Proceedings 2005 IEEE International Symposium on
Computational Intelligence in Robotics and Automation
June 27-30, 2005, Espoo, Finland

A Simple Algorithm of Pitch Detection by using Fast Direct Transform

Yuuki Yazama
Faculty of Engineering
The University of Tokushima
rabbit@is.tokushima-u.ac.jp

Yasue Mitsukura
Faculty of Education
Okayama University
mitsue@cc.okayama-u.ac.jp

Minoru Fukumi
Faculty of Engineering
The University of Tokushima
fukumi@is.tokushima-u.ac.jp

Norio Akamatsu
Faculty of Engineering
The University of Tokushima
akamatsu@is.tokushima-u.ac.jp

***Abstract*— There are some fundamental frequency detection methods for speech processing such as cepstrum analysis. However, the traditional methods need a lot of computing time and arithmetic processing. Moreover, a speech signal is converted into frequency domain by using an analysis section. In this paper, we propose a fast direct transformation (FDT). FDT extracts an amplitude feature of a signal by a simple computation. We perform the fundamental frequency detection of speech signal by using FDT. We compare the FDT algorithm with the conventional fundamental frequency detection methods by using teacher data(fundamental frequency) detected by the inspection. We perform an improvement as compared with a autocorrelation method by using FDT algorithm.**

***Index Terms*— speech processing, pitch detection algorithm, fast direct transform.**

I. Introduction

The speech recognition is researched as a means of communications between man and machine. However, it has never been widespread though high accuracy can be comparatively obtained by adding a condition for the speech recognition as a practical system. The phenomenon of nonlinear individual characteristic is changeable[1]. In the research of speech recognition, a typical traditional analysis method is a frequency analysis method such as FFT, LPC analysis and cepstrum analysis. An auditory analysis section is set for these methods. However, we think that these methods cannot necessarily extract a specialized feature of voice. Because these methods have the analysis section and a frequency feature is a comprehensive character in the analysis section. Therefore, we propose FDT that is a fast computation method using an amplitude of signal in time domain. We perform the fundamental frequency detection by using FDT. The fundamental frequency of voice signal is the most basic information and indispensable information for the auditory analysis[2][3]. Moreover, there have been the researches to build the pattern of fundamental frequency into the speech processing[4]-[6].

First, we describe FDT. FDT represents the voice signal by a rugged waveform. The rugged waveform is a shape of signal which consists of a mountain and a valley. Second, we explain a fundamental frequency detection algorithm by using the rugged spectrum and the standard deviation spectrum. The proposed fundamental frequency detection algorithm (FDT algorithm) uses a simple computation by a smoothing process and a threshold processing. We detect the starting point of a periodic waveform from the teacher data by the inspection. Finally, we compare the FDT algorithm with traditional methods by using some teacher data.

II. Fast Direct Transformation(FDT)

Fast Direct Transformation converts a signal to a shape feature. The signal is expressed by a mountain and a valley of the signal as a pulsed signal. Hereafter, we show a generation method of ruggedness waveform by using FDT. Moreover, the feature of the standard deviation of the signal is used for FDT algorithm of the fundamental frequency extraction that we propose.

A. Ruggedness Spectrum

Ruggedness spectrum is a spectrum that is expressed by the valley or the mountain. Ruggedness spectrum is computed based on whether an amplitude x_k at a certain time k is the mountain or the valley by using threshold α_k (as shown in Fig.1). α_k is calculated by the following expression,

$$\alpha_k = \frac{\sum_{i=k-n}^{k-1} x_i + \sum_{i=k+1}^{k+n} x_i}{N}. \tag{1}$$

where N is a length of section calculated the threshold.

