PERIODICITY EMPHASIS OF VOICE WAVE USING NONLINEAR IIR DIGITAL FILTERS AND ITS APPLICATIONS

Hiroyuki KAMATA, Akira KANEKO and Yoshihisa ISHIDA

Department of Electronics and Communication, School of Science and Technology, Meiji University

ABSTRACT

We propose a new method for emphasizing the periodicity of voice wave using chaotic neurons, and propose a practical method to detect the fundamental frequency of human voice. The chaotic neuron is a kind of nonlinear recursive mapping proposed in the field of nonlinear theory and is usually used to generate the chaotic signal. Besides, when the chaotic neuron is considered in the theory of linear signal processing, we can interpret that the chaotic neuron is a positive feedback IIR digital filter of first order, therefore, it gives a spectrum slope to the target spectrum of input speech signal. In this study, we try to tune up the chaotic neuron to amplify the low frequency components to emphasize the component of fundamental frequency. As the result, spectrum peaks based on the formants are canceled, the spectrum peak corresponded to the fundamental frequency of voiced speech can be detected easily. In addition, a nonlinear function that has a dead band is included in the feedback loop of the chaotic neuron. As its effect, noise components of unvoiced speech are not amplified.

1. INTRODUCTION

In case of Japanese and Chinese languages, the intonation accent of speech is important information because meanings of words are decided by the information of intonation. The intonation of voiced speech corresponds to the fundamental frequency that is generated by the glottis.

Many methods for detecting the fundamental frequency of human voice have been proposed[1]. The fundamental frequency of human voice has the delicate sway, so the detection of the sway becomes an important problem. However, the methods that use the auto-correlation function or Fourier transformation detect the average frequency of an analyzed section. Therefore, if the sway of fundamental frequency is required to detect,

(a) The analysis of the fundamental frequency should be detected on the domain of the time series signal.

(b) The length of analysis section should be shortened as much as possible if the auto-correlation function or Fourier transformation is used.

However, the detection precision is reduced in case of (a) when the fundamental frequency is high frequency. Moreover, in case of (b), the spectrum peaks based on formants have a great influence on the fundamental frequency detection.

Then, we propose a new method that detects the fundamental frequency. First, we try to arrange the waveform of the target time series signal using nonlinear IIR digital filters. In this paper, we first try to use the chaotic neuron[2-3] as an example of nonlinear IIR digital filters.

The chaotic neuron is a kind of chaotic signal generator devised in the field of the nonlinear theory, the equation of the chaotic neuron is expressed by the recursive mapping that includes a nonlinear function. The nonlinear function is chosen from some discontinuous functions. Therefore, the input signal is modulated to the chaotic signal like a random signal.

Besides, when the equation of the chaotic neuron is evaluated by the theory of linear signal processing, the structure of the chaotic neuron is a first order positive feedback digital filter. Therefore, we can suppose that the characteristic is effective in the emphasis of the low frequency components that include the fundamental frequency.

Linear digital filters may also express the same effect. However, the effect will not become a suitable one when the unvoiced speech is processed because the noisy signal is also emphasized, therefore, the periodicity detection of signal becomes a difficult situation.

Besides, we have found that the chaotic neuron emphasizes the periodicity of the target signal without emphasizing the noisy signal when a combination of parameters of the nonlinear function included in the chaotic neuron is chosen suitably.

In this study, we first examine the effect of the chaotic neuron and choose suitable parameters of the nonlinear function. Next, we apply the proposed method to the fundamental frequency detection of human voice, and show that the proposed method is effective.

2. CHAOTIC NEURONS

The chaotic neuron[2-3] is shown in Eq. (1).

$$x(n) = s(n) + f(x(n-1)) \tag{1}$$

where, $s(n)$ is the input signal, $x(n)$ is the internal state, and the $f(x)$ is the nonlinear function.

