cases} -j & 0 < \omega < \pi \\ 0 & \omega = 0,\ \pi \\ +j & -\pi < \omega < 0\ . \end{cases} \qquad \cdots(1)$$

The corresponding impulse response is given by

$$h(n) = \begin{cases} \dfrac{2\sin^2(\pi n/2)}{\pi n} & n \neq 0 \\ 0 & n = 0\ . \end{cases} \qquad \cdots(2)$$

Since $h(n) \neq 0$ for $n < 0$ (for n odd), an ideal Hilbert transformer is not causal and physically realizable. In order to approximately realize such a filter, we first divide the impulse response into two parts, i.e.,

$$h(n) = h_p(n) + h_m(n) \qquad \cdots(3)$$

where $h_p(n)$ is causal and $h_m(n)$ is not causal. Using the z-transform, we get

$$H_p(z) = -H_m(z^{-1})\ . \qquad \cdots(4)$$

Considering causality, the Hilbert transformer can be block-diagrammed as Figure 1. The subfilter $H_p(z)$ is a realizable filter and the boxes labeled TIME REVERSAL have input-output relations of the form

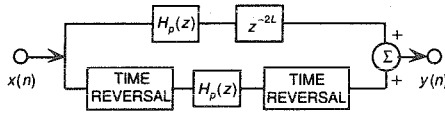


Figure 1. Block diagram of the Hilbert transformer

$$z(n) = w(-n) \qquad \cdots(5)$$

where $w(n)$ is the input sequence and $z(n)$ is the output sequence.

PITCH EXTRACTION BY A LOW-PASS FILTER AND A HILBERT TRANSFORMER

Figure 2 shows the pitch extraction system using a low-pass filter and a Hilbert transformer with approximately ideal frequency responses. In this system, the speech signal is sampled at 10kHz by using a 12 bits A/D converter, and then the sampling rate is reduced to 2 kHz by a decimation process. The decimated output is segmented in frames of 32 ms and is filtered by the low-pass filter, which can be designed by the same method as that used for the Hilbert transformer. The cut-off frequency of LPF is automatically tuned with neural networks so as to follow the fundamental frequency of speaker. The filtered output is then transferred to the Hilbert transformer.

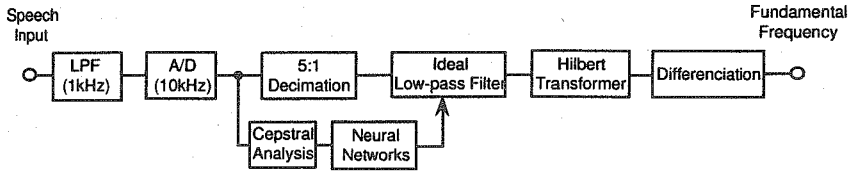


Figure 2. Structure of the extraction system for the fundamental frequency

The Hilbert transformer can be used for calculating instantaneous attributes of a time series, in particular the amplitude and frequency[3]. The instantaneous amplitude is the amplitude of the complex Hilbert transform. On the other hand, the instantaneous frequency is the time rate of change of the instantaneous phase angle. These two signals are obtained from the analytic signal. The analytic signal $a_x(t)$ of a real signal $x(t)$ is defined as:

$$a_x(t) = x(t) + jh_x(t) \,. \qquad \cdots(6)$$

Being $h_x(t)$ the Hilbert transform of $x(t)$

$$h_x(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(t')}{(t-t')}\,dt' \,. \qquad \cdots(7)$$

Then, the analytic signal can be expressed in modulus-argument form:

$$a_x(t) = e_x(t)\exp(j\phi_x(t)) \,. \qquad \cdots(8)$$

The amplitude and instantaneous phase of the signal $x(t)$ are

$$e_x^2(t) = |a_x(t)|^2 = x^2(t) + h_x^2(t)\,, \qquad \cdots(9a)$$

$$\phi_x(t) = \tan^{-1}(h_x(t)/x(t)) = \text{phase of } a_x(t)\,. \qquad \cdots(9b)$$

The instantaneous frequency is then given by

$$f_x(t) = \frac{1}{2\pi}\frac{d\phi_x(t)}{dt}\,. \qquad \cdots(10)$$

Based on this principle, the fundamental frequency of the speech signal can be obtained by differentiatng the angle trajectory of the analytic signal with respect to time.

