International Journal of Innovative
Computing, Information and Control
Volume **13**, Number **5**, October **2017**

ICIC International ©2017 ISSN 1349-4198
pp. 1441–1452

STATISTICAL ANALYSIS AND MODELING OF FORMANT FREQUENCIES OF VOWELS PHONATED BY TRADITIONAL JAPANESE *SHIGIN* SINGERS

Masashi Nakayama[1], Kosuke Kato[2] and Masaru Matsunaga[3]

[1]Graduate School of Information Sciences
Hiroshima City University
3-4-1, Ozuka-higashi, Asaminami, Hiroshima 731-3194, Japan
masashi@hiroshima-cu.ac.jp

[2]Office for Industry-University Co-creation
Osaka University
2-8-A201, Yamadaoka, Suita, Osaka 565-0871, Japan
kato@uic.osaka-u.ac.jp

[3]Department of Electrical and Computer Engineering
National Institute of Technology, Kagawa College
355 Chokushi, Takamatsu, Kagawa 761-8058, Japan

Received February 2017; revised June 2017

Abstract. *In order to understand the articulation associated with changes in the shape of the vocal fold of traditional Japanese Shigin singers, this study investigated the formant frequencies of anechoic recordings of the quasi-steady-state portions of vowels phonated by six trained Shigin singers. Results show that each of the four formant frequencies (F_1, F_2, F_3, and F_4) of the vowels (/u/, /o/, /a/, /e/, and /i/) phonated by Shigin singers differed from those of other singing styles and those of the normal speech mode. The F_1, F_2, and F_3 of the vowels phonated by Shigin singers were statistically modeled as a function of the singer's gender, phonated vowels, loudness level, and phonation frequency. Each of the models was described using a simple linear equation. These findings are expected to be useful for synthesis of the singing voice, development of training support systems, and understanding of the voice production mechanisms of singers.*

Keywords: *Shigin* singing, Speech, Formant frequency, Statistical modeling, Linear prediction model

1. **Introduction.** Human vocalization plays an important role in human communications. In order to understand human vocalization, it is essential to examine the mechanism of articulation: the maneuvers made to adjust the shape of the vocal tract during phonation. The shape of the vocal tract – the lips, tongue, jaw, velum, and larynx – is changed by articulations. Several methods have been proposed to examine the vocal tract shapes, e.g., magnetic resonance imaging [1], computer tomography [2,3], direct measurement of airway resonance [4], and analysis of formants [5-7]. One of the least invasive and most convenient methods is to analyze formants. In [5], formants were defined as the poles of the transfer function of the supraglottal vocal tract; the pole frequencies were labeled $F_1, \ldots, F_n$ and their bandwidths $B_1, \ldots, B_n$. This definition was followed by many studies [6,7]. Therefore, a number of studies have investigated the formant frequencies of speech and singing [6,8,9]. However, there have been no systematic attempts to examine the formant frequencies of the vowels phonated by traditional Japanese *Shigin* singers.

Shigin is a traditional Japanese singing style for the recital of Japanese or Chinese poetry in Japan. The reading conforms to a melodic line called *Seicho*. *Shigin* has been practiced continuously since its establishment in the 19th century [10]. Early *Shigin* singers were not required to follow any melodic rules, and originally, performances were relatively freeform. *Shigin* was formalized in the early 20th century by a group of *Shigin* singers who invented melodic rules and established schools to train singers according to these rules [11]. Over time, as schools and instructors developed their own styles, the melodies produced by these formative rules diverged. To achieve a full, rich *Shigin* style, students must undergo special training in breathing and phonation as well as develop techniques such as *yuri* (vibrato) and *fushi* (control of the melodic trajectory and singing volume). Correct accents must also be learned because each word in a poem must be sung clearly and correctly to effectively convey the meaning. To this end, the Nippon Ginkenshibu Foundation (an association dedicated to preserving the traditional arts of *Shigin* and sword dancing) has established a standard set of accents for *Shigin* [12]. Musically, *Shigin* uses a relative score, expressed in the form of numbers, such as 2'-2-3. In standard musical notation, this is equivalent to C-D-E.

This study investigates the formant frequencies of vowels phonated by traditional Japanese *Shigin* singers. Its purpose is to answer the following research questions.

[Q1] Does each of the four formant frequencies (F_1, F_2, F_3, and F_4) of the vowels (/u/, /o/, /a/, /e/, and /i/) phonated by *Shigin* singers differ from those of other singing styles and those of the normal speech mode?

[Q2] Can the F_1, F_2, F_3, and F_4 of the vowels phonated by *Shigin* singers be statistically modeled as a function of the singer's gender, phonated vowels, loudness level, and phonation frequency?

[Q3] Can each of the models be described using a simple linear equation?

This paper is organized as follows. Section 2 describes the method of this study. In Section 3, the results of this study are shown and discussed in order to answer each of the above research questions. Section 4 summarizes the conclusions of this study.

2. Method.

2.1. Anechoic recordings of *Shigin* singing.

Anechoic recordings of the vowels phonated by *Shigin* singers were used for the acoustic analysis. Six trained singers participated in the recording sessions. The phonated vowels were recorded in an anechoic chamber using a 1/2-inch microphone (type 4189; Bruel and Kjaer) at a 48 kHz sampling rate and 16-bit resolution. The distance between the singer and the microphone was 50 cm, which is a typical distance for recording singing. The database was constructed from the phonations of five Japanese vowels (/u/, /o/, /a/, /e/, and /i/) in three different pitches (low, medium, and high) at three different strengths (weak, medium, and strong), which resulted in a total of 1,620 samples from the six singers (six trials of the five vowels for each of the nine pitches and strength combinations). Each vowel was phonated by the singer when a pitch was provided as a reference scale sound, and hence, the pitches of the recorded phonations differed accordingly. Half of the database was used for the subsequent acoustic analysis by extracting three initial trials. This is done because three trial samples were considered sufficient for conducting the subsequent statistical analysis.

