Physiological Observations and Synthesis of Subharmonic Voices

Ken-Ichi Sakakibara*, Hiroshi Imagawa† , Hisayuki Yokonishi† , Miwako Kimura‡, and Niro Tayama§
* Department of Communication Disorders, Health Sciences University of Hokkaido, Sapporo, Japan
E-mail: kis@hoku-iryo-u.ac.jp
† Department of Otolaryngology, The University of Tokyo, Tokyo, Japan
‡ Tokyo Voice Center, Sanno Hospital, International University of Health and Welfare, Tokyo, Japan
§ Department of Otolaryngology, Head and Neck Surgery, National Center for Global Health and Medicine, Tokyo, Japan

***Abstract*—A subharmonic voice is a voice generated by subharmonic vibrations which include integer-multiple periodicity, such as periodic-double and periodic-triple. In this paper, various types of subharmonic voices were observed using high-speed digital imaging. A new laryngeal source model for synthesis of subharmonic voices were proposed based on the results of physiological observations. The proposed model consists of several modules, such as dumping function, and amplitude modulator. Parameters of the model were estimated by referring to the glottal area function, kymograph, and laryngotopography obtained by analysis of high-speed images. Synthesized sounds were scaled in similarity to diplophonia and the R (roughness) parameter of the GRBAS scale in listening tests. As a result, the proposed model was effective for synthesis of subharmonic voices, such as diplohphonia and rough voices.**

I. Introduction

A subharmonic voice is defined as a voice which includes subharmonic frequential components or is generated by a subharmonic vibration. A subharmonic vibration is a vibration which includes vibratory modes with lower frequency than fundamental frequency in a ratio of $1/n$ where $n \in \mathbf{N}$. A subharmonic vibration in F_0/n is referred to as a period-n vibration.

Subharmonic voices are frequently observed in pathological voices, such as diplophonia [11]. Even in non-pathological voices, subharmonic voices are found under the boundary conditions for a vocal fold oscillation, such as of low-power or high-power oscillation. For example, subharmonic vibratory patterns are observed in vocal fray, as well as periodic and aperiodic vibratory patterns [14]. Furthermore subharmonic voices are also found in some singing styles of folk music, such as Tyvan khöömei [5], [18], South African umngqokolo [15], Sardinian chants [1], and so on. In khöömei and Sardinian chants, subharmonic voices are generated by vocal-ventricular phonations, in which the vocal and ventricular folds simultaneously vibrate and the ventricular folds vibrate in F_0/n. In growl, such as umngqolo, subharmonic voices are generated by vocal-aryepiglottic phonations, in which the vocal and aryepiglottic folds simultaneously vibrate and the aryepiglottic folds vibrate in F_0/n.

In this paper, we observe vocal fold vibratory patterns of subharmonic voices of pathological voices using high-speed digital imaging (HSDI) and analyze physical mechanisms of generation of subharmonic voices. We also propose a new laryngeal source model for pathological and non-pathological subharmonic voices based on a many-parameter model for laryngeal source in [16]. A method of parameter estimation is processed based on physiological observations. We also evaluate effectiveness of the model for synthesis of subharmonic voices by listening tests.

II. Physiological Observeations of Subharmonic Voices in Pathological Cases

For synthesis of subharmonic voices, we observe and analyze vocal fold vibratory patterns of pathological voices which include subharmonic vibrations using HSDI. We extract useful information for the parameter estimation of a laryngeal source model from a glottal area function , glottal width, and vibratory modes, and synthesize subharmonic voices in the framework of source-filter formant synthesis [4].

The high-speed digital camera Photoron, FASTCAM-1024PCI at a frame rate 4500 fps, image resolution of 400×512 pixels, 8bit grayscale, and memory size of 12 GB allowing sampling duration of 11.1 s, was employed. A rigid endoscope (# 4450.501, Richard Wolf) was attached to an attatchment lens ($f = 35$ mm, Nagashima Med. Instrument Corp.) connected to the camera. HSDI (high-speed digital images) were simultaneouosly recorded with EGG (electroglottography) and sound signals.

A. Vocal fold paralysis

Fig. 1 shows the multi-line kymograph, sound waveform, and EGG waveform of a patient with left vocal fold paralysis (recurrent nerve paralysis), female, age of 22, perceived as diplophonia.

The multi-line kymograph in Fig. 1 has five kymographs which are displayed in correspondence with red lines from top to bottom in the laryngeal view at the left. In the laryngeal view of the left, the anterior part of the larynx is seen at the bottom and the posterior is seen at the top. Therefore, the right vocal fold is seen at the left, and the left vocal fold is seen at the right of Fig. 1.