The nonlinear function $f(x)$ is usually represented by the discontinuous function shown as follows.

$$f(x) = \begin{cases} \kappa \cdot x + \sigma & : \quad x \leq -\varepsilon \\ \dfrac{\kappa \cdot \varepsilon - \sigma}{\varepsilon} x & : \quad -\varepsilon < x < \varepsilon \\ \kappa \cdot x - \sigma & : \quad \varepsilon \leq x \end{cases} \qquad (2)$$

In this study, we try to use the nonlinear function $f(x)$ shown in Eq. (3) for making the gap in which the function $f(x)$ outputs the value zero.

$$f(x) = \begin{cases} \kappa \cdot x + \sigma \ : \ x \leq -\varepsilon \\ 0 \quad : \ -\varepsilon < x < \varepsilon \\ \kappa \cdot x - \sigma \ : \ \varepsilon \leq x \end{cases} \qquad (3)$$

By the above, we expect that the amplification of small signal such as unvoiced signal is avoided.

Besides, if $\kappa = a > 0$ and $\varepsilon = 0$, Eq. (1) corresponds with Eq. (4).

$$x(n) = s(n) + a \cdot x(n-1) \qquad (4)$$

As Eq.(4) has the pole $(a,0)$ in Z-plane, low frequency components that include direct (0[Hz]) component are amplified.

However, the characteristic shows that the input signal is integrated (Figure 1(b)). Therefore, if the input signal has the direct component by any reason, the output signal of Eq. (1) increases monotonously. Furthermore, in case of the linear digital filter such as Eq. (4), the noise components that are included in low frequency area are also amplified (Figure 1(d)).

We can expect that the chaotic neuron can resolve the problems that emphasize the noise. However, it is difficult to evaluate the advantage of Eqs. (1) and (3) theoretically because these equations are nonlinear systems. Then, we try to evaluate the systems by some experiments.

2.1. The Effects of the Coefficients κ, σ and ε

Figure 2 shows some experimental results for various coefficients. As the figure shows, the coefficient σ has the effect that keeps away integrating the signal(Figure 2 (b), (c) and (d)). Usually, σ is set to 10.0 when the chaotic signal is generated using the chaotic neuron [3]. Besides, we think that σ is suitable to be set to small value for amplifying the low frequency component when the linear system such as Eq. (4) is considered.

The coefficient ε has the effect that keeps away the amplification of the small signal such as unvoiced signal as expected (Figure 3 (b) and (c)).

It is an ideal technique that can process without distinction between voiced and unvoiced speech, therefore, the effect of the coefficient ε is an useful one.

Furthermore, we have confirmed that the effect of the coefficient κ is almost the same with the coefficient a in Eq. (4), therefore, if the κ is larger than 1.0, the output signal $x(n)$ of Eq. (1) has the tendency of divergence.

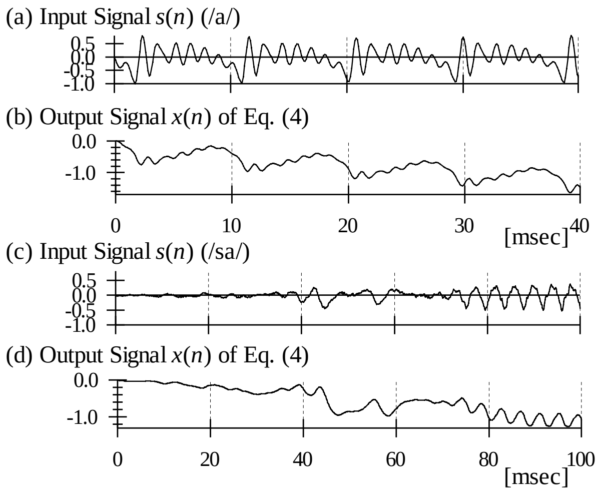

Figure 1: Periodicity emphasis by the linear digital filter Eq.(4). (a) and (c) are the input signals $s(n)$. (b) and (d) are the output signals $x(n)$ by Eq.(4)(the coefficient a = 1.0). (a) and (c) are /a/ and /sa/ of human voices, respectively.