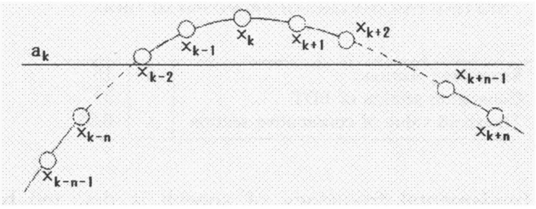


Fig. 1. Threshold α_k and x_k

Ruggedness spectrum is expressed by Eq(2).

$$B_k = \begin{cases} 1 & (\alpha_k < x_k) \\ 0 & (\alpha_k > x_k). \end{cases} \tag{2}$$

In addition, Eq(1) can be transformed into Eq(3).

$$\alpha_k = \alpha_{k-1} + \frac{x_{k-1} + x_{k+n} - x_k - x_{k-1-n}}{N}. \tag{3}$$

0-7803-9355-4 / 05 / $20.00 ©2005 IEEE.

B. Standard Deviation Spectrum

We define standard deviation spectrum that shows amplitude strength information on the signal. This spectrum is calculated by the expression (shown in Eq(4)) calculating the standard deviation.

$$V_k = \sqrt{\frac{\sum_{i=k-n}^{k+n} x_i^2}{N+1} - \bar{x}_k^2}. \qquad (4)$$

where V_k is a value of standard deviation spectrum and $\bar{x}_k$ is an average value of the signal over N length. By calculating the standard deviation spectrum, we can decide whether the signal of N length is a noise or not. Because the signal has a certain level of amplitude.

III. Fundamental Frequency Detection using FDT Algorithm

The similar waveform (fundamental frequency) periodically repeated in the speech is detected by the FDT algorithm. The speech has a feature that the amplitude level is small and the shape is the valley in the part of fundamental frequency. If this valley part can be detected, the fundamental frequency can be calculated and the detection of the repeated similar waveform can be carried out on the time domain.

The detection algorithm of fundamental frequency candidate using FDT is performed by the following process.

1) Double smoothing of speech waveform using smoothing section β
2) Generation of ruggedness spectrum and standard deviation spectrum using FDT section γ
3) First correction of ruggedness spectrum and standard deviation spectrum by using threshold of standard deviation spectrum λ
4) Second correction of ruggedness spectrum and standard deviation spectrum by using the maximum value of each section

The third process is a threshold processing, and the part of V_k less than λ is changed to "0.0" and B_k is changed to "1.0". Moreover, the fourth process is shown Figure 2.

TABLE I

Setting Parameters of Proposed Method.

Smoothing section	β	16
Generation section of FDT	γ	32
Threshold value of consecutive section	λ	0.05

The fundamental frequency of speech is detected by FDT algorithm enclosed with the frame. Talbe I shows the parameters using FDT algorithm.

The detection of fundamental frequency using FDT is performed by the following process.

1) Detection of pitch point and calculation of fundamental frequency by using ruggedness spectrum
2) Correction of pitch point by using average fundamental frequency

The first process of the fundamental frequency detection algorithm is to detect the position of the minimum value of an original signal in the valley section of the ruggedness spectrum. Then, this minimum value point is a starting point of the fundamental frequency which we call "pitch point".

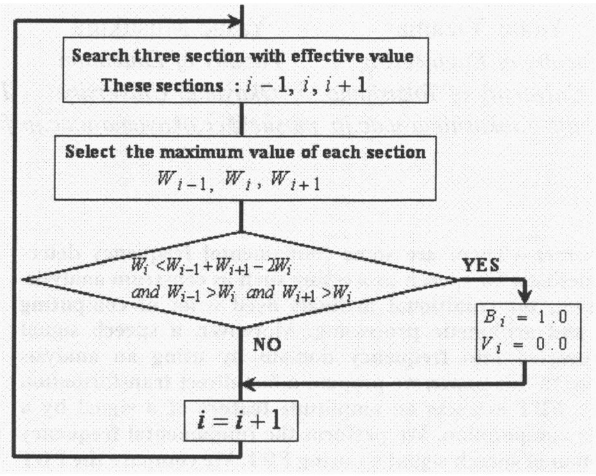

Search three section with effective value
These sections : $i-1$, i, $i+1$
Select the maximum value of each section
W_{i-1}, W_i, W_{i+1}
$W_i < W_{i-1} + W_{i+1} - 2W_i$
and $W_{i-1} > W_i$ and $W_{i+1} > W_i$
YES
NO
$B_i = 1.0$
$V_i = 0.0$
$i = i + 1$