This research was partly supported by Japan and Grant-in-Aid (KAKENHI:20500161) from the MEXT, Japan.

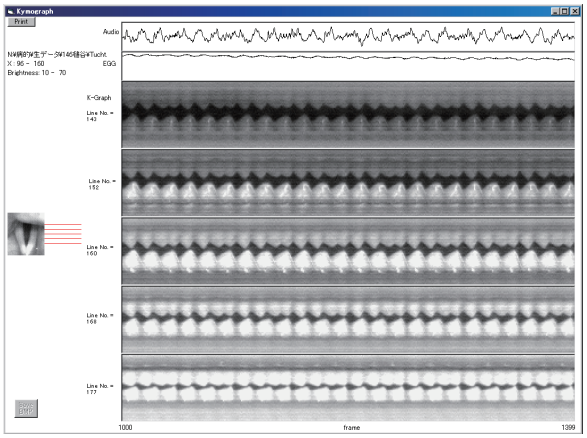

Fig. 1. Multi-line kymograph of HSDI of a patient with left vocal fold paralysis (recurrent nerve paralysis). Left: laryngeal view, Top of right: sound waveform, Middle of right: EGG waveform, Bottom of right: multi-line kymographs (five kymographs correspond with red lines in the laryngeal view of the left).

In kymograph, lines extracted from HSDI are rotated by -90 degree and laterally displayed along with time axis,. Therefore, the left vocal fold is seen on the top side and the right vocal fold is seen on the bottom side.

Fig. 2 shows an enlarged kymograph in Fig. 1. A ratio of the vibratory frequencies of the left vocal fold to the right vocal fold observed in Fig. 1 is 3:4. This means that the vocal fold vibration includes subharmonic vibrations. In fact, the left vocal fold vibrates in $F_0/4$ and the right vocal fold vibrates in $F_0/3$.

The amplitudes of the vibrations of the left and right vocal folds are dumped in each period. A white dashed line in Fig. 2 depicts the amplitude envelope of the left vocal fold vibration. The amplitude of the left vocal fold decreases in a period. The glottal closure of this case is incomplete.

The vibratory frequencies of the left and right vocal folds obtained by analyzing HSDI are 450 Hz and 360 Hz respectively.

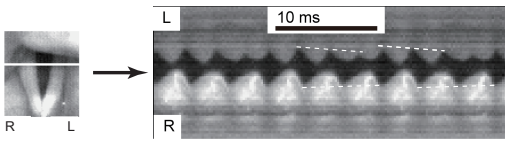

Fig. 2. Kymograph of a pathient with left vocal fold paralysis

B. Vocal fold cyst

Fig. 3 shows the multi-line kymograph, sound waveform, and EGG waveform of a patient with a cyst of the right vocal fold, female, age of 70, perceived as diplophonia and rough. The cyst is formed in the middle of the right vocal fold and prevents mucosal wave propagation.

In comparison with the case of the vocal fold paralysis in Fig. 1, differences of vibratory frequencies between the

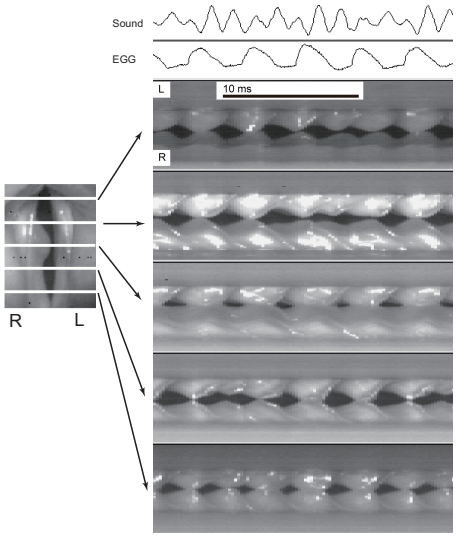

Fig. 3. Multi-line kymograph of HSDI of a patient with the right vocal fold cyst. Left: laryngeal view, Top of right: sound waveform, Middle of right: EGG waveform, Bottom of right: multi-line kymographs (five kymographs correspond with red lines in the laryngeal view of the left).

left and right vocal folds are unclearly observed from the kymography.

Fig. 4 shows the laryngotopograph of the same patient with the right vocal fold cyst.