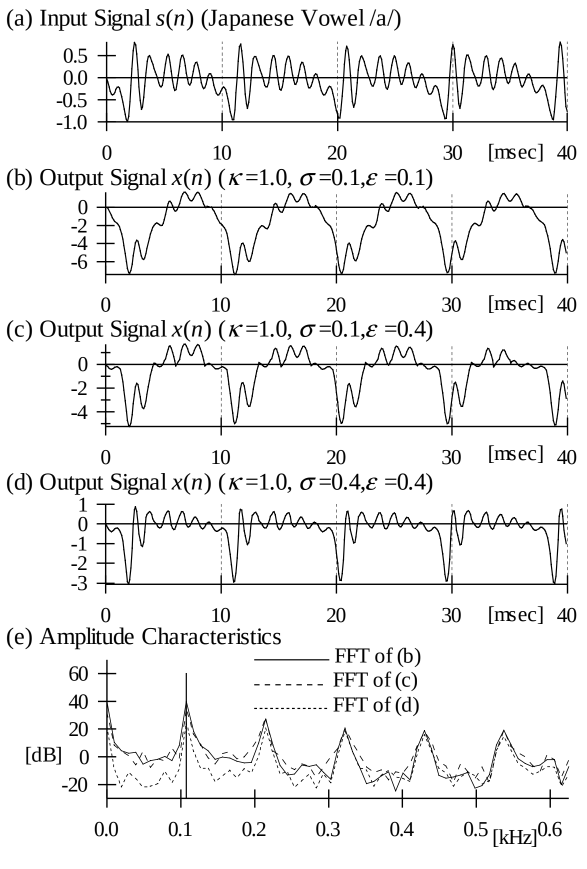

Figure2: Periodicity emphasis by the chaotic neuron Eq.(1) and Eq. (3). (a) is the input signal $s(n)$. (b) ,(c) and (d) are the output signals $x(n)$ by Eq.(1). (e) is the amplitude characteristics of the output signal(b), (c) and (d).

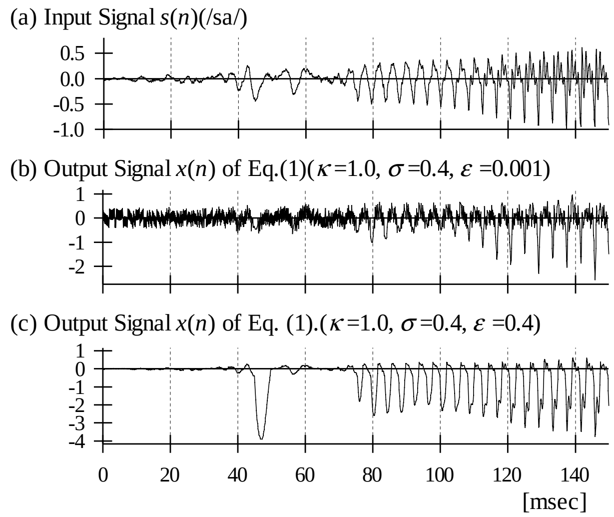

Figure3: Periodicity emphasis by the chaotic neuron Eq.(1) and Eq. (3). (a) is the input signal $s(n)$ (/sa/). (b) and (c) are the output signals $x(n)$ by Eq. (1) and Eq. (3).

3. FUNDAMENTAL FREQUENCY DETECTION USING THE CHAOTIC NEURON AND MODIFIED FFT

In case of human voice, the fundamental frequency is observed in the frequency area from 80[Hz] to 1200[Hz]. If we try to detect the fundamental frequency based on the spectrum domain, the spectrum peak of lowest frequency generated by the glottal wave is needed to detect. Usually, spectrum peaks of the glottal wave are resonated by the formants, therefore, the spectrum peak that shows the maximum value is not always the fundamental frequency's peak.

By using the chaotic neuron, the spectrum envelope that the amplification decreases as the frequency goes up is added to the target signal, so we can expect that the fundamental frequency is estimated by detecting the maximum spectrum peak in the target frequency area in case of our study. Furthermore, we can distinguish between voiced and unvoiced speech by the value of the detected spectrum peak.