Fig. 2. The fourth process of the detection algorithm of fundamental frequency candidate

IV. Computer Simulation

To verify the effectiveness of the FDT algorithm, the fundamental frequency detected by it is compared with the fundamental frequency detected by using the autocorrelation. Then, the fundamental frequency of teacher data is obtained by manual operation and inspection.

A. Generation of Teacher Data

The speech data are the utterance data of two men and two women recorded by 16 kHz. The number of speech data are 13. We detect the fundamental frequency from speech data by the inspection, and make the teacher data. A generation method of teacher data detects a section of valley of signal. The generation method of teacher data is shown in Figure 3.

B. Result of computer Simulation

The autocorrelation method of the waveform processing used for the computer simulation as a comparative method is described. A simple operation and a waveform processing on time domain are the reason to use the autocorrelation method. Besides, fundamental frequency and the fundamental frequency point are detected by the autocorrelation method as well as the proposed procedure. The fundamental frequency extraction algorithm using the autocorrelation is as follows.

1) The autocorrelation of each 200 points is calculated
2) The mean value of the maximum value and the minimum value of autocorrelation is computed as a threshold, and it generates a signal indicated by mountain or valley by using that threshold

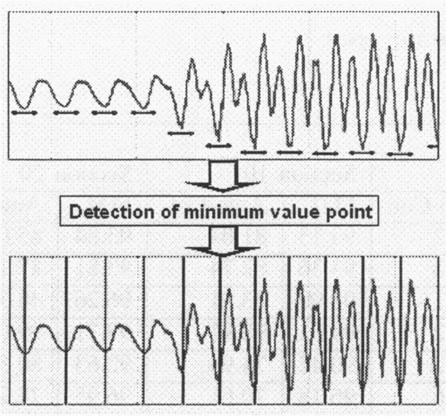


Fig. 3. Method of making teacher data

3) Similar processing to detect fundamental frequency of the FDT algorithm is performed

Moreover, we use the cesptrum method as frequency analysis for fundamental frequency detection. This method is used an analysis section of 1024 point, and maximum value of high quefrency is fundamental frequency.

TABLE II
Error margin of fundamental frequency of teacher data and fundamental frequency of each method

Data	Cepstrum	Auto-correlation	FDT algorithm
Data 1	9.28	8.74	8.17
Data 2	14.83	43.42	7.51
Data 3	5.64	33.56	4.79
Data 4	3.34	1.23	0.78
Data 5	6.55	23.43	8.66
Data 6	9.66	22.66	4.91
Data 7	15.26	10.36	4.09
Data 8	12.91	81.33	11.48
Data 9	10.31	75.97	10.14
Data 10	36.59	19.86	5.10
Data 11	12.84	33.67	16.26
Data 12	90.72	24.61	1.41
Data 13	10.30	49.65	3.54
Average margin	18.25	32.96	6.68

Table II shows the error margin of fundamental frequency between teacher data and each method. Table III shows a conformity of the fundamental frequency detected by the FDT algorithm. To measure the conformity of the fundamental frequency point, a tolerance is used in the fundamental frequency point of the teacher data (shown in Figure 5). The conformity is calculated by judging whether the fundamental frequency point detected by FDT algorithm exist in the permissible section. The number of the conformed fundamental frequency is counted if the pitch point exist in the permissible section, and the conformity is calculated by the ratio to the number of all fundamental frequency of teacher data.