Laryngotopography is a method for analyzing HSDI based on pixel-wise Fourier analysis of a time-varying brightness curve for each pixel across images [6], [17]. The procedure of analysis is as follows: (i) a rectangular area is selected; (ii) a time-varying raw brightness curve is extracted from the images (Figure 1); (iii) for each pixel, the average level of brightness of consecutive frames (512 or 256 frames) are calculated; (iv) the brightness curve is normalized by subtracting the average level from the original raw brightness curve; (v) The Hamming window is applied to the normalized brightness curve and discrete Fourier transform (DFT) of 1024 points by adding zero-padding to increase a frequency resolution is applied to the normalized brightness curve. When the high-speed imaging is conducted at 4500 fps, the frame size for analysis is 0.114 s (512 frames) or 0.057 s (256 frames) and the frequency resolution is 3.7 Hz.

The laryngotopgraphic image at (g) of Fig. 4 shows an amplitude spectrum of the brightness curve at the point slightly behind the anterior commissure. Three different significant peaks are observed in the spectrum. The first component is at 251 Hz, the second is at 313 Hz, and the third is at 379 Hz.

In Fig. 4, (a) and (b) show distributions of amplitude and phase of the component of 251 Hz, respectively. (c) and (d) show distributions of amplitude and phase of the component of 313 Hz, respectively. (e) and (f) show distributions of amplitude and phase of the component of 379 Hz, respectively.

In this case, the vocal fold vibration is mostly the superposition of three different vibratory modes of 251, 313, 379 Hz. More precisely, if the horizontal vibratory mode with degree

h and vertical vibratory mode with degree v is denoted by (h, v)-mode in [20], the vocal fold vibration of this case is the superposition of $(1, 0)$-mode in 251 Hz, $(2, 0)$-mode in 313 Hz, and $(1, 0)$-mode and weak $(3, 0)$-mode in 379 Hz.

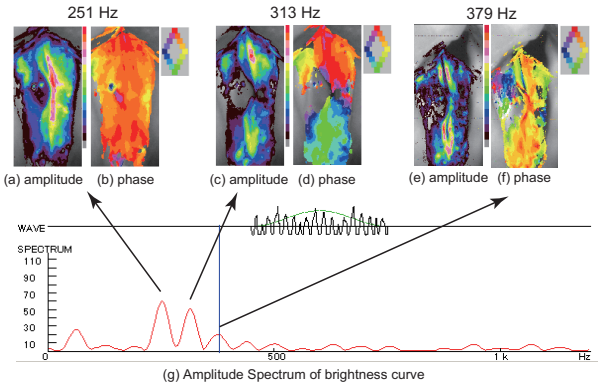

Fig. 4. Laryngotopograph of a patient with the right vocal fold cyst. Amplitudes and phases of selected frequencies. (a), (b) are amplitude and phase of the component of 251 Hz, (c), (d) are amplitude and phase of the component of 313 Hz, (e), (f) are amplitude and phase of the component of 379 Hz, respectively. (g): amplitude spectrum of the brightness curve at the point slightly behind the anterior commissure

C. Laryngeal papilloma

Fig. 5 shows the multi-line kymograph, sound waveform, and EGG waveform of a patient with left laryngeal papilloma, female, age of 28, perceived as diplophonia and rough.

Normally, the vocal folds vibrate mainly in $(1, 1)$-mode with phase difference between the upper and lower lips. However, in this case, the vocal fold vibration includes $(1, 0)$-mode in 454 Hz and $(0, 1)$-mode in 563 Hz.

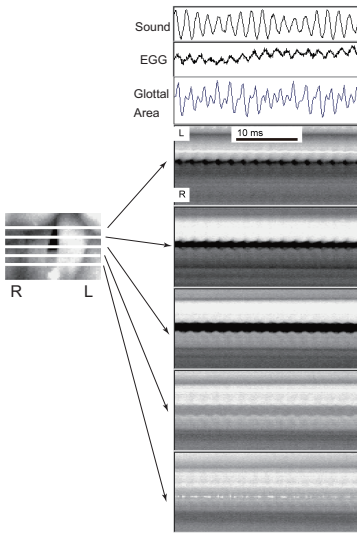

Fig. 5. Multi-line kymograph of HSDI of a patient with the right vocal fold papilloma. Left: laryngeal view, Top of right: sound waveform, Middle of right: EGG waveform, Bottom of right: multi-line kymographs (five kymographs correspond with red lines in the laryngeal view of the left).