Table 1 lists the singers who participated in the experiment and the target pitches phonated by the *Shigin* singers. The following acoustic features were considered: F_0, an equivalent continuous A-weighted sound pressure level L_{Aeq}, and the formant frequencies F_1, F_2, F_3, and F_4. Here, F_0 indicates the pitch and L_{Aeq} indicates the loudness.

2.2. Extraction of quasi-steady-state portions.

Figure 1 shows two representations (waveform and spectrogram) of the Japanese vowel /a/ phonated by Singer 2 (7-hon, female) in a medium pitch at medium strength. From the spectrogram, the voiced section

TABLE 1. Singers who participated in the experiment and the target pitches phonated by the *Shigin* singers

Singer (*honsu*, gender)	Pitch		
	Low	Mid.	High
Singer 1 (8-hon, female)	A3 (220.00 Hz)	E4 (329.63 Hz)	A4 (440.00 Hz)
Singer 2 (7-hon, female)	G3 (196.00 Hz)	D4 (293.66 Hz)	G#4 (415.30 Hz)
Singer 3 (6-hon, female)	A3 (220.00 Hz)	D4 (293.66 Hz)	G4 (392.00 Hz)
Singer 4 (6-hon, female)	G3 (196.00 Hz)	D4 (293.66 Hz)	G4 (392.00 Hz)
Singer 5 (3-hon, male)	F3 (174.61 Hz)	C4 (261.63 Hz)	G4 (392.00 Hz)
Singer 6 (2-hon, male)	D#3 (155.56 Hz)	A#3 (233.08 Hz)	D#4 (311.13 Hz)

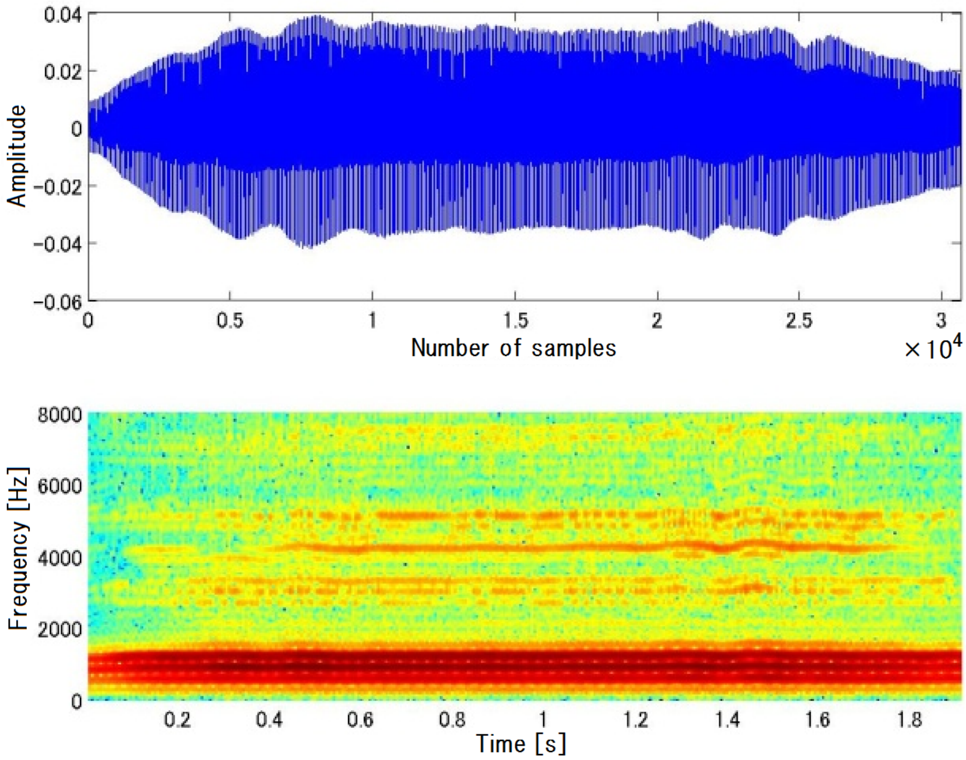


FIGURE 1. Waveform (upper) and spectrogram (lower) of the Japanese vowel /a/ phonated by Singer 2 (7-hon, female) in a medium pitch at medium strength

had a fundamental frequency of approximately 300 Hz. However, the volume of the head and tail sections are at an unsteady-state level, and transient vibrato signatures were found. Therefore, this study focused on the portions which were acoustically quasi-steady-state in terms of amplitude and frequency. To statistically isolate the quasi-steady-state portion of a vocal segment, two types of sub-segments were trimmed from the beginning and end of the vocalization in advance: Type I, segments whose L_{Aeq} was lower than the lowest vocalization level; and Type II, segments whose F_0 was higher or lower than the median of the time series data of the vocalized vowel by 150 cents (approximate maximum range of the vibrato [13]).

2.3. **Extraction of F_1, F_2, F_3, and F_4 from quasi-steady-state portions.** In order to analyze the four formant frequencies (F_1, F_2, F_3, and F_4) from the quasi-steady-state portions of the phonations, the robust formant tracking function using linear predictive coding analysis [14] in Praat [15] was used with the following default conditions:

- time step: automatic
- maximum number of formants: 5
- maximum formant frequency: 5,500 Hz
- window length: 0.025 s
- pre-emphasis: from 50 Hz
- standard deviation: 1.5
- maximum number of iterations: 5
- tolerance: 0.000001

The dimensions of F_1, F_2, F_3, and F_4 were compressed by calculating their geometric mean in the quasi-steady-state portion of their time series data.

