

EVALUATION BASED ON 3D DISPLAYS OF VOCAL VOWELS

T.Moriyama¹ and H.Tsuda²

1 Tokushima Bunri University, Kagawa 769-2193, Japan

2 Ebara Corporation.Ltd, Haneda, 144-0042, Japan

Abstract: The aim of the study is the recognition on the voice-prints of vocal vowels using instrument and control engineering method in order to utilize for CIM, security system, the evaluation of the vocal training and the voice compensation for voice handicapped persons. The five stages evaluation on the vocal voice is decided by a professor of the department of music for the lived voice and output of DFT. The authors obtain three dimensional diagrams composed of the frequency, the spectrum and time using DFT.

Keywords: DFT, evaluation of vocal voice, 3D display

1 INTRODUCTION

The aim of this study is the recognition on the voice-prints using instrument and control engineering methods in order to utilize for Computer-Integrated-Manufacturing (CIM), security systems, the evaluation of the vocal training and the voice compensation for voice handicapped persons is requires. M.Takasu (1993) in [1] evaluated the vocalness of female voice as singing formant and I.R.Titze et al (1994) in [2] evaluated the acoustics of the tenor high voice using Fast Fourier Transform (FFT) on a 'resonant voice', respectively.

The authors evaluated the female and male vocal voices evaluated by the professor of the department of music using three dimensional discrete Fourier Transform (3D FFT) on formant f1~f4 and the 'resonant' fr1~fr4.

2 TEST CONDITIONS AND EQUIPMENTS

2.1 Test Condition

The five stage stages (5, 4, 3, 2, 1) evaluation on the vocal voice was classified by the professor of the department of music for A, B and C components of the resonant excellent vocal voices. A, B and C show breathing, making resonant voice, using 'vocal chords' of wind pipe of human body, respectively. The evaluation was decided by a professor of the department of music in our university for the lived voices and output of Digital Audio Tape (DAT).

The vocal voices of /a/, /e/, /i/, /o/, and /u/ per five seconds were gathered in 0 ~ 20 kHz micro phone of DAT and converted to AD converter for a personal computer (PC). The voices were displayed for 0 ~ 10 kHz on the 3D display on the frequency, the spectrum and the time using a discrete Fourier Transform (DFT).

2.2 Equipments

The authors selected NEC PC-9801RA, analyzing software ERMEC's SAS-01, EC-6600 and AD and DA boards with the accuracy of 0.1 % full scale in 0 ~ 20 kHz. The efforts of the analyzing the vocal voices were displayed for 0 ~ 10 kHz based on Shannon's sampling theorem on the 3D display on the frequency, the spectrum and time using DFT.

3 THE EFFORTS OF TEST

The authors obtained three dimensional diagram composed of the frequency, the spectrum and the time using DFT for the total vocal voices and made the DFT for each 51.2 mill seconds of the top in 400 ~ 900 mill seconds of the sampled voice. The data in 0 ~ 10 kHz were utilized for the evaluation based on Shannon's sampling theorem as 1/2 of gathered data in 0 ~ 20 kHz. A 3D display on vocal voices for /a/, /e/, /i/, /o/ and /u/ for five second of the voice with a continuous resonant is shown in **Figure 1**. The horizontal axis means the frequency 0 to 10,000 Hz. The vertical axis means the spectrum of sound level. The top is 100 dB (0 in Figure 1) and 0 dB (-100 in Figure 1). The third axis means time 0 to five seconds.

```

return
* 3D Mode Display *
Cursor line = -1 ( 0.0000 - 51.1500 ms : 0 - 1023 dt)
sch max Cursor Point X = 0.0000 KHz
Y = 0.06496
100 line ( 5068.8000 ms : 101376 dt)
    
```

