

ANALYSIS AND MODELING OF ENTRAINMENT IN CHORUS SINGING

Motonari Kawagishi[†], Shota Kawabuchi[†], Chiyomi Miyajima[†], Norihide Kitaoka[†], and Kazuya Takeda[†]

[†]: Graduate School of Information Science, Nagoya University, Aichi, Japan

ABSTRACT

The dynamics of the contour of the fundamental frequency (F_0) of singing voices in a chorus is analyzed from the view point of 'entrainment' in singing behavior. The One-Mass-Two-Spring (OMTS) coupled system is used as the mathematical model of the contour of the F_0 of singing voices that are concurrently singing the same melody. Using this model, the characteristics of the F_0 dynamics of a voice singing in a chorus are parameterized by the mass, the coefficients of friction, and the spring factors of an OMTS system. It is experimentally confirmed that a steepest decent method can estimate the four model parameters, so that the model can generate the F_0 contour of a voice singing in a chorus with a less than 44.4 cents of RMS error. Preliminary experiments also show that experienced and novice singers can be correctly identified using the parameters of our model, because their entrainment behaviors are significantly different.

Index Terms— Singing voice, Fundamental frequency contour, Chorus, Entrainment, Second-order linear system

1. INTRODUCTION

Unlike solo singing, singing behavior in a chorus is affected by other singers, therefore, acoustical characteristics may differ from that of solo singing. In [1], Rossing reported that the intensity of singing-formants [2] are different in solo and chorus singing behaviors. This work also reported that singers adjust the power of their voices to adapt to the loudness of the voices of the other singers. These phenomena are the consequence of the fact that controlling singing behavior becomes more difficult when the voices of other singers' voice are louder [3]. Other research has studied individual preferences such as the Self-to-Other Ratio (SOR) in chorus singing [4, 5]. Such differences in singing behavior are an important issue for musical signal processing, particularly for generating natural sounding chorus voices.

Such adaptive changes can be related to 'entrainment' behaviors [6, 7, 8], which have been studied in relation to various aspects of musical signal processing. Entrainment can be defined as changes in behavior that results in movement towards synchronization of two or more phenomena. In [9], the authors analyzed the interaction between the breathing timing of a performer and an audience. In [10], synchrony of the playing rhythm of two piano players was analyzed and modeled. However, there has been little research done on modeling the singing behavior of choruses.

The purpose of this study is to build a generative model that characterizes the singing behavior of a chorus in a mathematical form. As the first step, a model of F_0 dynamics, i.e., the dynamic properties of the fundamental frequency of a voice singing in a chorus is studied. The dynamic characteristics of the F_0 contour of a

singing voice play a major role in musical expressions such as vibrato [11, 12] and overshoot [13]. It is also known that the dynamics of the F_0 affects the perceptual impression of listeners [14, 15]. Modeling the dynamics of the F_0 is one of the most important issues in understanding and characterizing singing behavior.

In a previous work [16], the authors studied the F_0 dynamics of singing voices from the viewpoint of the behavior of the F_0 around attractors, using the joint distribution of the F_0 and its derivatives. A Gaussian Mixture Model (GMM) was used to parameterize the distribution:

$$(F_0, \dot{F}_0) \sim \sum_i \omega_i \mathcal{N}(F_0, \dot{F}_0; \bar{\mu}_i, \bar{\sigma}_i^2) \quad (1)$$

This work also makes use of the idea that the dynamics of the F_0 of a singing voice can be characterized by the trajectory of the F_0 in the phase plane around the attractor, but we extend it to chorus singing using a more explicit mathematical form which models the physical coupling system.

The mass-spring system is a simple model of second-order dynamics which has been used for modeling the contour of the F_0 of speech [17]. This model has also been used for the analysis of solo singing voices [14, 18, 19, 20]. A mass-spring system which can model the dynamics of the F_0 of singing voices can be represented by a second-order differential equation:

$$\alpha \frac{d^2 y(t)}{dt^2} + \beta \frac{dy(t)}{dt} + \gamma y(t) = u(t) \quad (2)$$

where $u(t)$ is external force and $y(t)$ is the displacement of the mass. When we apply this equation to F_0 dynamics, $u(t)$ can represent the contour of the 'target' F_0 of the musical score and $y(t)$ can represent the contour of the sung F_0 . α , β and γ control the stability and musical expression of the singing voice. The challenge faced in this work is to extend this basic mass-spring system to the coupling system, i.e., to the One-Mass-Two-Spring (OMTS) system, in order to model the F_0 of singing voices affected by entrainment.