Besides, the chaotic neuron can be implemented by personal computers in real time. However, the spectrum computation is not always easy to compute even if the fast Fourier transformation is used.

Then we try to use the techniques of FFT pruning and modify the algorithm in order to fit our purpose.

3.1. FFT Pruning

The large numbers of data are needed to calculate FFT for getting the high frequency resolution Besides, for observing the sway of the fundamental frequency, it is not suitable that the number of data is large. Usually, zero-padding is performed to the target signal shown in Eq. (5) for resolving the problem.

$$y(n) = \begin{cases} x(n) & : n = 0,1,\cdots, N-1 \\ 0 : n = N, N+1,\cdots, NP-1 \end{cases} \qquad (5)$$

However, in case of the data $y(n)$, the large number of times that zero are added and multiplied with any data increase. Then, the FFT pruning technique [4] has been proposed for avoiding waste. The concept of the FFT pruning is to avoid processes for computing with zero.

Besides, the frequency area that the fundamental frequency of human voice exits is limited to low frequency, so all of frequency area is not needed to calculate. However, the normal FFT and FFT pruning technique do not resolve the problem.

3.2. Modified FFT pruning

Then, we try to prune the radix-2 FFT algorithm more. First, three parameters *li*, *lp* and *la* are prepared. The parameters 2^{li}, 2^{lp} and 2^{la} show the number of input data, the resolution of Fourier transformation and the number of output data, respectively. For example, in case of *li*=3, *lp*=4 and *la*=2, the pruned algorithm that avoid processes for computing with zero is shown in Figure 4.

In case of out study, the sampling frequency is set to 22.050[kHz]. Then, we decide the parameters as follows:

The number of input data: *li*=9,(2^{li}=512): In case of male voice, 2~3 pitch periods of voiced speech are included in this data length. (If female voice is analyzed, this parameter should be set to *li*=8 or less.)

The number of resolution of FFT: *lp*=10, (2^{lp}=1024).

The number of the output data: *la*=5, (2^{la}=32): By this condition, the frequency area [0[Hz] - 689[Hz]] is analyzed. (In case of female voice, *la* should be set to 6.)

By the device, the quantity of computation is reduced to about 60[%] compared with normal FFT(2^{10}).

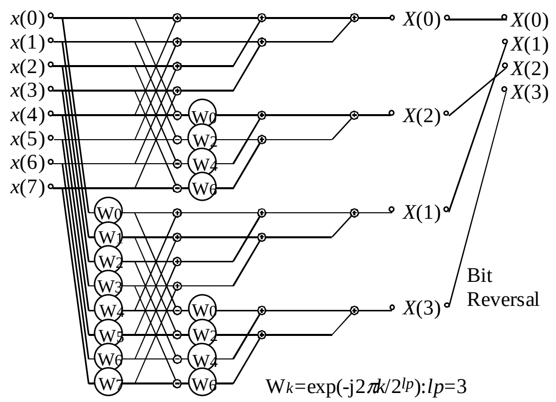

Figure 4: Block diagram of the modified FFT pruning. (*li*=3, *lp*=4, *la*=2)

3.3. Fundamental Frequency Detection

Figure 5 shows the algorithm for detecting the fundamental frequency using the periodicity emphasis by the chaotic neuron and modified pruned FFT. As we have mentioned above, the spectrum envelope is added to the target signal $s(n)$ by using the chaotic neuron, so the spectrum peak that shows the maximum volume in the low frequency area corresponds to the fundamental frequency's peak. Therefore, the algorithm that is shown in Figure 5 carries out the maximum volume detection. The difference between unvoiced and voiced speech signals is distinguished by the volume of the detected spectrum peak.

Furthermore, the peak picking method is used for improving the accuracy of the fundamental frequency detection in this algorithm.

The experimental results are shown in Figure 6. In case of that the chaotic neuron is used, the difference of the volume of the spectrum peaks between unvoiced and voiced speech signals increases (Figure 6(c)).

In addition, the percentage that the algorithm shown in figure 6 can detect the correct fundamental frequency has become 92[%] in the experimentation of 10,000 times. As the future works, we will try to improve the percentage of correct answers.