Figure 2 shows an example of the measured F_1, F_2, F_3, and F_4 as a function of time for the Japanese vowel /a/ phonated by Singer 2 (7-hon, female) in a medium pitch at medium strength. The distributions of F_1 and F_2 measured from the dataset are shown in Figure 3, while those of F_3 and F_4 are shown in Figure 4.

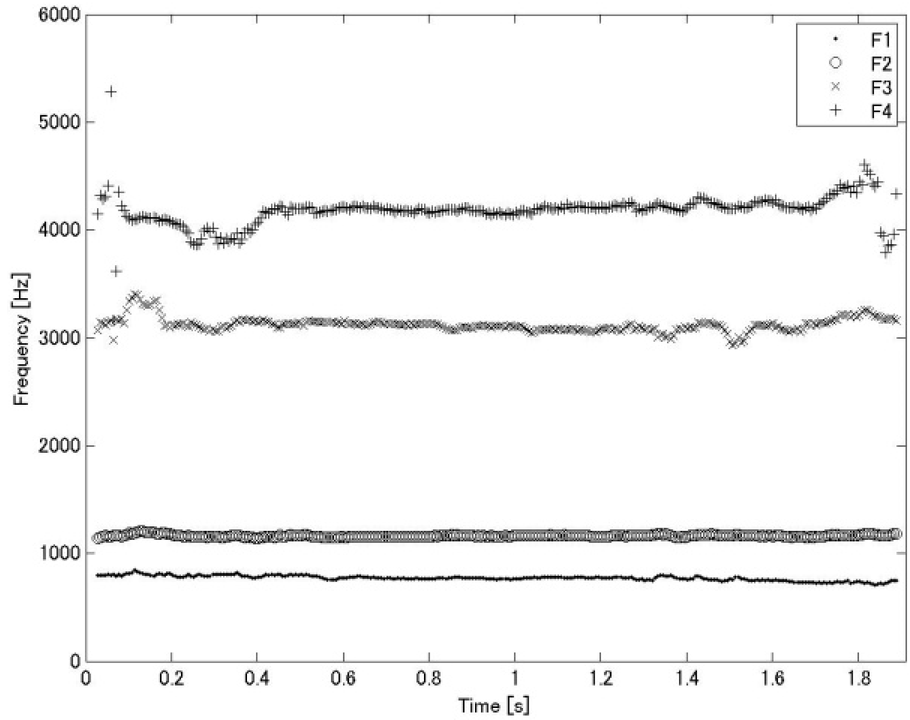


FIGURE 2. Example of measured F_1, F_2, F_3, and F_4 as a function of time extracted from the Japanese vowel /a/ phonated by Singer 2

2.4. **Dataset construction.** To describe F_1, F_2, F_3, and F_4 in relation to the qualitative attributes (*Singer*, *Gender*, *Honsu*, *Vowel*, *Pitch*, *Volume*, and *Trial*) and acoustic features (L_{Aeq}, F_0, F_1, F_2, F_3, and F_4), the feature vector for each utterance was defined as

$$\mathbf{F}_{Singer,Gender,Honsu,Vowel,Pitch,Volume,Trial} = (L_{Aeq}, F_0, F_1, F_2, F_3, F_4) \quad (1)$$

The dimensions of F_0, F_1, F_2, F_3, and F_4 were compressed by calculating the geometric mean of each in the quasi-steady-state portion. For the following analysis, seven samples with L_{Aeq} less than 55 dB were omitted from the dataset. Thus, the final dataset was composed of 13 variables (see Equation (1)) × 803 samples (6 singers × 5 vowels × 3 volume levels × 3 pitches × 3 trials – 7 samples).