The rest of this paper consists of four sections. In the next section, data recording and analysis of F_0 dynamics are discussed. In that section, we compare the F_0 frequencies in solo and chorus singing voices and show that the F_0 shifts toward that of accompanying voices. In Section 3, we introduce the OMTS model to represent the dynamics of the F_0 of singing voices, and discuss the parameter estimation methods of the model. Section 4 describes the experimental evaluations of the model, and we show that the proposed model can generate the contour of the F_0 of the voices singing in a chorus with an RMS error rate of less than 44.4 cents. It is also shown that experienced and the novice singers can be correctly identified using the model parameters. Based on those results, the effectiveness of the proposed model for characterizing individual singing behavior in a chorus is clarified. Section 5 summarizes the paper.

Table 1. Signal analysis conditions for F_0 estimation.

Signal sampling freq.	16 kHz
Quantization bit rate	16 bits
window function	Hanning window
F_0 estimation window length	64 ms
window shift	10 ms
F_0 contour smoothing	30 ms

Table 2. Average μ and standard deviation σ of $d(t)$

	μ [cent]		σ [cent]	
	Solo	Chorus	Solo	Chorus
Novice	25.18	22.81	38.13	37.27
Experienced	12.72	7.93	34.93	33.08

2. MELODIC CONTOURS OF CHORAL SINGING

2.1. Data recording and analysis

In this paper, we discuss the entrainment of singing behavior, the way in which voices singing the same melody adapt to and move towards one another, from the viewpoint of the dynamics of the F_0 , i.e., F_0 entrainment. In order to analyze F_0 entrainment, we recorded the singing voices of six subjects singing in unison. Three of six subjects were members of the university chorus club, and had had 3-7 years of vocal training (Experienced). The other three had no formal singing experience (Novice). The bass part of an old, Japanese folk song, 'Fu-ru-sa-to' (My Hometown), is used for the recording. Each singer sang the song three times, under both the solo and chorus conditions. First, the voices of the singers singing solo, but with instrumental accompaniment, were recorded. Next, they sang as a chorus, using a prerecorded vocal of an experienced singer singing the same melody, as well as an instrumental accompaniment. The prerecorded vocal was the singing voice of another experienced subject, who sang along with a professional singer's vocal and his instrumental accompaniment. Both the accompanying vocal and the instrumental accompaniment were heard through headphones. In total, six subjects (three experienced singers and three novices) each sang the song six times, and songs which were sung with vocal accompaniment were used as samples to analyze for F_0 entrainment.

2.2. Analysis of F_0 entrainment

The contour of the F_0 of each singing signal is extracted using TANDEM-STRAIGHT [21]. The conditions for F_0 extraction are listed in Table 1. In this paper, F_0 is represented in logarithmic frequency units [cent], given by the following derivation:

$$f \text{ [cent]} = 1200 \log_2 \frac{f \text{ [Hz]}}{440 \times 2^{\frac{3}{12} - 5}} \quad (3)$$

A musical semitone is 100 in [cent]. In Figure 1, an example of extracted F_0 contours of solo and chorus singing are depicted in the time and phase domains [16, 22]. In the figures, the F_0 contours of the original musical score and that of the accompanying vocal, which are denoted U and V , respectively, are also plotted. From the phase plane representation, it can be observed that the center of the vortex,

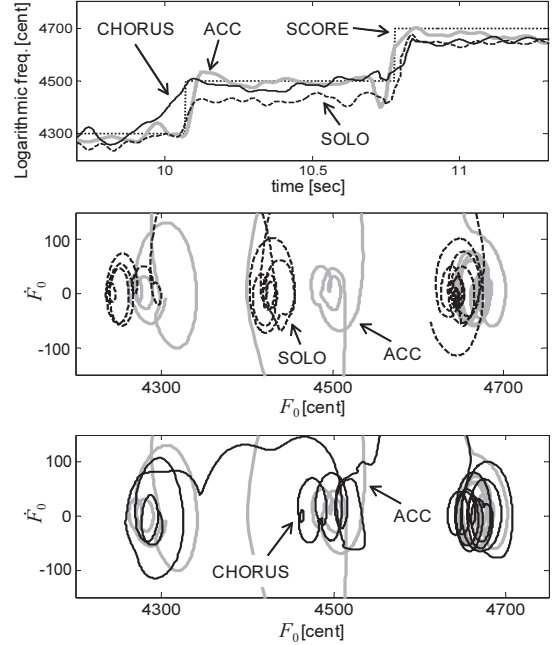


Fig. 1. Melodic contours (top) and corresponding phase plane for F_0 - \dot{F}_0 in solo and chorus singing (middle and bottom). During chorus singing, the attractor position shifts closer to that of the accompanying vocal. Dotted line: original musical score (SCORE); solid gray line: accompanying vocal (ACC); dashed line: solo singing (SOLO); solid black line: chorus singing (CHORUS).