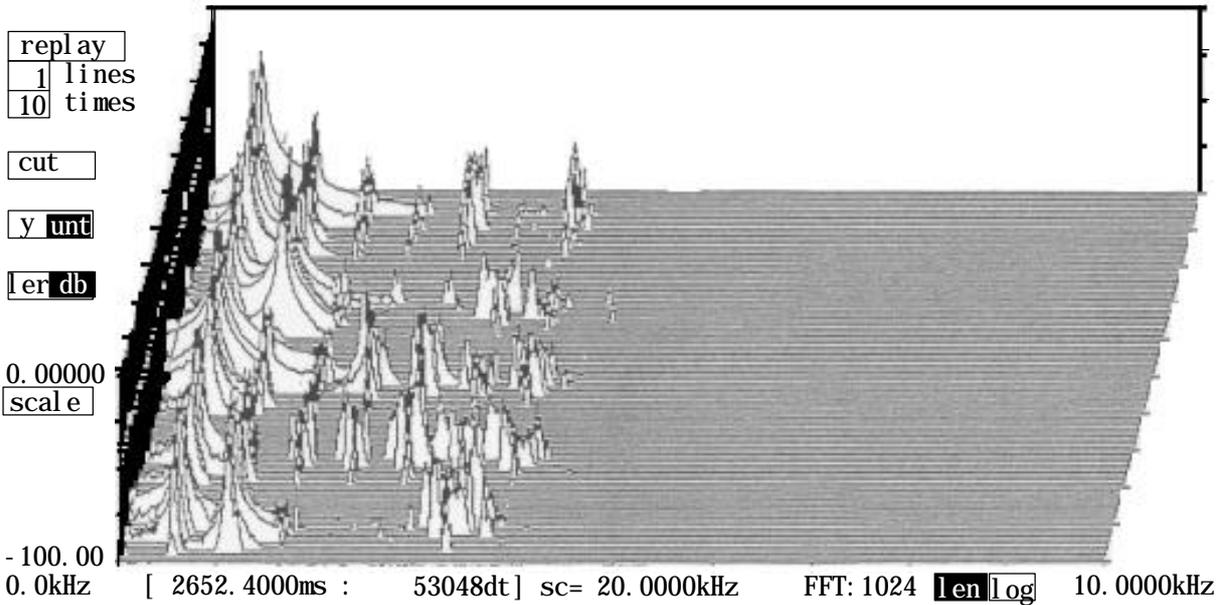


Figure 1 3D display on vocal voices with continuous resonant

A sample of the DFT for the 3D display is shown in Figure 2. It is found out the excellent voice has the additional spectrum in 2 ~ 4 kHz. The spectrum as a 'resonant voice' fr1 ~ fr4 for the formant f1 ~ f4.