III. Laryngeal Source Model for Subharmonic Voices

As we have seen, the vocal fold vibratory patterns of subharmonic voices, which are perceived as diplophonia and rough, are characterized by the following characteristcs:

1) incomplete glottal closure
2) dumping of the amplitude of the vocal fold vibration in each period (sub-period)
3) superposition of vibratory mode with different frequencies
4) modulation of the amplitude of the vocal fold vibration

For synthesis of subharmonic voices with the above characteristics, consecutive pulses are non-linearly connected and, therefore, it is impossible to synthesize subharmonic voices using a model with assumption of harmonic vibration of the vocal folds, such as R-model [12] and LF-model [3]. These harmonic models have been widely used for formant synthesis of voices. However, even though temporal modulations, such as jitter and shimmer, are added into the voice sources generated by the harmonic model, it is impossible to synthesize subharmonic voices which has non-linear connection between pulses.

A many parameter model of laryngeal source [16] is effective for synthesis of subharmonic voices. However, this model has many independent parameters and, in general, it is difficult to estimate these many parameters by acoustical analysis.

In this paper, we propose a method of parameter estimation for a many-parameter model based on physiological observations, such as kymography, glottal area function, and modal analysis using laryngotopography. In addition, we revise a many parameter model proposed in [16] along with the parameter estimation method.

A. Basic pulse function

Let $u[T, V, O_q, \theta](t)$ be a glottal souce model of KL-GLOTT88 [10]. Then,

$$u[T, V, O_q](t) := \begin{cases} at^2 - bt^3 & \text{if } 0 \le t \le TO_q \\ 0 & \text{if } TO_q \le t < T \end{cases} \qquad (1)$$

where

$$a = \frac{27V}{4{O_q}^2T^2}, \qquad b = \frac{27V}{4{O_q}^3T^3},$$

T: fundamental period, V: maximal amplitude, and O_q: open quotient.

By using an initial phase θ, the glottal source model is written by four parameters T, V, O_q, θ as follows:

$$u[T, V, O_q, \theta](t) \equiv u[T, V, O_q](t - t_0) \mod T \qquad (2)$$

$$\text{where} \qquad t_0 = \frac{\theta}{2\pi}T$$

B. Dumping function

From results of analysis of HSDI, the amplitude of the vocal fold vibration is almost linearly dumped in each subharmonic period. Then, a dumping function of the model is formulated as a sawtooth-like function written as follows:

$$S[T,V,r](t) := \begin{cases} \dfrac{V}{rT}t & \text{if } 0 \le t < rT \\ \dfrac{V}{1-r}\left(1-\dfrac{t}{T}\right) & \text{if } rT \le t < T \end{cases} \tag{3}$$

where V: a modulation index, T: a modulation period, and r: a ratio of increase interval to a period of a sawtooth function, such that $r \in (0,1)$. Hence, $S[T,V,r](t)$ increases in $0 \le t \le rT$ and decreases in $rT \le t \le T$ linearly.

Let θ be a phase parameter, a dumping function is defined as follows:

$$S[T,V,r,\theta](t) \equiv S[T,V,r]\left(t-\frac{\theta}{2\pi}T\right) \mod T \tag{4}$$

C. Amplitude modulator

A modulator of the vocal fold vibratory amplitude is defined by using sinusoidal function as follows:

$$M[T,V,\theta](t) := \left(1-\frac{V}{2}\right)+\frac{V}{2}\sin\left(\frac{2\pi}{T}t-\theta\right) \tag{5}$$

where T: a modulation period, V: a parameter related to modulation index, and θ: a phase parameter.

D. Structure of a source model

Fig. 6 shows a block diagram of the proposed model. $\bigoplus$ denotes addition of signals and $\bigotimes$ denotes multiplication of signals.

First, basic pulse functions of the left and right vocal folds are determined depending on the vibratory frequencies of the left and right vocal folds.

The basic pulse functions of the left and right vocal folds are denoted by $u[T_1,V_1,O_{q1},\theta_1]$, $u[T_2,V_2,O_{q2},\theta_2]$. If the left and right vocal folds vibrate in the same frequency by synchronization, only one function can be used as a basic pulse function. If the glottal closure is incomplete, u_o is added to the fundamental pulse functions.

If the amplitudes of the vocal fold vibrations are damped in each sub-period, then, the laryngeal source signal is dumped by dumping functions $S[T_3,V_3,\theta_3]$ and $S[T_4,V_4,\theta_4]$.

Furthermore, if different modes are superimposed, the laryngeal source signal is modulated by $M[T_5,V_5,\theta_5]$, $M[T_6,V_6,\theta_6]$, ...depending on the number of the superimposed modes and their frequencies, amplitudes, and phase differences.

