ISSPA 92, Signal Processing and
Its Applications.
Gold Coast, Australia, August 16-21, 1992.
Editor: D Gray

A Method of Speech Signal Analysis by Zero-Crossings

T Arai
Y Yoshida
Faculty of Science and Technology
Sophia University, Tokyo, JAPAN

We have studied zero-crossings of speech signal because it can be easily implemented in hardware. The information lost by signal limiting are recovered by our proposal method of "spectrum-reversal zero-crossings". We also present the results of a vowel recognition experiment using this method combined with neural networks.

1. INTRODUCTION

We have studied a zero-crossing method of speech signal processing, because it can be easily implemented in hardware. Although it is known that the hard limited signal has certain aural intelligibility [1], it has been found by other researchers that the zero-crossing rate of a speech signal is related to the frequency of the formant. The information in the hard limited signal are not sufficient for good speech recognition in many other applications. The information lost by limiting is recovered by our proposal method of "spectrum-reversal zero-crossings." In practice, the spectrum-reversal is approximated by a linear interpolation of the π radian spectrum-shifted signal. The zero-crossing wave of the spectrum reversal signal is used to supplement the information from the original zero-crossing wave. We study the theory, and implement the hardware I/O to get the informations of the zero-crossings. Vowel recognitions are examined.

2. ZEROS OF A SPEECH SIGNAL

We have modeled speech as an analytic signal which can be represented by its envelope and phase. These provide information about the pattern of the complex zeros of the analytic signal[3]. For a periodic, band-limited real signal $s(t)$, its analytic signal $m(z)$, where $z = e^{j\omega\zeta}, \zeta = t + j\sigma$, is called the *entire function* [2] and is expressed by its zeros z_i $(i = 1, 2, \cdots, N)$ in a factored form,

$$m(z) = c_0 \prod_{i=1}^{N} (1 - z_i^{-1} z), \tag{1}$$

where c_0 is a real constant. A spectrum reversal signal $\overline{m}(z)$ is written in terms of the same zeros z_i,

$$\overline{m}(z) = c_0 \prod_{i=1}^{N} (z - 1/z_i^*). \tag{2}$$

Because the zero-crossings of $s(t)$ are correspond to the zeros of $m(z)$ on the interior of the unit circle in the $z-$plane, we considered that the information lost by clipping is preserved by the zero-crossings of $\bar{s}(t) = \text{Re}\{\overline{m}(t)\}$. Figs. 1 and 2 shows an example for the Japanese vowel /a/, where $\bar{s}(t)$ is approximated by a linear interpolation of a π radian spectrum-shifted signal.

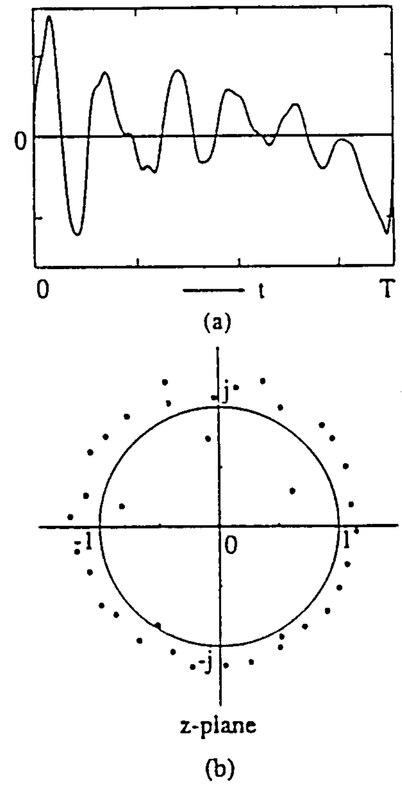

Fig. 1 Original signal of /a/, (a)waveform (b)zeros.

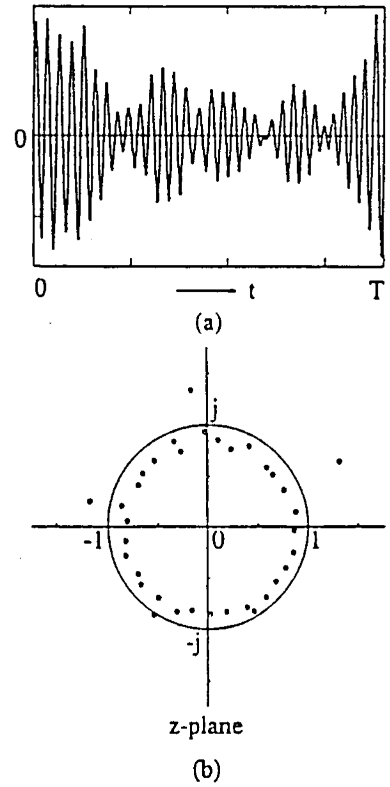

Fig. 2 Spectrum-reversal signal of /a/, (a)waveform (b)zeros.

3. VOWEL RECOGNITION

In a sample period we obtained a certain amount of aural intelligibility from a restored signal derived from a set of one-bit each for a zero-crossing waveform of original signals and one-bit each for spectrum reversal zero-crossing position[2,3]. Fig. 3 shows our recognition system. We use hardware autocorrelators to process these two types of zero-crossing information[4]. The system uses the autocorrelation functions of two sliced signals.