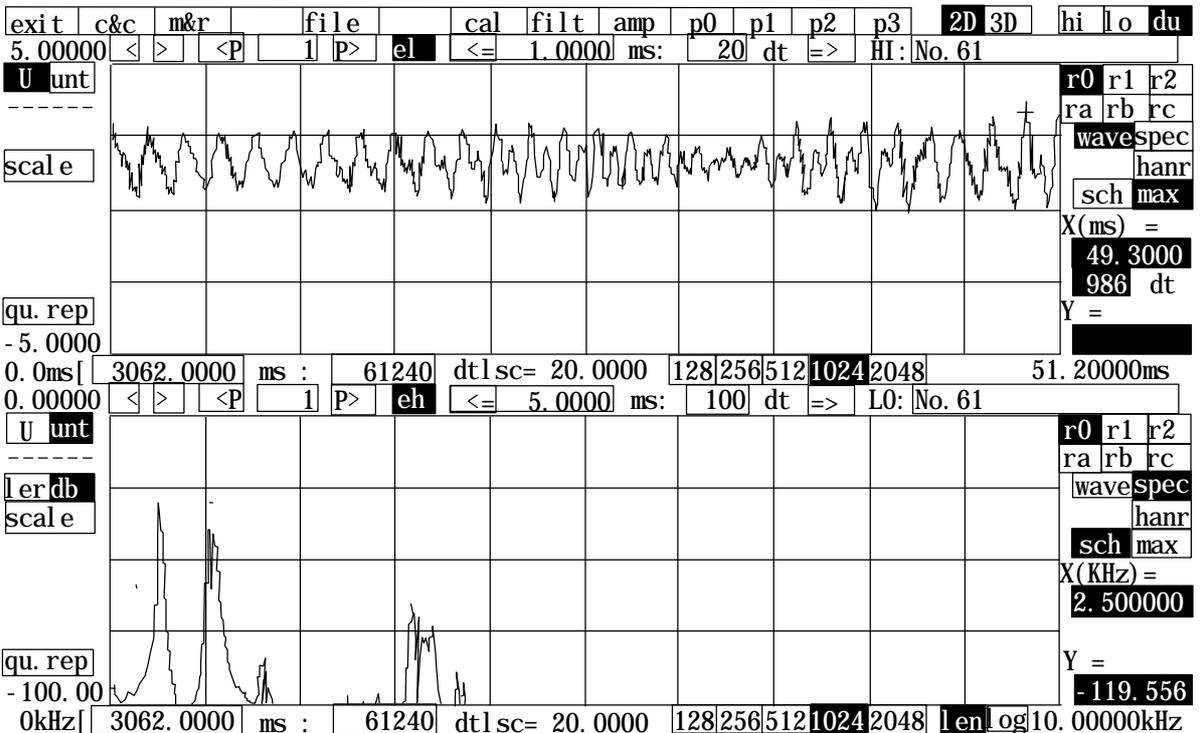
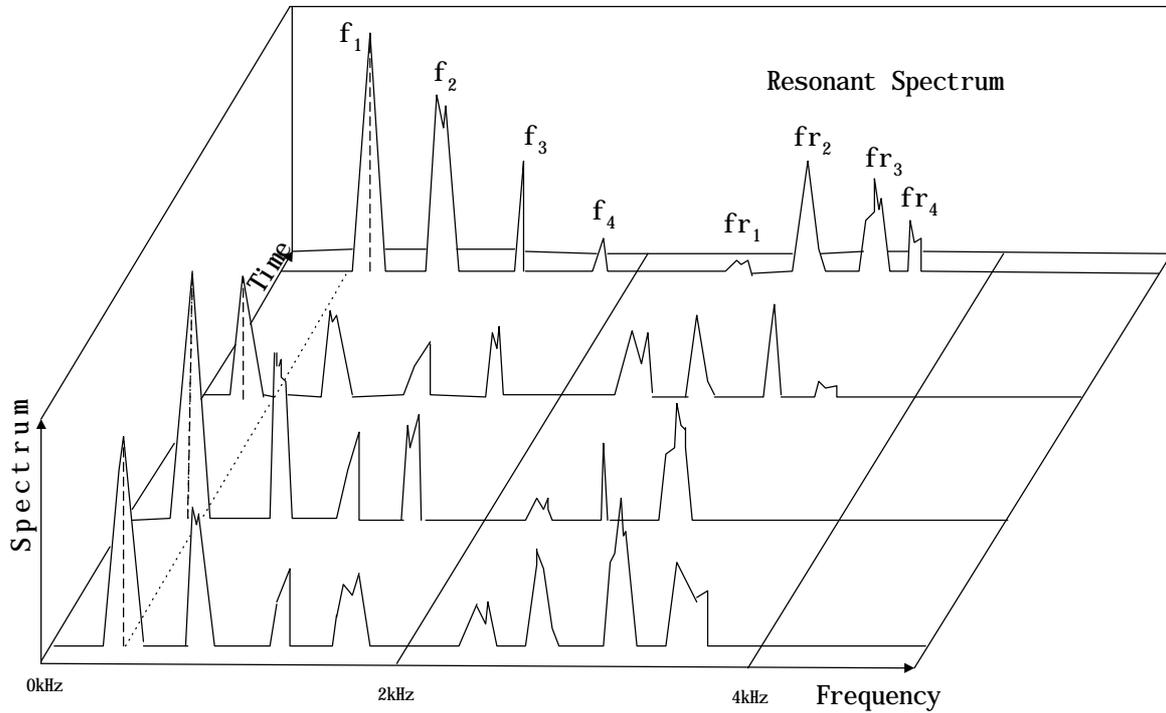


Figure 2 DFT display on vocal voices with continuous resonant

The explanation on the resonant voice is shown in **Figure 3**. It is found out the resonant voice will be depend on the fluctuation of the peak and the frequency of the first formant and the resonant voice is classified as continuous, discontinuous and weak discontinuous resonant.