i.e., the attractor, shifts closer to that of the accompanying vocal in the chorus signal. This shift can be regarded as a result of typical F_0 entrainment in chorus singing. In order to discuss a change in F_0 in the presence of an accompanying vocal, we calculate the difference between the F_0 of the prerecorded accompanying vocal and the sung vocal, under both the solo and chorus conditions:

$$d(t) = V(t) - F_0(t) \quad (4)$$

The average and the standard deviation of $d(t)$ are listed in Table 2. Statistical significance at a 1% critical rate in the difference between solo and chorus versions for both mean and variance were confirmed using a t-test and an f-test. From this result, it is clear that the accompanying vocal affects singing behavior. Since the variance of $d(t)$ decreases in chorus singing, we can conclude that the F_0 of a singer gets closer to that of the accompanying vocal statistically.

3. MODELING F_0 DYNAMICS OF CHORUS SINGING

3.1. One-Mass-Two-Spring (OMTS) model of chorus singing

In the previous section, the statistical change in the height of the F_0 in chorus singing as it gets closer to that of accompanying vocals is confirmed. In this section, we further analyze the dynamic characteristics of F_0 behavior. For this purpose, we use the one-mass-two spring (OMTS) system as a simple coupling system, as shown in

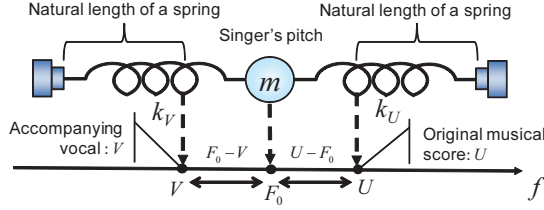


Fig. 2. One-Mass-Two-Spring (OMTS) model of chorus singing

Figure 3. The equation of motion in this system is given by:

$$m \frac{d^2 F_0}{dt^2} = -\lambda \frac{dF_0}{dt} + k_U (U - F_0) + k_V (V - F_0) \quad (5)$$

where m represents the weight of the mass, λ is the coefficient of friction and k_U and k_V are constants of springs connected to the vibrators corresponding to the F_0 contour of the original score and that of the accompanying vocal, respectively. A new term, expressing the influence of the accompanying vocal, $k_V (V - F_0)$, is added to equation (2), the model representing the F_0 dynamics of the solo singing voice. The stiffness of each of the two springs represents the strength of the coupling with the musical score and the accompanying vocal, respectively. In other words, the larger k_V is, the more chorus singing is affected by the accompanying vocal. Therefore, a set of model parameters, $\Theta = \{m, \lambda, k_U, k_V\}$, characterizes the behavior of F_0 entrainment in chorus singing.

3.2. Estimating the OMTS parameters

As one of several methods for identifying the model parameters for the observed signals, we implemented the steepest descent method as follows. The square error given by equation (6) is used for the objective function, i.e., $J(\Theta)$:

$$J(\Theta) = \frac{1}{N} \sum_{n=1}^N (F_0(t_n) - y(t_n, \Theta))^2 \quad (6)$$

where $y(t)$ is a function which is given by solving equation (5). The repeated update of Θ can be represented by:

$$\Theta^{(\tau+1)} = \Theta^{(\tau)} - \eta \frac{\partial J(\Theta^{(\tau)})}{\partial \Theta} \quad (7)$$

where

$$\frac{\partial J(\Theta)}{\partial \Theta} = -\frac{2}{N} \sum_{n=1}^N (F_0(t_n) - y(t_n, \Theta)) \frac{\partial y(t_n, \Theta)}{\partial \Theta} \quad (8)$$

where η is a vector comprised of a learning rate parameter for each model parameter. The derivative of the error function with respect to Θ can be calculated by solving equation (5) under the given parameter set, Θ . Equation (5) can be solved for the given parameters by Fourier series expansion and Laplace transform as follows.