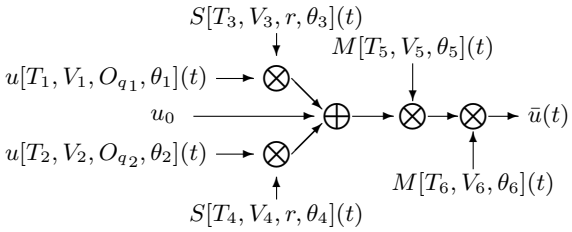

Fig. 6. A block diagram of the proposed model

IV. Parameter estimation and synthesis of a laryngeal source

The parameters in Fig. 6 are mathematically independent and the number of the parameters is more than twenty. Obviously, it is not easy to determine suitable values of the parameters to synthesize a target voice.

In this paper, we propose a new method of parameter estimation of the many-parameter model of laryngeal source based on analysis of HSDI.

The parameter estimation of the model is processed by the following steps:

- Step 1. If the left and right vocal folds vibrate in different frequencies, two different basic pulse functions are used. Otherwise, only one basic pulse function is used.
- Step 2. If the glottal closure is incomplete, positive value u_0 is added.
- Step 3. By observing the glottal area function and glottal width function, if the dumping of the amplitude of the vocal fold vibrations in each sub-period is significant, a laryngeal source is dumped using a dumping function. The parameters of the dumping function are estimated by analysis of HSDI.
- Step 4. If different modes are superimposed in the vocal fold vibration, a laryngeal source is modulated by sinusoidal functions with suitable modulation frequencies determined by the frequencies of the vibratory modes observed in the laryngotopograph.
- Step 5. Parameters which minimize differences between a laryngeal source and the glottal area function are estimated.

The parameters of the model are estimated through this procedure, partly manually.

In general, as well-known, the shape of the glottal area function is not equal to that of the laryngeal flow [13], [19]. Non-linear interaction between the vocal fold vibration and the vocal tract plays an essential role to regulate the laryngeal flow. The proposed model directly referred the glottal area functions for determining the laryngeal flow shape and, therefore, realistic physical phenomena are not manipulated into the model. However, in a basic pulse function used here, the Klatt-model, takes differences between the laryngeal flow and the glottal area into considerations. We use the glottal

area functions and other characteristics obtained by analysis of HSDI only for controlling modulation and dumping.

A. Examples of synthesized voices

1) Vocal fold paralysis: The following parameter setting is used for synthesis of a voice with the vocal fold paralysis:

$$[T_1, V_1, O_{q_1}, \theta_1] = [1/270,\ 0.60,\ 0.8,\ 2\pi/45]$$
$$[T_2, V_2, O_{q_2}, \theta_2] = [1/360,\ 0.32,\ 0.9,\ 0]$$
$$u_0 = 0.1$$
$$[T_4, V_4, r, \theta_4] = [1/90,\ 0.50,\ 0.08,\ -5\pi/6]$$

In the case of the vocal fold paralysis in Fig. 1, frequencies of the left and right vocal fold vibration are different. Therefore, two basic pulse functions in 270 Hz and 360 Hz. Dumping of the amplitude of the vocal fold vibrations was observed in Fig. 1, therefore, a dumping function used.

Fig. 7 shows synthesized laryngeal source and the glottal area function of the case with the vocal fold paralysis.

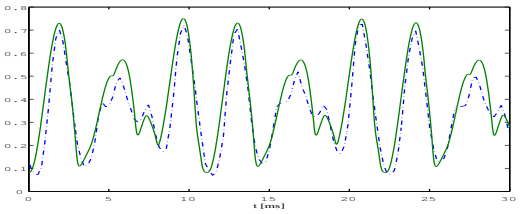

Fig. 7. Synthesized laryngeal source (green solid line) and glottal area function (blue dashed line) for the vocal fold paralysis

2) Vocal fold cyst: The following parameter setting is used for synthesis of a voice with the vocal fold cyst:

$$[T_1, V_1, O_{q_1}, \theta_1] = [1/251,\ 1.00,\ 0.8,\ 0]$$
$$u_0 = 0.05$$
$$[T_5, V_5, \theta_5] = [1/313,\ 0.10,\ -7\pi/18]$$
$$[T_6, V_6, \theta_6] = [1/379,\ 0.30,\ -7/9\pi]$$

In the case of the vocal fold cyst in Fig. 3, the mode in 251 Hz was dominant in the vocal fold vibration, and, hence, only a basic pulse function in 251 Hz is used. Two modulation functions are used to manipulate other vibratory modes in 313 Hz and 379 Hz.