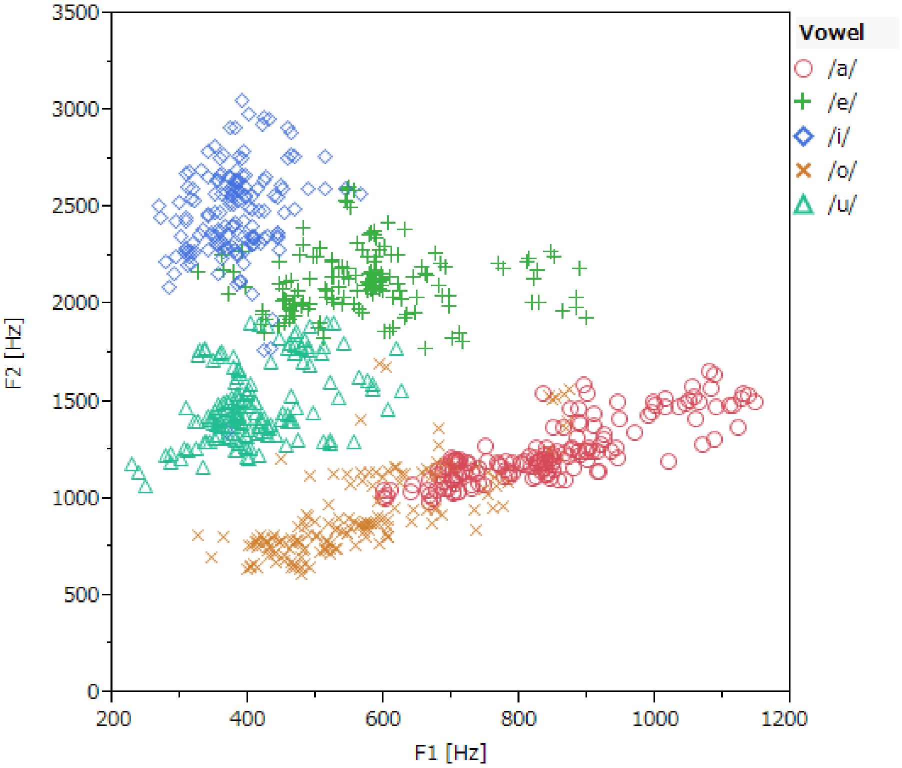


FIGURE 3. Distributions of F_1 and F_2

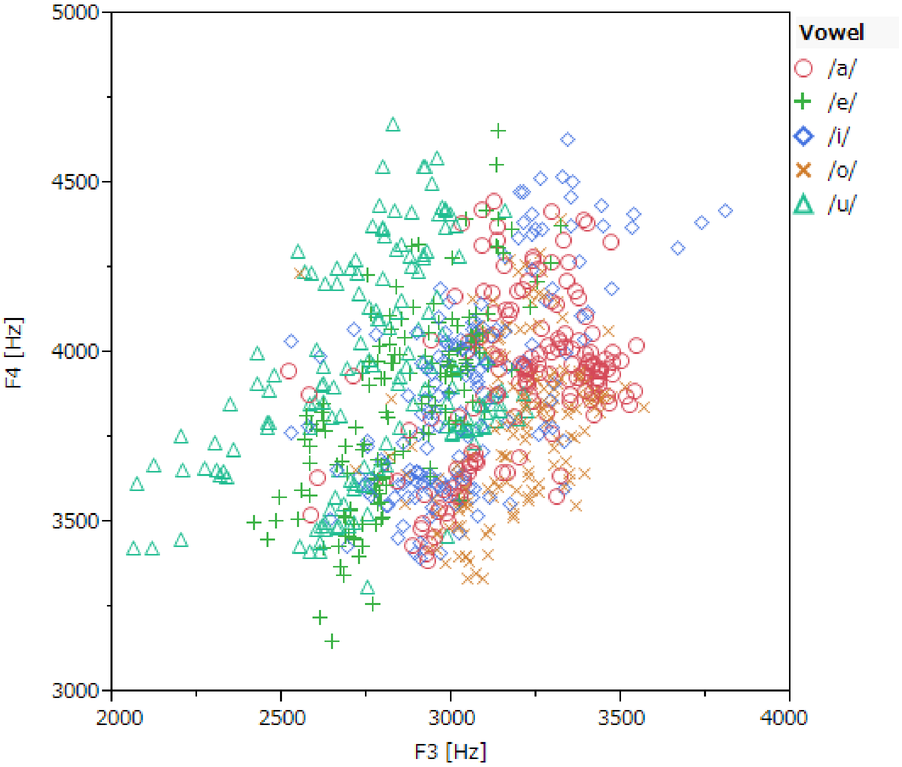


FIGURE 4. Distributions of F_3 and F_4

2.5. **Statistical analysis.** The statistical analysis was conducted using JMP (Version 10) [16].

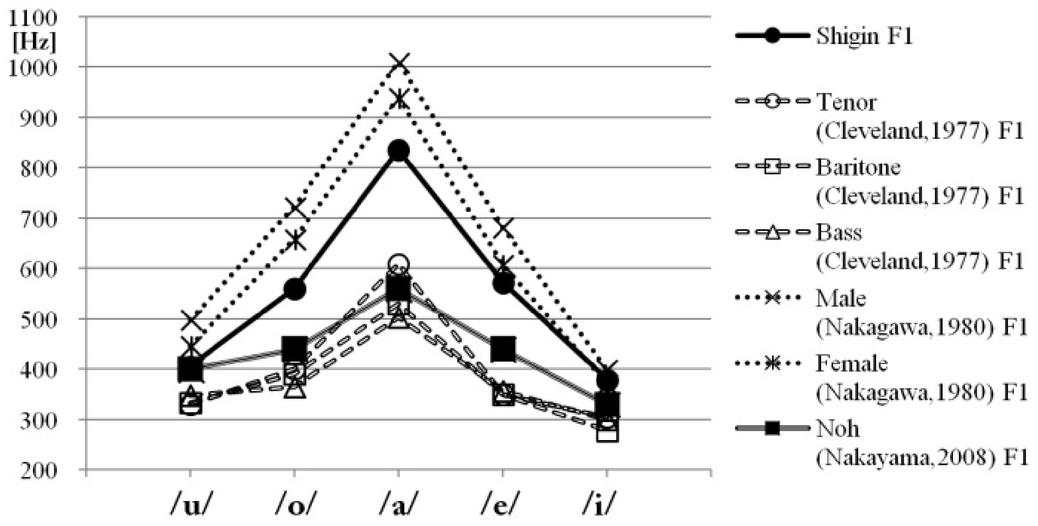
2.5.1. *Four-way ANOVA and linear prediction model of F_1, F_2, and F_3.* On the basis of previous studies on the formant frequencies of Western operatic singing [8], each of the three formant frequencies (F_1, F_2, and F_3) of the sung vowels was described as a function of *Vowel*, *Gender*, L_{Aeq}, and F_0. Hence, *Vowel*, *Gender*, L_{Aeq}, and F_0 were treated as explanatory variables, while F_1, F_2, and F_3 were regarded as objective variables. Consequently, a four-way ANOVA was performed. The linear prediction models for F_1, F_2, and F_3 employing the four variables *Vowel*, *Gender*, L_{Aeq}, and F_0 were also analyzed because each model is useful for the understanding of the shape of the vocal tract during phonation.