First, rewrite equation (5) in the following form:

$$\ddot{y}(t) + \alpha \dot{y}(t) + (\beta_U + \beta_V) y(t) = \beta_U U(t) + \beta_V V(t) \quad (9)$$

where α , β_U , and β_V are the normalized values of the coefficient of friction and spring factors by the mass. By expanding $U(t)$ and $V(t)$ in Fourier series,

$$U(t) = u_0 + \sum_{k=1}^{\infty} (u_{1k} \cos k\omega_0 t + u_{2k} \sin k\omega_0 t) \quad (10)$$

$$V(t) = v_0 + \sum_{k=1}^{\infty} (v_{1k} \cos k\omega_0 t + v_{2k} \sin k\omega_0 t) \quad (11)$$

and substituting to equation (9), we obtain the differential equation,

$$\begin{aligned} \ddot{y}(t) + \alpha \dot{y}(t) + (\beta_U + \beta_V) y(t) \\ = c_0 + \sum_{k=1}^{\infty} (c_{1k} \cos k\omega_0 t + c_{2k} \sin k\omega_0 t), \end{aligned} \quad (12)$$

where c_k are the weighting sums of the Fourier coefficients, i.e., $c_k = \beta_U u_k + \beta_V v_k$. The solution of equation (12) is given as:

$$y(t) = \mathcal{L}^{-1} [\{(s + \alpha)y(0) + \dot{y}(0) + R(s)\} Q(s)] \quad (13)$$

where $R(s)$ is the Laplace transform of the right hand side of equation (12) and $Q(s)$ is given by:

$$Q(s) = \frac{1}{s^2 + \alpha s + (\beta_V + \beta_U)}. \quad (14)$$

The solution of the equation (12) is finally given by the following equation:

$$\begin{aligned} y(t) = & \frac{1}{\xi_1 - \xi_2} \left\{ (y(0)\xi_1 + \dot{y}(0) - \alpha y(0))e^{\xi_1 t} \right. \\ & \left. - (y(0)\xi_2 + \dot{y}(0) - \alpha y(0))e^{\xi_2 t} \right\} \\ & + \frac{c_0}{\xi_1 \xi_2} + \frac{c_0 e^{\xi_1 t}}{(\xi_1 - \xi_2)\xi_1} + \frac{c_0 e^{\xi_2 t}}{(\xi_2 - \xi_1)\xi_2} \\ & + \sum_{k=1}^{\infty} \left\{ \frac{(\xi_1 c_{1k} + k\omega_0 c_{2k})e^{\xi_1 t}}{(k^2\omega_0^2 + \xi_1^2)(\xi_1 - \xi_2)} + \frac{(\xi_2 c_{1k} + k\omega_0 c_{2k})e^{\xi_2 t}}{(k^2\omega_0^2 + \xi_2^2)(\xi_2 - \xi_1)} \right. \\ & + \frac{(\xi_1 \xi_2 - k^2\omega_0^2)c_{1k} + k\omega_0(\xi_1 + \xi_2)c_{2k}}{(k^2\omega_0^2 + \xi_1^2)(k^2\omega_0^2 + \xi_2^2)} \cos k\omega_0 t \\ & \left. - \frac{(k^2\omega_0^2 - \xi_1 \xi_2)c_{2k} + k\omega_0(\xi_1 + \xi_2)c_{1k}}{(k^2\omega_0^2 + \xi_1^2)(k^2\omega_0^2 + \xi_2^2)} \sin k\omega_0 t \right\} \end{aligned} \quad (15)$$

where ξ_i is given by:

$$\xi_{1,2} = \frac{\alpha \pm \sqrt{\alpha^2 - 4(\beta_U + \beta_V)}}{2}.$$

By repeatedly the updating Θ we can finally reach the optimal estimate of the parameters for the given signals, i.e., F_0 , U , and V . Θ was updated 100 times based on the steepest descent method, because it was confirmed experimentally that Θ converges after about 30 iterations.

4. EXPERIMENTAL EVALUATION

4.1. Parameter Estimation

In order to test the OMTS model's ability to characterize individuals in entrainment during chorus singing, we evaluated the model experimentally. The test data consisted of the first 17 seconds of the

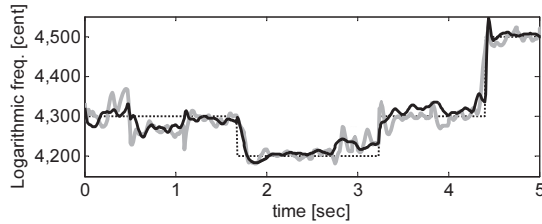


Fig. 3. An example of the actually sung signal $F_0(t)$ and F_0 contour $y(t)$ derived from the estimated model parameters. Dotted line: original musical score; solid gray line: $F_0(t)$; solid black line: $y(t)$.