Fig. 8 shows the synthesized laryngeal source and the glottal area function of the case with the vocal fold cyst.

3) Laryngeal papilloma: The following parameter setting is used for synthesis of a voice with the laryngeal papilloma:

$$[T_1, V_1, O_{q_1}, \theta_1] = [1/484,\ 0.20,\ 1.0,\ 0]$$
$$u_0 = 0.8$$
$$[T_5, V_5, \theta_5] = [1/563,\ 0.30,\ -\pi/9]$$
$$[T_6, V_6, \theta_6] = [1/971,\ 0.15,\ 5\pi/9]$$

As well as the case of the vocal fold cyst, first, one basic pulse function in 484 Hz is used and modulation function in 563 Hz is used as $(0, 1)$-mode of 563 Hz.

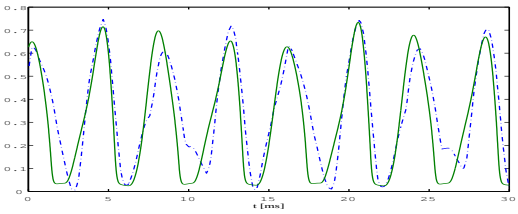

Fig. 8. Synthesized laryngeal source (green solid line) and glottal area function (blue dashed line) for the vocal fold cyst

In the case of the laryngeal papilloma, a higher frequential component was observed at 971 Hz in both the glottal area function and the laryngotopograph. Therefore, a modulation function with 971 Hz is added.

Fig. 9 shows the synthesized laryngeal source and the glottal area function of the case with the laryngeal papilloma.

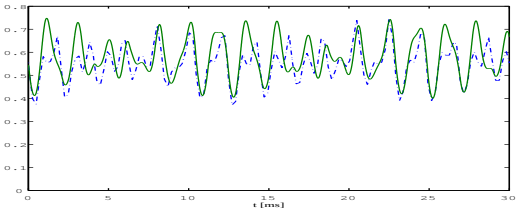

Fig. 9. Synthesized laryngeal source (green solid line) and glottal area function (blue dashed line) for the laryngeal papilloma.

V. Perceptual Evaluation

We evaluate the effectiveness of our laryngeal source model by listening tests.

In the sense of synthesis of subharmonic voices, the proposed model is incomparable with the harmonic models, such as R-model and LF-model, because the harmonic models is not able to synthesize subharmonic sounds. Therefore, the effectiveness of the model in synthesis of subharmonic voices is evaluated by using the R (rough) parameter of the GRBAS scale [7] and diplophonic perception.

Diplophonia is a pathological voice in which two or more pitches are simultaneously perceived. There are a variety of production mechanisms of diplophonic voices, as seen in this paper and [11]. However, perceptual evaluation of diplophonic voices is simply carried out by asking whether two or more pitches are perceived or not, and therefore, it is a binary decision. Hence, perceptual evaluation of diplophonic voices are simpler and easier than the R parameter with 4-point scale of the GRBAS scale.

A. Method

For synthesis of stimuli for listening tests, two basic pulse functions are used. The set of stimuli was synthesized by changing the following parameters: (i) a ratio of the frequencies; (ii) a ratio of the amplitudes; (iii) the fundamental frequencies of two basic pulse functions.

Six participants consisting of medical doctors of the otolaryngology, speech therapists, and speech scientists joined the listening tests. They were instructed to scale the synthesize voices using both the R parameter with 4-point scale from 0 to 3 of the GRBAS scale and the diplophonic binary scale.

As the values of parameters for synthesis in Fig. 6, we set $O_{q1} = O_{q2} = 0.96$, $\theta_1 = \theta_2 = 0$, $u_0 = 0$ to be constant, and change the values of (T_1, V_1, T_2, V_2). The synthesized laryngeal source is convoluted with transfer function of the vowel [a], and the glottal noises are not added in the framework of Klatt synthesizer [9].

B. Results

Fig. 10, 11, 12 show results of the perceptual evaluations with the R parameter and diplophonic perception for $T_1 : T_2 = 2 : 1$, $T_1 : T_2 = 3 : 2$, and $T_1 : T_2 = 4 : 3$, respectively. The x-axes represent the logarithmic ratio of the amplitudes of two basic pulse functions, i. e., $20 \log_{10}(V_1/V_2)$.