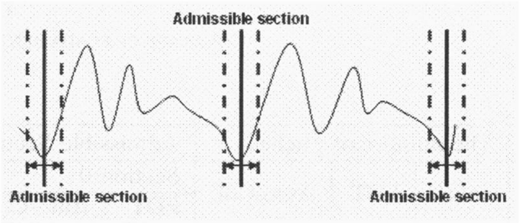


Fig. 4. Searching range of fundamental frequency detection point by using permissible section

V. Consideration

The proposed method obtained the stable result more than the autocorrelation method in error margin between the teacher data and the calculated result(shown Table II). The caused decreased accuracy is excessive deletion of the pitch point. It relates to the change of amplitude level, which are "from the no sound part into a voiced part" and "from a voiced part into the no sound part". When the amplitude level is low in speech analysis, the possibility including enough information of voice is low. Therefore, it is important to give priority to fundamental frequency of an important part, and to detect the part surely.

In general, it is said that the noise talerance of waveform processing is low. Then, we perform a simulation of noise talerance. We perform a simulation of noise talerance. Added noise is 5dB, 10dB, 20dB and 40dB. Noise performance obeys a gaucian distribution and we perform the noise immunity simulation for the cepstrum method and FDT method. As a result of noise talerance, we show Table IV. When strength of the noise is 10dB or more, the result of the FDT method is inferior to that of the cepstral analysis. If the noise is strong, it is necessary to add a processing to decrease the noise.

TABLE IV
The err margin result of noise immunity computer simulation with the teacher data

Noise	FDT method	Cepstrum method
5dB	43.91	39.92
10dB	32.50	29.15
20dB	18.03	20.08
40dB	6.22	18.07

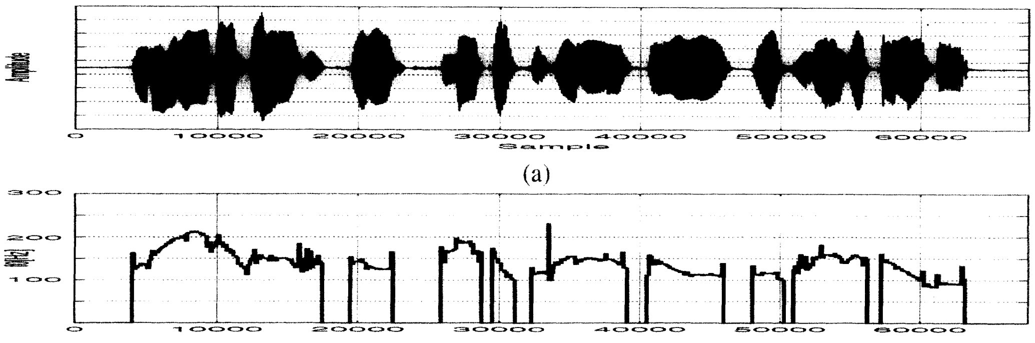
Moreover, we think that there is a problem of a generation method in teacher data. It is necessary to decide the position of the pitch point assumed to be an ideal of the teacher data referring to the amplitude level. However, we make teacher data to be an ideal without considering the amplitude level in this simulation. Therefore, it is thought that establishment of the generation method of teacher data is necessary.

In addition, we describe about the detection accuracy of the pitch point. However, the accuracy of data 5 is lower than that of other data detection accuracy of the pitch point. That reason is that the pitch points are detected in the