NEURAL NETWORKS FOR CONTROLLING THE CUT-OFF FREQUENCY OF LOW-PASS FILTER

Figure 3 shows a part of neural networks for controlling the cut-off frequency of LPF. Networks are trained as follows:

Let the transfer function of an approximately ideal LPF in Eq.(4) be described by

$$H_p(z) = 0.5\frac{\prod_l (z - z_l)(z - z_l^*)}{\prod_l (z - p_l)(z - p_l^*)} \qquad \cdots(11)$$

where $\{p_i\}$ and $\{z_j\}$ are pole and zero respectively, and in case of LPF $H_p(z) = H_m(z^{-1})$. Each location of poles and zeros is non-linear for changes of the cut-off frequency. For this reason we use neural networks to control the cut-off frequency. As poles and zeros are complex values described by

$$p_i = \mathrm{Re}\{p_i\} + j\mathrm{Im}\{p_i\} \qquad \cdots(12a)$$

$$z_j = \mathrm{Re}\{z_j\} + j\mathrm{Im}\{z_j\}\,, \qquad \cdots(12b)$$

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. Therefore, after learning we can get filter coefficients corresponding to each cut-off frequency.

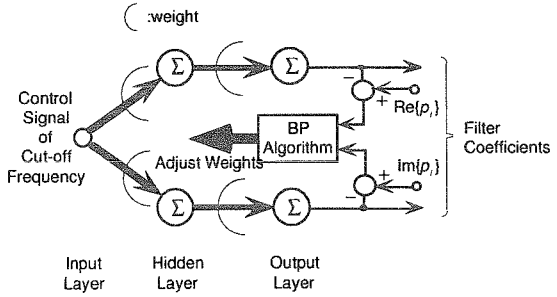
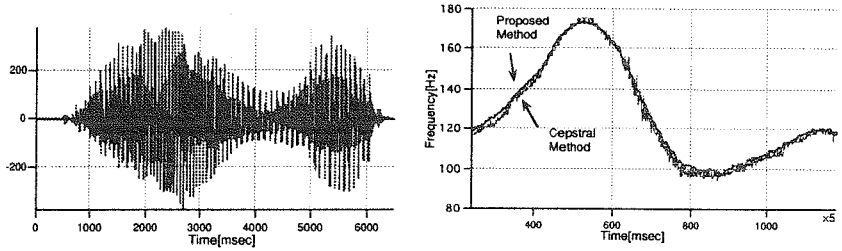


Figure 3. Neural networks for controlling the cut-off frequency of LPF

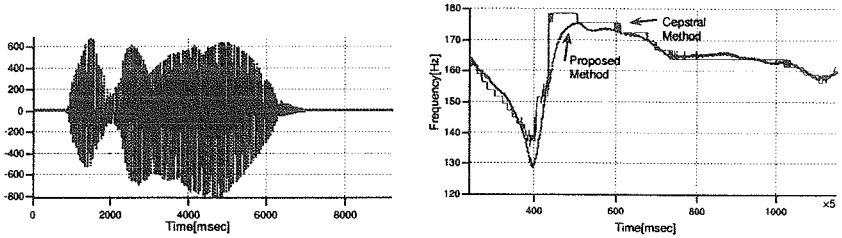
EXPERIMENTAL RESULTS

Figure 4 shows an example of the extraction of fundamental frequency for the utterance */aoie/* (this means "blue picture" in English). Figure 5 illustrates the fundamental frequency of */ohayoo/* ("good morning").



(a)speech signal (b)fundamental frequency

Figure 4. Extraction of the fundamental frequency for the utterance */aoie/*



(a)speech signal (b)fundamental frequency

Figure 5. Extraction of the fundamental frequency for the utterance */ohayoo/*

CONCLUSIONS

We have proposed a new method of estimating the fundamental frequency of the real speech. The method in this paper uses the analytic signal obtained by the Hilbert transformer. From some experiments, it is shown that our method is useful to the pitch extraction.

ACKNOWLEDGMENT

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

REFERENCES

[1]H. W. Schüssler, and J. Weitch (1987) *On the Design of Recursive Hilbert Transformers*, Proc. ICASSP-87, 876-879.

[2]A. Hiroi, H. Kamata and Y. Ishida (1994) *Linear Phase IIR Hilbert Transformers Using Time Reversal Techniques*, Trans. IEICE, 864-867.

[3]J. N. Little and L. Shure (1992) *Signal Processing Toolbox User's Guide*, (The MATH WORKS Inc.).