Both of the values of the R parameter and diplophonic perceptual rate increase with decreasing differences between the amplitudes of two basic pulse functions in all cases. As F_0 decreases, the value of the R parameter has a tendency to increase. However, in the case that the frequential ratio is $2 : 1$, the diplophonic perceptual rate is not large at the lowest F_0 with 90 Hz :45 Hz.

If $T_1 : T_2 = 3 : 2$ or $4 : 3$, the diplophonic perceptual rate is almost 1.0, if the difference of the amplitudes of two basic pulse functions is small.

The maximal values of the R parameter when $T_1 : T_2 = 2 : 1$ or $3 : 2$ are less than 2.0. Tthe maximum value of the R parameters is largest when $T_1 : T_2 = 4 : 3$, $T_1 = 120$ Hz, and $T_2 = 90$ Hz.

The Pearson product-moment correlation coefficient r between the diplophonic perception and the R parameter is small when $T_1 : T_2 = 2 : 1$ ($r = 0.32$), and large when $T_1 : T_2 = 3 : 2$ ($r = 0.92$) and $T_1 : T_2 = 4 : 3$ ($r = 0.91$).

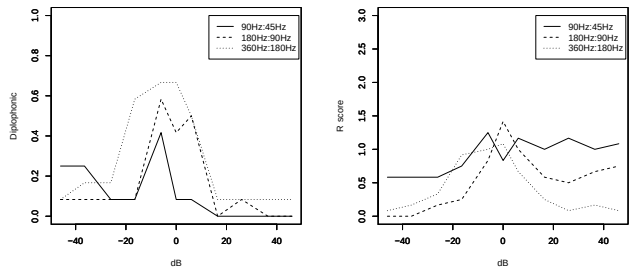

Fig. 10. Results of perceptual evaluation: $T_1 : T_2 = 2 : 1$. Left: a rate of the diplophonic perception. Right: the R parameter with 4-point scale of the GRABS scale. x-axis represents the logarithmic ratio of $V1 : V2$.

VI. Discussions

The synthesized laryngeal sources using the proposed model are almost similar to the glottal area functions.

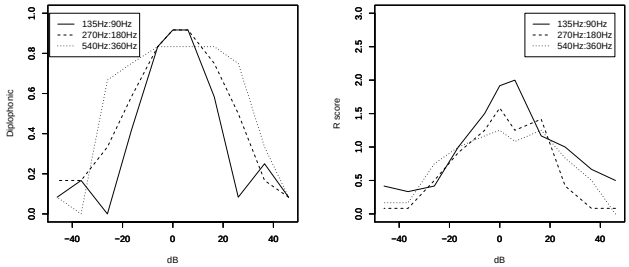

Fig. 11. Results of perceptual evaluation: $T_1 : T_2 = 3 : 2$. Left: a rate of the diplophonic perception. Right: the R parameter with 4-point scale of the GRABS scale. x-axis represents the logarithmic ratio of $V1 : V2$.

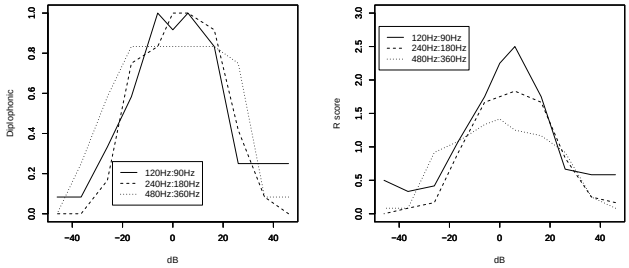

Fig. 12. Results of perceptual evaluation: $T_1 : T_2 = 4 : 3$. Left: a rate of the diplophonic perception. Right: the R parameter with 4-point scale of the GRABS scale. x-axis represents the logarithmic ratio of $V1 : V2$.

The proposed model is very flexible to synthesize subharmonic voices. On the other hand, the model has many parameters and, hence, the parameter estimation is seemingly difficult. However, using results of analysis of the HSDI, the parameters are systematically estimated and the fine structures of modulation and dumping of the glottal area functions obtained from the HSDI are mostly reproduced three different pathological voices by the proposed model.

A many-parameter model proposed in [16] has many parameter, and the automatic estimation of its parameters are available only for subharmonic voices with period-2. However, the proposed method of parameter estimation is almost automatically done even in the case of subharmonic voices with different vibratory modes. Therefore, the proposed model and the method of parameter estimation are considered to be useful for synthesis of subharmonic voices and have succeeded to synthesize voices with diplohphonic perception.