.

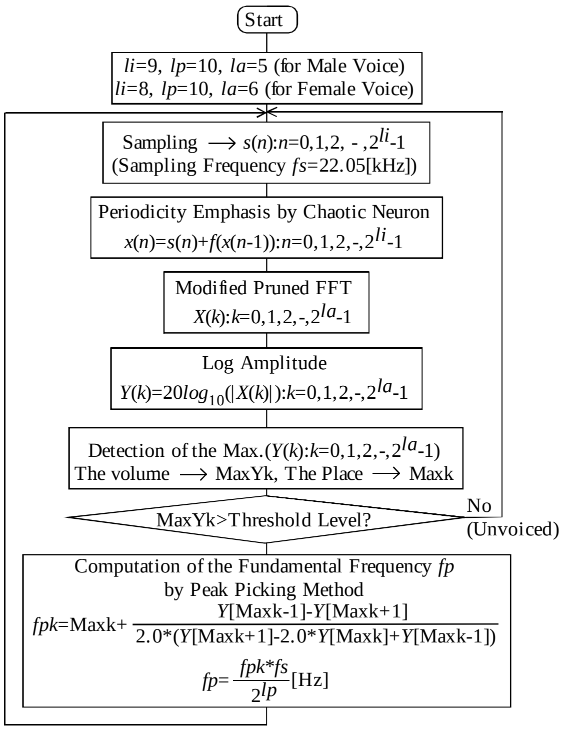

Figure 5: Algorithm of the fundamental frequency detection using the chaotic neuron and the modified pruned FFT.

4. CONCLUSION

We have proposed a method for emphasizing the periodicity of voiced speech using the chaotic neuron and have improved the potential of the fundamental frequency detection with FFT.

In case of the proposed method, it can process the speech signal without distinction between voiced and unvoiced speech signals by selecting the parameters of the chaotic neuron, so the process for fundamental frequency detection becomes simple.

In addition, we have also tried to improve the calculation efficiency of FFT using pruning techniques and have confirmed that the quantity of computation reduces to 60[%] compared with normal FFT algorithm.

As future works, (a) it is needed to improve the percentage of correct answers by the proposed algorithm, (b) the real time processing using personal computer is needed to realize. We would like to report the problem in next opportunity.

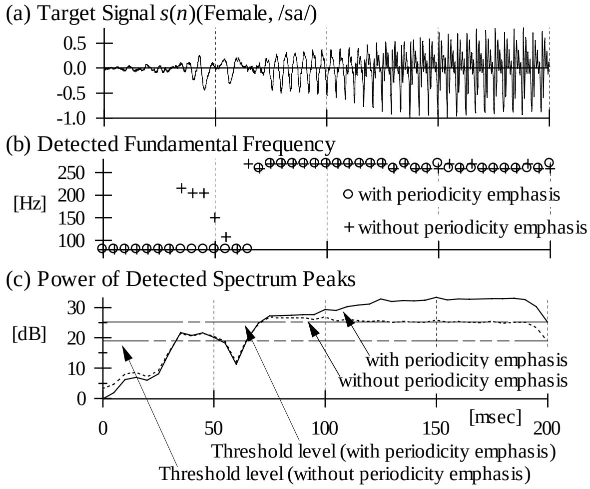

Figure 6: Experimental results of the fundamental frequency detection by the algorithm shown in figure 5. (κ=1.0, σ=0.4, ε=0.4)

5. REFERENCES

1. Hess W., *Pitch Determination of Speech Signal*, Springer-Verlag, 1983.
2. Aihara K., Takebe T. and Toyoda M., "Chaotic Neural Networks", Physical Review Letters A, Vol. 144, 1990.
3. Kamata H., Endo T. and Ishida Y., "Communication with Chaos via DSP Implementation", Proceedings of ISCAS'97, Vol. 2, 1997.
4. Markel J. D., "FFT Pruning", *IEEE Trans., AU* 19, No. 4, 1971.