2.5.2. *Four-way ANOVA on F_4.* With reference to previous studies of Western operatic singing [17], the formant frequency F_4 of the sung vowels was described as a function of *Vowel*, *Singer*, L_{Aeq}, and F_0. A four-way ANOVA was performed by treating *Vowel*, *Singer*, L_{Aeq}, and F_0 as explanatory variables and F_4 as an objective variable.

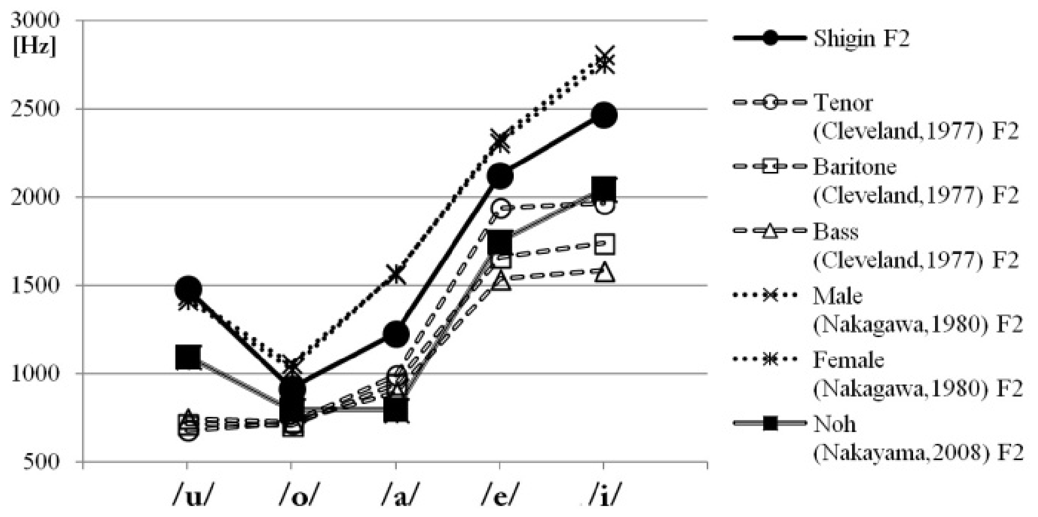
3. **Results and Discussions.** In order to answer research question Q1 posed in the Introduction, Figures 5 and 6 show distributions of the extracted F_1, F_2, F_3, and F_4 values of each of the five vowels in *Shigin* singing compared to those in Western operatic singing [17,18], choir singing and choir speech [19], the traditional Japanese singing style of *Noh* [9], and the normal speech mode in Japanese [20].

Figure 5(a) demonstrates that the F_1 values of *Shigin* singing were between those of the normal Japanese [20] speech mode and those of both Western operatic singing [17,18] and *Noh* singing [9]. The F_2 value of the /u/ vowel (Figure 5(b)) of *Shigin* singing was similar to that of the normal Japanese [20] speech mode but approximately 400 Hz and 800 Hz higher than that of *Noh* singing [9] and Western operatic singing, respectively [17,18]. Furthermore, the F_2 values of the other vowels (/o/, /a/, /e/, and /i/) of *Shigin* singing were between the values of the normal Japanese [20] speech mode and those of both Western operatic singing [17,18] and *Noh* singing [9]. In the case of F_3 (Figure 6(a)), the values of all five Japanese vowels of *Shigin* singing were approximately 400-500 Hz higher than those of Western operatic singing [17,18]. Figure 6(b) shows that the F_4 values of *Shigin* singing were approximately 1,000 Hz and 600-900 Hz higher than those of Western operatic singing [17,18] and choir singing [19], respectively. Based on the published results, a *Shigin* singer may phonate the vowel /u/ with their tongue in a position similar to that for the normal Japanese speech mode rather than that for *Noh* singing or Western operatic singing, such that F_2 normally increases when the tongue moves from a forward to backward position [21].

In order to answer research question Q2 posed in the Introduction, Tables 2, 3, and 4 list the four-way ANOVA results for F_1, F_2, and F_3, respectively. A full factorial ANOVA was also performed in which the main effects of each variable and the interactions between the variables were included in the model. The results for F_1, F_2, and F_3 all indicate that the *Vowel* was the most significant factor and *Gender* was the fourth significant factor. These two factors had the largest and fourth largest sum of squares, respectively (Tables 2(b), 3(b), and 4(b)). However, the factors with the second and third largest sum of squares were different for each formant frequency: the cross-effect L_{Aeq}*F_0 and F_0 were the second and third significant factors for F_1, respectively, F_0 and the cross-effect *Vowel**F_0 for F_2, respectively, and L_{Aeq} and F_0 for F_3, respectively. The results of the four-way ANOVA on F_4 are listed in Table 5, and a full factorial ANOVA was also performed in this case.



(a) F_1



(b) F_2

FIGURE 5. Distributions of the extracted (a) F_1 and (b) F_2 of the five vowels in *Shigin* singing compared with the distributions for the normal Japanese speech mode, Western operatic singing, and *Noh* singing