It is reported in [8] that the R parameter of the GRBAS scale increases with increasing perturbation parameters, such as PPQ and APQ, and also increases with decreasing F_0. In [2], subharmonic voice with the $F_0/2$ component was synthesized using simple source model, and it is concluded that the scale of roughness increases with increasing modulation indices of AM and FM, and increases with decreasing F_0. The results about the R parameters in this paper are in good agreement with the results in [2].

VII. Conclusion

The vocal fold vibratory patterns of pathological subharmonic voices were observed using high-speed digital imaging (HSDI) . Production mechanisms of subharmonic voices were analyzed. A new laryngeal source model for pathological and non-pathological subharmonic voices was proposed and the parameter estimation was processed based on physiological observations.

We also evaluated the effectiveness of the model for synthesis of subharmonic voices by listening tests. The proposed model has succeeded to synthesize subharmonic voices perceived as rough or diplophonic.

It is addressed as a future study to propose a full automatic parameter estimation method in the framework of A-b-S in the acoustic domain and method to add turbulent noises based on physiological and physical mechanisms.

Acknowledgment

We thank Takaharu Nito for his helpful supports.

References

[1] L. Bailly, N. Henrich, and X. Pelorson, Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling, *J. Acoust. Soc. Am.* , 127(5):3212–3222, 2010.

[2] C. C. Bergan and I. R. Titze, Perception of pitch and roughness in vocal signals with subharmonics, *J. Voice*, 15(2):165–175, 2001.

[3] G. Fant and J. Liljencrants and Q.-g. Lin, A four-parameter model of glottal flow, *KTH STL-QPSR*, 4/1985:1–14, 1985.

[4] J. L. Flanagan, *Speech analysis synthesis and perception*, 2nd Ed., Springer-Verlag, 1972.

[5] L. Fuks, B. Hammarberg, and J. Sundberg, A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences, *KTH TMH-QPSR*, 3/1998:49–59, 1998.

[6] S. Granqvist and P.-Å. Lindestad, A method of applying Fourier analysis to high-speed laryngoscopy, *J. Acoust. Soc. Am.*, 110(6):3193–3197, 2001.

[7] M. Hirano, *Clinical examination of the voice*, Springer, 1982.

[8] S. Imaizumi, Acoustiic measures of roughness in pathological voice, *J. Phonetics*, 14:457–462, 1986.

[9] D. H. Klatt, Software for a cascade/paraleel formant synthesizer, *J. Acoust. Soc. Am.*, 67(3):971–995, 1980.

[10] D. H. Klatt and L. C. Klatt, Analysis, synthesis, and perception of voice quality variations among female and male talkers, *J. Acoust. Soc. Am.*, 87(2):820–857, 1990.

[11] S. Niimi and M. Miyaji, Vocal fold vibration and voice quality, *Folia Phoniat.*, 52:32–38, 2000.

[12] A. E. Rosenberg, Glottal shape on the quality of natural vowels, *J. Acoust. Soc. Am.*, 49(2):583–590, 1970.

[13] M. Rothenberg, Acoustic interaction between the glottal source and the vocal tract, *Vocal fold physiology*, K. N. Stevens and M. Hirano Ed., pp. 305–328, Univ. Tokyo Press. 1981.

[14] K.-I. Sakakibara, Production mechanism of voice quality in singing, *J. Phonetic Soc. Jpn.*, 7(3):27–39, 2003.

[15] K.-I. Sakakibara, L. Fuks, H. Imagawa, and N. Tayama, Growl voice in pop and ethnic styles, *Proc. International Symposium on Musical Acoustics, 2004*, 2004.

[16] K.-I. Sakakibara and H. Imagawa, A many-parameter model of laryngeal flow with ventricular resonance and supraglottal vibration, *Proc. Forum Acusticum 2005*, 2005.

[17] K.-I. Sakakibara and H. Imagawa, Modal analysis of vocal fold vibrations using laryngotopography, *Proc. Interspeech 2010*, 2010.

[18] K.-I. Sakakibara, K. Kondo, E. Z. Murano, M. Kumada, H. Imagawa, and S. Niimi, Vocal and false vocal fold vibrations and synthesis of khoomei, *Proc. International Computer Music Conference*, 135–138, 2001.

[19] I. R. Titze, *Principles of voice production*, Prentice-Hall, 1994.

[20] I. R. Titze and D. T. Talkin, A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation, *J. Acoust. Soc. Am.*, 66(1):60–74, 1979.