Figure 3 Explanation on the resonant voice



The comparison table on the classification using continuous, discontinuous and weak discontinuous resonant voice is shown in **Table 1**. The evaluation on the excellent resonant voice by a professor for vocal training is equal to the authors' a simple evaluation based on arithmetic mean for the compornents of A, B and C for fifteen female and male students.

Table 1 Comparison table between the efforts and the evaluation by a professor

	No.of Samples	The evaluation and No.of Samples															*			
		A					B					C								
		No.	5	4	3	2	1	5	4	3	2	1	5	4	3	2		1		
Continuous Resonant	4	1	.					.						.					İ	
		2	.					.						.					İ	
		3				.			.						.				D	
		4				.				.					.					D
Discontinuous Resonant	9	5				.				.							.		x	
		6				.				.							.		x	
		7	.						.					.					İ	
		8	.						.					.					İ	
		9				.			.					.					D	
		10									D
		11	.					.						.						İ
		12					.				.							.		x
Weak Discontinuous Resonant	2	14				.				.							.		x	
		15					.			.							.		x	

* :The authors' a simple evaluation based on arithmetic mean for the compornents of A,B and C.
İ , Δ and x are excellent, good and poor, respectively.

4 CONCLUSIONS

The efforts of the the analysis for vocal vowels are as follows:

- 4.1 The excellent evaluated vocal voices have four peaks as a continuous or discontinuous resonant voices that are 0 ~ 70 % of the sound level of the fundamental frequency in 2.0 ~ 4.0 kHz.
- 3.2 The center frequency and the sound level of the peaks are fluctuated in the time duration.

5 FUTURE STUDY

- 5.1 The neural network method will be applied.
- 5.2 The relation between this evaluation and the evaluation by a professor for vocal training will be studied.

ACKNOWLEDGEMENTS

The authors are grateful to
Prof. Yukako Inatomi / Department of Music in Tokushima Bunri University and her students for their faithful teaching and vocal training.

The authors are grateful to
Prof. Toshie Takahashi and Prof. Kiichi Tsuchiya / Waseda University,
Prof. Emeritus Yoshikazu Sawaragi / Kyoto University.
Prof. Takashi Soeda / President of Tokushima Bunri University and
Prof. Takeshi Koyama / Dean of The Faculty of The university
for their stimulating discussions.

REFERENCES

- [1] M.Takasu, Vocalness of female voice, Society of Acoustics, Japan, Vol.49-no.6, p.381-388, 1993
- [2] I.R.Titze, S. Mapes and B. Story, Acoustics of the tenor high voice, Vol.95-No.2, p.1133-1142, 1994
- [3] H.Tsuda, A.Fukuma and T.Moriyama Short time frequency spectrum of vowels for difference between voice-prints, Proc. EIC, 1995, p.257-257
- [4] H.Tsuda, A.Fukuma and T.Moriyama, Voice characteristics for the different speaking methods, Proc. Joint Convention Record of Electric Engineers, Shikoku-section, 1995, p.117-117
- [5] H.Tsuda, A.Fukuma and T.moriyama, Study on the evaluation of vocal voices using 3D analysis, Proc. EIC, 1996, p.461-461
- [6] H.Tsuda, A.Fukuma and T.moriyama, Evaluation based on dynamic characteristics of vocal vowels, proc. SICE'97 216A-2, 1997, p.1147-1150

AUTHORS: Takashi MORIYAMA: Professor, Dr. Eng., Head of Department of Mechanical Electronic Engineering,

Faculty of Engineering, Tokushima Bunri University, Kagawa 769-2193, Japan,
Phone +81-(0)87-894-5111 ex 3401, Fax +81-(0)87-894-4201,
E-mail:moriyama@me.bunri-u.ac.jp

<http://www.bunri-u.ac.jp/eng/moriyama/MORIYAMA.htm>

Hiroki TSUDA, M.S., Department of Control System Design, Division of information, Communication and Control, Ebara Corporation Ltd., Haneda 144-0042, Japan

Phone:+81-(0)45-565-4123

E-mail:tuda01@ed.ebara.co.jp