Table 3. Estimated model parameters and RMSE

	α	β_V	β_U	$\frac{k_V}{k_U + k_V}$	RMSE [cent]
Novice	1.53	1.27	1.21	0.51	44.4
Experienced	1.41	1.29	1.37	0.48	39.4

Note: RMSE is calculated between the sung signal and the signal derived from the estimated model parameters. $k_V/(k_U + k_V)$ is the ratio of the spring constant calculated using β_U, β_V and represents the adaptation ratio to the accompanying vocal.

contour of the F_0 as described in Section 2.1. The number of iterations in the steepest decent algorithm is fixed at 100. An example of the F_0 contour derived from the estimated model parameters is shown in Figure 3, together with that of the signal actual sung. For this evaluation, the model parameters were estimated using the three recordings of the chorus versions sung by each subject. Two recordings were used to train the model and the other one was used for evaluation (trial-open, song-close, and singer-close paradigm).

The averaged values of the estimated model parameters of experienced and novice subjects are listed in Table 3. Each average is calculated using nine signals sung by three subjects. We can see that the spring factor which connects the mass to the source of vibration in our model (corresponding the accompanying vocal) is higher in novice subjects than in experienced subjects. Although statistical significance is not tested, this result suggests that novice subjects are more sensitive to the accompanying vocal than experienced singers. It is also suggested, by the difference in the coefficient of friction, α , that experienced subjects use more dynamic expression than novice subjects.

4.2. Singer Discrimination

Singing behavior in a chorus differs between subjects, and it may be possible to characterize the differences using OTMS model parameters. In this section, we discuss the individuality of singing behavior by evaluating the root mean square error (RMSE) of the F_0 contour generated using the OTMS model with respect to the observed F_0 signal. Since we have three signals for each singer, we used a three-fold-trial paradigm where the model parameters were estimated using two signals and the remaining signal is used for calculating the RMSE value.

The results are listed in Table 4. The OTMS model that gave the

Table 4. Calculation results of RMSE

	No-A	No-B	No-C	Ex-A	Ex-B	Ex-C
No-A	69.26	71.01	69.97	70.90	69.69	69.47
No-B	69.88	68.53	68.58	74.15	72.07	71.46
No-C	51.55	51.08	51.62	56.05	53.61	52.78
Ex-A	50.46	56.22	54.05	48.55	48.94	49.15
Ex-B	53.20	57.96	55.64	53.55	52.30	52.27
Ex-C	46.99	51.66	49.04	48.05	46.39	46.36

Note: “ No ” and “ Ex ” represent Novice and Experienced singers, respectively. Horizontal and vertical labels are the subject IDs of the models and test signals, respectively. The lowest error value is expected to occur when a singer’s test signal is compared to the same singer’s model.

minimum RMSE is highlighted using bold fonts. As shown in the table, four out of six subjects had the minimum RMSE to our model as compared to the individual model. Even in Novice C (No-C) and Experienced B (Ex-B), whose smallest RMSE are not achieved by their individual models, a relatively lower RMSE value is achieved by the individual models. It is also confirmed that experienced and novice subjects can be correctly identified based on the RMSE. In addition, we confirmed statistical significance in the difference in the mean between experienced and novice subjects of the model at a 5% critical rate using a t-test. From these results, it is confirmed that the OTMS model can properly characterize individual singers by analyzing F_0 entrainment during chorus singing.

5. SUMMARY AND FUTURE WORK

In this paper, we proposed a One-Mass-Two-Spring model for describing the F_0 entrainment of chorus singing behavior. We introduced a steepest decent method for estimating the model parameters from the observed signals. Four out of six subjects were correctly identified using the model parameters. Experienced and novice subjects were also correctly identified. Therefore, the effectiveness of the model as a parametric representation of F_0 entrainment is strongly suggested.

However, future work is needed for applying the model to wider applications. First of all, we need to confirm the effectiveness of the model under bilateral interaction, because, in this study, experimental conditions were limited to cases where each subject sang along with a ‘ pre-recorded ’ accompanying vocal. In order to evaluate the general effectiveness of the model, we also need to extend the size of the experiments, in regards to both the number of subjects and the number of songs.