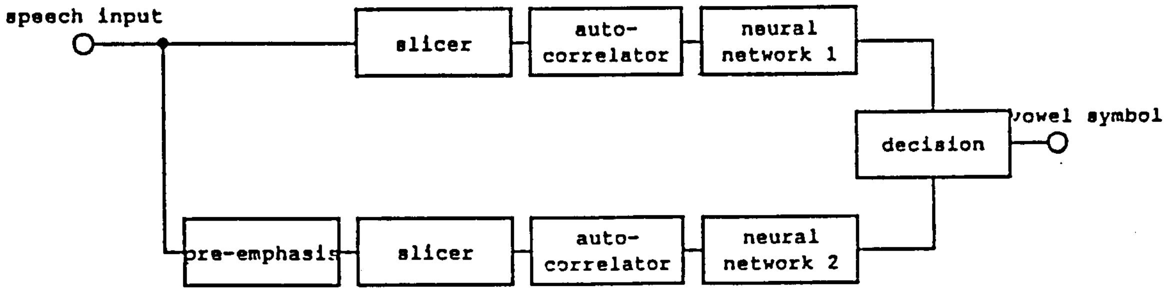

Fig. 3 Block diagram of our recognition system. Pre-emphasis effects as spectrum-reversal.

The input to one zero-crossing-autocorrelator is the raw speech waveform, and that to the other passes first through a high frequency enhancement filter. Each of the 32 samples of the two autocorrelations are introduced directly to the neural networks. We found that the output from neural network 1 distinguishes the five Japanese vowels by the first formant into three groups, {/i/, /u/}, {/e/, /o/} and {/a/}, and the output from neural network 2 distinguishes the five Japanese vowels by the second formant into two groups, {/a/, /o/, /u/} and {/i/, /e/}. As shown in Fig. 4, these two outputs are logically combined to complete the recognition. Table 1 shows the conditions for this experiment. Fig. 5 shows the output of vowels obtained when a speaker repeatedly enunciated five Japanese vowel sounds (/a,i,u,e,o/) per second over a two second period.

Table 1 Conditions for the Experiment.

sampling frequency	10 kHz
frame length	25.6 ms
frame period	25.6 ms

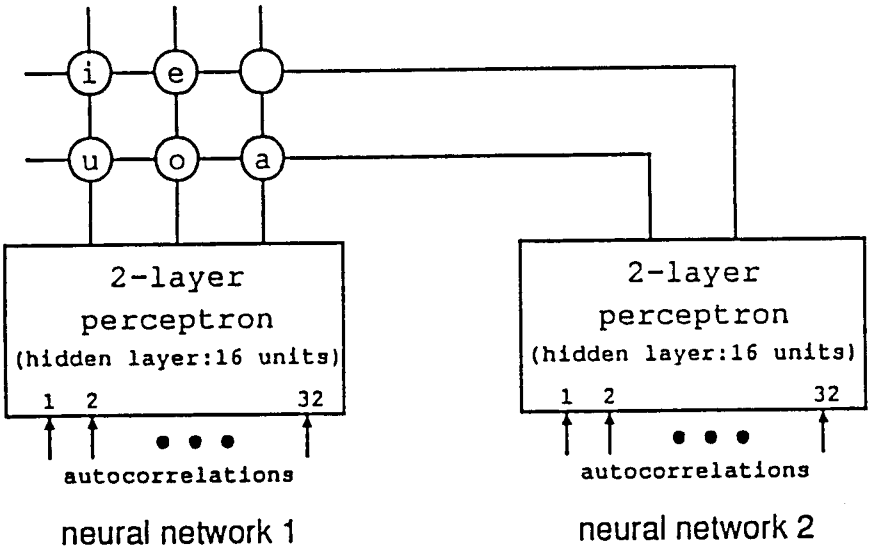

Fig. 4 The neural networks.

aaaaaaaeiiiiiiiuuuuuuieeeeeeeeoooooooo
aaaaaaaoeiiiiiiiuuuuuuieeeeeeeeoooooooo

Fig. 5 The Output of Vowels (for two seconds).

4. CONCLUSION

A new method of vowel recognition based on zero-crossings is proposed. Using this method, we obtained satisfactory results for Japanese vowel recognition. Because of the simplicity of the processing, the recognition speed is very fast. We plan to try to recognize consonants as well as vowels by using simple amplitude information.

REFERENCES

1. Licklider, J. C. R., and Pollack, I., "Effects of differentiation, integration, and infinite peak clipping on the intelligibility of speech," J. Acoust. Soc. Am., 20, January 1948, pp. 42-51.

2. Voelcker, H. B., "Toward a unified theory of modulation," in Proc. IEEE, 54, March 1966, pp. 340-353, and May 1966, pp. 735-755.

3. Arai, T., and Yoshida, Y., "Study on zero- crossings of speech signals by means of analytic signal," J. Acoust. Soc. Jpn. (J), 46, March 1990, pp. 242-244.

4. Arai, T., and Yoshida, Y., "A Restoration of Speech Signal from Zero-crossings," Proc. Autumn Meet. Acoust. Soc. Jpn., September 1990, pp. 199-200.

5. Arai, T., and Yoshida, Y., "A Restoration of Speech Signal from Zero-crossings," Tech. Rep. Speech Soc. Jpn., SP90-78, January 1991, pp. 7-13.

6. Kamio, H., Arai, T., and Yoshida, Y., "Real Time Vowel Recognition using Zero-Crossings Autocorrelator," Proc. Autumn Meet. Acoust. Soc. Jpn., October 1991, pp. 163-164.