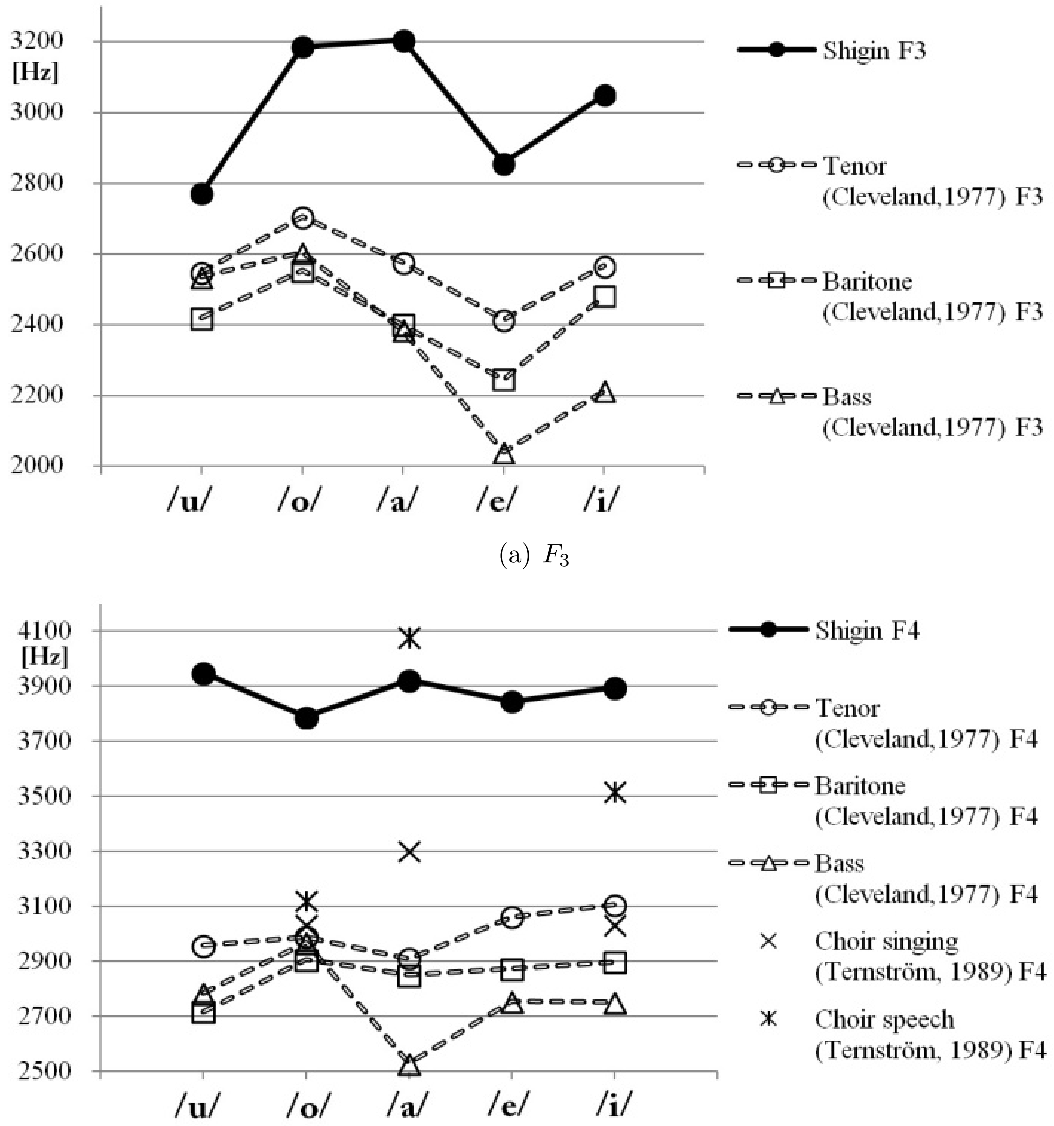
The four most significant factors, in order, were *Singer*, *Singer*Vowel*, $Vowel*L_{Aeq}$, and $Singer*L_{Aeq}$.

In order to answer research question Q3 posed in the Introduction, a simple linear equation was formulated as

$$F_{\{1,2,3\}} = a_0 + a_1(Vowel) + a_2(Gender) + a_3(L_{Aeq}) + a_4(F_0) \tag{2}$$

where a_0, $a_1(Vowel)$, $a_2(Gender)$, $a_3(L_{Aeq})$, and $a_4(F_0)$ are the values calculated from multiple regression analysis with dummy variables; these values were fitted to the data.

Table 6 lists the values of a_0, $a_1(Vowel)$, $a_2(Gender)$, $a_3(L_{Aeq})$, and $a_4(F_0)$ for F_1, F_2, and F_3. The R^2 values varied between 0.565 ($p < 0.001$) and 0.921 ($p < 0.001$). The a_2 coefficient shows that the F_1, F_2, and F_3 of female singers were 20-70 Hz higher than those of male singers. The a_3 coefficient indicates that a louder voice caused an increase in F_1 and a decrease in both F_2 and F_3. The *Shigin* singer may open their mouth and



(a) F_3

(b) F_4

Figure 6. Distributions of the extracted (a) F_3 and (b) F_4 of the five vowels in *Shigin* singing compared with the distributions for Western operatic singing and choir singing

jaw wider and move their tongue from a backward to forward position when producing a louder voice [8].

Finally, the a_4 coefficient demonstrates that a higher pitched voice caused an increase in F_1, F_2, and F_3, indicating the existence of formant tuning. *Shigin* singers may accurately adjust the opening of their mouth and jaw and the positioning of their tongue according to the targeted vowel.

4. **Conclusions.** In order to understand the articulations associated with changes in the shape of the vocal fold of traditional Japanese *Shigin* singers, this study investigated the formant frequencies of anechoic recordings of the quasi-steady-state portions of vowels phonated by six trained *Shigin* singers.