TABLE III

COVERAGE OF FUNDAMENTAL FREQUENCY DETECTION POINT

	The number of pitches			Admissible Section(%)							
Data	Teach	FDT	Auto-Col.	Section 0		Section 5		Section 10		Section 20	
				FDT	Auto-Col.	FDT	Auto-Col.	FDT	Auto-Col.	FDT	Auto-Col.
Data1	422	404	419	92.18	72.75	92.42	75.12	93.13	81.04	93.84	85.07
Data2	226	223	294	92.48	74.77	92.48	77.43	93.36	82.74	93.81	87.17
Data3	561	553	632	95.37	19.43	95.90	23.71	95.90	33.51	96.26	48.31
Data4	1021	1040	1020	99.31	98.63	99.41	98.72	99.61	98.92	99.71	99.02
Data5	407	467	417	75.43	64.13	83.05	70.02	90.42	74.94	92.63	80.59
Data6	524	549	515	95.23	56.87	95.80	65.46	96.18	70.61	96.95	78.24
Data7	616	780	635	95.78	75.49	96.26	81.81	96.26	83.93	97.24	87.34
Data8	913	976	701	89.06	61.16	91.03	68.27	93.11	70.90	94.75	73.30
Data9	513	557	379	91.03	36.26	91.81	46.00	92.40	56.73	93.76	61.60
Data10	332	345	359	94.88	47.80	94.88	54.52	94.88	59.94	94.88	64.46
Data11	1061	1041	994	85.11	19.23	88.22	37.70	89.25	58.15	94.72	83.88
Data12	271	271	257	95.57	16.60	95.57	32.10	95.57	53.50	95.57	69.37
Data13	458	463	390	97.60	15.07	98.25	24.24	98.25	36.24	98.47	57.64



(a)

(b)

Fig. 5. Speech waveform and presumption of fundamental frequency by each method(speech waveform (a), fundamental frequency using teacher data (b)

gap of the point to be the fundamental frequency. This phenomenon is also caused in the autocorrelation method (shown Table III). This is a phenomenon caused easily in the autocorrelation method. Even if it is possible to detect the pitch point by autocorrelation method, it is difficult to detect those point correctly in the viewpoint of detection of the pitch point. This is a reason that the result of the pitch point detection accuracy of the autocorrelation method is worse than the proposed method.

VI. CONCLUSION

We proposed the FDT algorithm as a fundamental frequency detection method. Fundamental frequency is detected by using this method, and the effectiveness is verified by comparing with the autocorrelation method. We performed computer simulation comparing this method with the method using a spectrum waveform processing, and the effectiveness of this method is verified as a future work. Moreover, we will consider an algorithm to detect formant.

REFERENCES

[1] Shigeharu YOSHIO, Qigang ZHAO, Tetsuya SHIMAMURA and Jouji SUZUKI, "Pitch Detection Based on Autocorrelation of Root and Fourth-Root Power Spectra", IEICE, Vol.J84-A No.3, March 2001

[2] Hiroshi TOMOMITSU, Hiromitsu KASHIWAGI, KAzuhiro OHTSUKA and Morihiro TAKANASHI , "The Effect of Pitch Frequencies on Speaker Identification", IEICE, Vol.J76-A No.11 pp.1641-1643, November 1993

[3] Yuichi ISHIMOTO, Kentaro ISHIZUKA, KIyoaki AIKAWA and Masato AKAGI, "Fundamental FrequencyEstimation for Noisy Speech Using Entropy-Weighted Periodic and Harmonic Features", IEICE, VOL.E87-D, NO.1, January 2004

[4] YAMASHITA, tomoyoshi ISHIDA and Kazuki SHIMADERA, "A Stochastic F0 Contuor Model Based on Clustering and a Probabilistic Measure", IEICE, VOL.E86-D NO.3, March 2003

[5] Mitsuru NAKAI, Hiroshi SHIMODAIRA and Shigeki SAGAYAMA, "Prosodic Phrase Segmentation Based on Pitch-Pattern Clustering", IEICE, Vol.J77-A No.2 pp.206-214, February 1994

[6] Mitsuru NAKAI, Harald SINGER, Yoshinori SAGISAKA and Hiroshi SHIMODAIRA, "Accent Phrase Segmentaion Based on F_0 Template Using a Superpositional Prosodic Model", IEICE, Vol.J80-D-II No.10 pp.2605-2614, October 1997