Acknowledgements: This work has been partially supported by the Grant-in-Aid for Challenging Exploratory Research (23650088).

6. REFERENCES

- [1] T. D. Rossing, J. Sundberg, and S. Ternstrom, "Acoustic comparison of voice use in solo and choir singing," *Acoustical Society of America*, Vol. 79, pp. 1975–1981, 1986.
- [2] Johan Sundberg, *The Science of the Singing Voice*. the Northern Illinois University Press, 1987.
- [3] S. Ternstrom and J. Sundberg, "Intonation precision of choir singing," *Acoustical Society of America*, Vol. 84, pp. 59–69, 1988.
- [4] S. Ternstrom, "Preferred self-to-other ratios in choir singing," *Acoustical Society of America*, Vol. 105 pp. 3563–3574, 1999.
- [5] S. Ternstrom and J. Sundberg, "Self-to-other ratios measured in an opera chorus in performance," *Acoustical Society of America*, Vol. 118, pp. 3903–3911, 2005.
- [6] J. Buck and E. Buck, "Mechanism of Rhythmic Synchronous Flashing of Fireflies," *Science*, Vol.159, pp. 1319–1327, Mar. 1968.
- [7] T. J. Walker, "Acoustic Synchrony: Two Mechanisms in the Snowy Tree Cricket," *Science*, Vol.166, pp. 891–894, Nov. 1969.
- [8] Z. Neda, E. Ravasz, Y. Brechet, T. Vicsek, and A.L. Barabasi, "The sound of many hands clapping," *Nature*, Vol.403, pp. 849–850, 2000.
- [9] T. Yamamoto and Y. Miyake, "Analysis of interaction in musical communication and its modeling," *In Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC 2000)*, Vol. 2, pp. 763–768, 2000.
- [10] Y. Kobayashi and Y. Miyake, "New ensemble system based on mutual entrainment," *In Proc. IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2003)*, pp. 235–240, 2003.
- [11] C. E. Seashore, "A musical ornament, the vibrato," *Psychology of Music*, McGraw-Hill Book Company, pp. 33–52, 1938.
- [12] H. B. Rothman and A. A. Arroyo, "Acoustic variability in vibrato and its perceptual significance," *In Proc. Psychology of Music*, Vol. 1, No. 2, pp. 123–141, 1987.
- [13] G. de Krom and G. Bloothoof, "Timing and accuracy of fundamental frequency changes in singing," *ICPhS*, 1995, pp. 206–209.
- [14] T. Saitou, M. Goto, M. Unoki, and M. Akagi, "Speech-To-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices," *WASSPA*, pp. 215–218, 2007.
- [15] T. Saitou and M. Goto, "Acoustic and Perceptual Effects of Vocal training in Amateur Male Singing," *International Conference on Spoken Language Processing (INTERSPEECH 2009)*, pp. 832–835, 2009.
- [16] Y. Ohishi, M. Goto, K. Itou, and K. Takeda, "A Stochastic Representation of the Dynamics of Sung Melody," *In Proc. International Conference on Music Information Retrieval (ISMIR 2007)*, pp. 371–372, Sept. 2007.
- [17] H. Fujisaki, "A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour," *Vocal Physiology:Voice Production, Mechanisms and Functions*, (O.Fujimura, ed.), Raven Press, pp. 347–355, 1988.
- [18] N. Minematsu, B. Matsuoka, and K. Hirose, "Prosodic Modeling of Nagauta Singing and Its Evaluation," *SpeechProsody 2004*, pp. 487–490, Sept. 2004.
- [19] H. Mori, W. Odagiri, and H. Hirose, "F0 dynamics in singing: Evidence from the data of a baritone singer," *IEICE Trans. Inf. and Syst.*, Vol.E87-D, No.5, pp. 1086–1092, 2004.
- [20] Y. Ohishi, H. Kameoka, D. Mochihashi, and K. Kashino, "A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components," *International Conference on Spoken Language Processing (INTERSPEECH 2012)*, Sept. 2012.
- [21] H. Kawahara, M. Morise, T. Takahashi, R. Nishimura, T. Irino, and H. Banno, "Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation," *ICASSP 2008*, pp. 3933–3936, April. 2008.
- [22] T. Kako, Y. Ohishi, H. Kameoka, K. Kashino, and K. Takeda, "Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane," *International Conference on Music Information Retrieval (ISMIR 2009)*, pp. 393–397, Oct. 2009.