TABLE 2. Results of four-way ANOVA for F_1

(a) ANOVA results

Factor	DF	Sum of Squares	Mean square	F-ratio
Model	22	26,584,082	1,208,367	276.0734
Error	780	3,414,044	4,377	
Total	802	29,998,125		(Prob > F) < 0.0001

(b) Effect test results [a]

Factor	DF	Sum of Squares	Prob > F
Vowel	**4**	**7,669,540**	< **0.0001**
Gender	**1**	**370,199**	< **0.0001**
L_{Aeq}	1	225,304	< 0.0001
$\boldsymbol{F_0}$	**1**	**538,093**	< **0.0001**
*Vowel*Gender*	4	292,091	< 0.0001
*Vowel**L_{Aeq}	4	128,238	< 0.0001
*Vowel**F_0	4	79,181	0.0013
*Gender**L_{Aeq}	1	270,322	< 0.0001
*Gender**F_0	1	130,592	< 0.0001
$\boldsymbol{L_{Aeq}*F_0}$	**1**	**549,022**	< **0.0001**

[a]Four most significant factors are shown in bold.

TABLE 3. Results of four-way ANOVA for F_2

(a) ANOVA results

Factor	DF	Sum of Squares	Mean square	F-ratio
Model	22	274,839,947	12,492,725	516.1370
Error	780	18,879,338	24,204	
Total	802	293,719,284		(Prob > F) < 0.0001

(b) Effect test results [a]

Factor	DF	Sum of Squares	Prob > F
Vowel	**4**	**162,587,452**	< **0.0001**
Gender	**1**	**1,191,813**	< **0.0001**
L_{Aeq}	1	365,640	0.0001
$\boldsymbol{F_0}$	**1**	**1,441,123**	< **0.0001**
*Vowel*Gender*	4	530,506	0.0002
Vowel*$\boldsymbol{L_{Aeq}}$	**4**	**585,346**	< **0.0001**
*Vowel**F_0	4	1,256,313	< 0.0001
*Gender**L_{Aeq}	1	255,547	0.0012
*Gender**F_0	1	124,687	0.0235
L_{Aeq}*F_0	1	247,228	0.0014

[a]Four most significant factors are shown in bold.

The following three aspects were examined: (1) Does each of the four formant frequencies (F_1, F_2, F_3, and F_4) of the vowels (/u/, /o/, /a/, /e/, and /i/) phonated by *Shigin* singers differ from those of other singing styles and those of the normal speech mode? (2) Can the F_1, F_2, F_3, and F_4 of the vowels phonated by *Shigin* singers be statistically modeled as a function of the singer's gender, phonated vowels, loudness level, and phonation frequency? (3) Can each of the models be described using a simple linear equation?

TABLE 4. Results of four-way ANOVA for F_3

(a) ANOVA results

Factor	DF	Sum of Squares	Mean square	F-ratio
Model	22	38,141,602	1,733,709	67.1817
Error	780	20,128,893	25,806	
Total	802	58,270,494		(Prob > F) < 0.0001

(b) Effect test results [a]

Factor	DF	Sum of Squares	Prob > F
Vowel	**4**	**16,309,348**	**< 0.0001**
Gender	**1**	**1,278,541**	**< 0.0001**
$\boldsymbol{L_{Aeq}}$	**1**	**1,439,901**	**< 0.0001**
$\boldsymbol{F_0}$	**1**	**1,309,296**	**< 0.0001**
*Vowel*Gender*	4	438,727	0.0021
*Vowel**L_{Aeq}	4	687,023	< 0.0001
*Vowel**F_0	4	784,349	< 0.0001
*Gender**L_{Aeq}	1	314	0.9122
*Gender**F_0	1	252,988	0.0018
L_{Aeq}*F_0	1	333,673	0.0003

[a]Four most significant factors are shown in bold.

TABLE 5. Results of four-way ANOVA for F_4

(a) ANOVA results

Factor	DF	Sum of Squares	Mean square	F-ratio
Model	50	41,815,179	836,304	33.1123
Error	752	18,992,972	25,257	
Total	802	60,808,150		(Prob > F) < 0.0001

(b) Effect test results [a]

Factor	DF	Sum of Squares	Prob > F
Singer	**5**	**5,626,861**	**< 0.0001**
Vowel	4	568,045	0.0002
L_{Aeq}	1	275,761	0.0010
F_0	1	464,196	< 0.0001
Singer* Vowel	**20**	**4,537,345**	**< 0.0001**
Singer* $\boldsymbol{L_{Aeq}}$	**5**	**948,181**	**< 0.0001**
*Singer**F_0	5	659,795	0.0001
Vowel* $\boldsymbol{L_{Aeq}}$	**4**	**1,020,305**	**< 0.0001**
*Vowel**F_0	4	833,034	< 0.0001
L_{Aeq}*F_0	1	317,063	0.0004

[a]Four most significant factors are shown in bold.

The results of this study led to the following conclusions.

1) Each of the four formant frequencies (F_1, F_2, F_3, and F_4) of the vowels (/u/, /o/, /a/, /e/, and /i/) phonated by *Shigin* singers differed from those of other singing styles and those of the normal speech mode.

Table 6. Values of a_0, $a_1(Vowel)$, $a_2(Gender)$, $a_3(L_{Aeq})$, and $a_4(F_0)$ in Equation (2); $R^2 = 0.830$ ($p < 0.001$) for F_1, $R^2 = 0.921$ ($p < 0.001$) for F_2, and $R^2 = 0.565$ ($p < 0.001$) for F_3

Factor (Item)	Category	Coefficient		
		F_1	F_2	F_3
a_0		2.77×10^2	1.92×10^3	3.50×10^3
$a_1(Vowel)$	/u/	0	0	0
	/o/	8.03×10^0	-7.13×10^2	1.88×10^2
	/a/	2.81×10^2	-3.67×10^2	2.46×10^2
	/e/	2.05×10^1	4.82×10^2	-1.39×10^2
	/i/	-1.70×10^2	7.97×10^2	-4.79×10^0
$a_2(Gender)$	Male	0	0	0
	Female	2.04×10^1	4.37×10^1	7.00×10^1
$a_3(L_{Aeq})$		9.41×10^{-1}	-8.06×10^0	-9.21×10^0
$a_4(F_0)$		6.92×10^{-1}	1.05×10^0	5.81×10^{-1}

2) The F_1, F_2, and F_3 of the vowels phonated by *Shigin* singers were statistically modeled as a function of the singer's gender, phonated vowels, loudness level, and phonation frequency.
3) Each of the models was described using a simple linear equation.

These conclusions may depend on the specific conditions of this study. They may be influenced by an increase or decrease in the number of singers who participated in the experiment. In this study, the phonation frequencies of the singers were between 152 and 433 Hz. The results may be affected if the singers had phonated lower/higher pitched vowels. Further investigations are needed to better understand these factors.

These findings are expected to be useful for synthesis of the singing voice, development of training support systems, and understanding of the voice production mechanisms of singers.

Acknowledgement. The authors thank Prof. Takashi Yano and Dr. Keiji Kawai of Kumamoto University for providing access to a database of anechoic recordings of *Shigin* phonations. The authors thank Mr. Syouhei Mizobuchi of Wakayama University for building a basic program for the acoustic analysis and also thank Prof. Seiichi Nakagawa of Toyohashi University of Technology for providing reference data on the formant frequencies in Japanese vowels with the corresponding age and gender information. A part of this work was presented at the Stockholm Music Acoustics Conference 2013 [22,23]. The authors are grateful to Professor Sten Ternström and an anonymous SMAC 2013 reviewer who suggested the valuable comments.

REFERENCES

[1] B. H. Story, I. R. Titze and E. A. Hoffman, Vocal tract area functions from magnetic resonance imaging, *J. Acoust. Soc. of Am.*, vol.100, no.1, pp.537-554, 1996.
[2] J. Stoffers, R. C. Neuschaefer and M. Kob, Comparison of vocal tract resonance characteristics using LPC and impedance measurements, *Acta Acust. Acust.*, vol.92, no.5, pp.689-699, 2006.
[3] T. Vampola, J. Horacek, A. M. Laukkanen and J. G. Svec, Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement, *Logoped. Phoniatr. Vocol.*, vol.40, no.1, pp.14-23, 2013.

[4] E. Joliveau, J. Smith and J. Wolfe, Vocal tract resonances in singing: The soprano voice, *J. Acoust. Soc. Am.*, vol.116, pp.2434-2439, 2004.
[5] G. Fant, *The Acoustic Theory of Speech Production*, Moulton, 1960.
[6] I. R. Titze, *Principles of Voice Production*, Prentice Hall, 1994.
[7] K. Stevens, *Acoustic Phonetics*, MIT Press, 1998.
[8] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, DeKalb, 1987.
[9] I. Nakayama and M. Yanagida, Introduction to database of traditional Japanese singing with examples of comparative studies on formant shifts and vibrato among genres, *Acoust. Sci. Technol.*, vol.29, no.1, pp.58-65, 2008.
[10] K. Sakakibara, *Encyclopedia of Shigin Styles*, Kin-ensha, Tokyo, 1981 (in Japanese).
[11] K. Hiroshige, *Practice Text: Collection 1*, Koufuuryu Sanyou Gineikai, Iwakuni, 1972 (in Japanese).
[12] *Collection of Chinese Poetry with Ginkenshibudo Accents (Quatrain Edition)*, Nippon Ginkenshibu Foundation, Tokyo, 1977 (in Japanese).
[13] E. Prame, Vibrato extent and intonation in professional Western lyric singing, *J. Acoust. Soc. Am.*, vol.102, no.1, pp.616-621, 1997.
[14] C. H. Lee, On robust linear prediction of speech, *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol.36, no.5, pp.642-649, 1988.
[15] P. Boersma and D. Weenink, *Praat Doing Phonetics by Computer [Computer Program]*, Version 5.3.66, http://www.fon.hum.uva.nl/praat/(Last viewed March 2014), 2014.
[16] SAS Institute Inc., *JMP [Computer Program]*, Version 10, http://www.jmp.com/ (Last viewed March 2014), 2014.
[17] J. Sundberg, Articulatory interpretation of the 'singing formant', *J. Acoust. Soc. Am.*, vol.55, no.4, pp.838-844, 1974.
[18] T. F. Cleveland, Acoustic properties of voice timbre types and their influence on voice classification, *J. Acoust. Soc. Am.*, vol.61, no.6, pp.1622-1629, 1977.
[19] S. Ternström and J. Sundberg, Formant frequencies of choir singers, *J. Acoust. Soc. Am.*, vol.86, no.2, pp.517-522, 1989.
[20] S. Nakagawa, H. Shirakata, M. Yamao and T. Sakai, Differences in feature parameters of Japanese vowels with sex and age, *Stud. Phonol.*, vol.14, pp.33-52, 1980.
[21] B. E. F. Lindblom and J. E. F. Sundberg, Acoustical consequences of lip, tongue, jaw, and larynx movements, *J. Acoust. Soc. Am.*, vol.50, no.4B, pp.1166-1179, 1971.
[22] M. Nakayama, K. Kato and M. Matsunaga, Formant frequencies of sung vowels intonated by six traditional Japanese Shigin singers. Part I: Dataset construction and analysis method, *SMAC Stockholm Music Acoustics Conference 2013 (SMAC13)*, pp.264-268, 2013.
[23] K. Kato, M. Nakayama and M. Matsunaga, Formant frequencies of sung vowels intonated by six traditional Japanese Shigin singers. Part II: Results of analysis and statistical investigations, *SMAC Stockholm Music Acoustics Conference 2013 (SMAC13)*, pp.269-274, 2013.