Estimating physical properties of vocal fold paralysis from high-speed filimg data

I. T. Tokuda[1], M. Kimura[2], H. Imagawa[3], K.-I. Sakakibara[4], N. Tayama[5]
[1] *Japan Advanced Institute of Science and Technology, Japan, Email: isao@jaist.ac.jp*
[2] *University of Texas Southwestern, USA;* [3] *University of Tokyo Hospital, Japan*
[4] *Health Sciences University of Hokkaido, Japan;* [5] *International Medical Center, Japan*

Introduction

Vocal fold atrophy causes a spindle-shaped glottal incompetence during phonation that results in a breathy, husky, or weak voice. Despite recent advanced medical technology, there still exist many unknown problems on vocal fold atrophy, since abnormal vocal folds vibrations induced by the vocal fold atrophy are highly complex and nonlinear. Up to date, two standard surgical techniques have been established: 1) aryteroid rotation and 2) injection medialization of vocal cords by biologic implants. Since it is nontrivial to predict the effect of the surgery beforehand, medical doctors must usually rely on their own experience about how to design the operation. Sometimes the surgical improvement can be much smaller than expected. If the simulation of the vocal fold atrophy before and after operation becomes possible, the effect of the operation can be predicted beforehand so as to improve the effect of the surgery. Moreover, such simulations can be used for the purpose of informed consent.

Analysis of pathological voices is composed of three important factors: sound phenomenon, vocal fold vibration, and vocal fold pathology. The vocal fold pathology causes abnormal vibrations of the vocal fold. The abnormal vibrations generate a pathological voice. The pathological voice is perceived as a sound phenomenon [1]. Although the relationship between the sound phenomenon and the vocal fold vibrations has been relatively well understood, the relationship between the abnormal vibrations and the pathological condition of the vocal folds has not yet been completely clarified. Kimura analyzed the digital high-speed camera data before and after the voice surgery to study the change in the vocal fold vibrations [1]. The digital high-speed camera data is useful for the analysis of the vocal fold vibrations, since it contains a rich information on the vibration patterns. Döllinger *et al.* developed a method for extracting vibration parameters from digital high-speed camera data by using an asymmetric two-mass model [5].

In this paper, parameters of the mathematical vocal fold model are estimated for digital high-speed data of vocal fold atrophy. By using Steinecke-Herzel asymmetric two-mass model [4], right and left vocal fold tensions, subglottal pressure and glottal area are estimated. Whether the effect of the voice surgery is reflected in the estimated parameter is judged by using a pair of data sets before and after the operation. Based on the estimation results, we consider whether the simulator of the voice surgery can be constructed by using the mathematical vocal fold model.

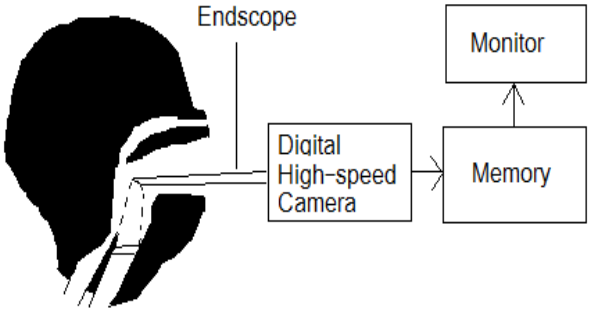

Figure 1: Block chart of the recording of vocal fold vibrations using digital high-speed camera

Digital High-Speed Camera

The digital high-speed camera directly records vocal fold vibrations through an endoscope. The sampling frequency was set to 4500 Hz. Each gray-scale image of the movie data is composed of 256 × 256 pixels. To extract the vocal fold edges, brightness of each pixel was scanned on a horizontal line that crosses the glottis in the recording image. A region where the pixel image is darker than a threshold was defined as the glottis. Edge points of the right and left vocal folds were extracted at the border of the glottis. Fig. 1 shows a block chart of the recording procedure of the vocal fold vibrations. In this paper, we used recording data of two patients A (female, 57 years old) and B (female, 70 years old) before and after the operation. Both patients suffer from vocal cord paralysis. As the voice surgery, injection medialization of vocal cords by biologic implants was carried out.

Mathematical Model

Two-mass model of the vocal fold was proposed by Ishizaka and Flanagan in 1972 [2]. In order for the simulation of pathological voices, Ishizaka and Isskhiki introduced an asymmetry between the right and left vocal folds into the Ishizaka-Flanagan model [3]. Steinecke and Herzel further simplified the Ishizaka-Isshiki model [4]. In this paper, we used the Steinecke-Herzel (SH) model, since it provides one of the simplest models that can describe the vocal fold atrophy. Fig. 2 shows a schematic representation of the SH model. Dynamics of

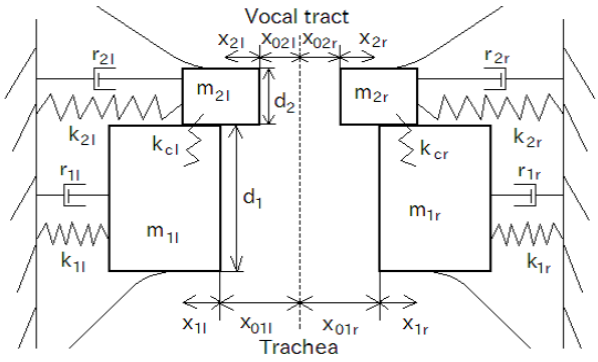

Figure 2: Schematic representation of SH model

the SH model is governed by the following equations:

$$m_{i\alpha}\ddot{x}_{i\alpha} + r_{i\alpha}\dot{x}_{i\alpha} + k_{i\alpha}x_{i\alpha} + \Theta(-a_i)c_{i\alpha}\left(\frac{a_i}{2l}\right) + k_{c\alpha}(x_{i\alpha} - x_{j\alpha}) = F_i(x_{1l}, x_{1r}, x_{2l}, x_{2r}), \quad (1)$$

$$\Theta(x) = \begin{cases} 1 & (x > 0), \\ 0 & (x \leq 1), \end{cases}$$

$$i, j = \begin{cases} 1: & \text{lower mass}, \\ 2: & \text{upper mass}, \end{cases} \quad \alpha = \begin{cases} l: & \text{left side}, \\ r: & \text{right side}, \end{cases}$$

where $m_{i\alpha}$ represents a mass size, $r_{i\alpha}$ is a damping constant, $k_{i\alpha}$ is a spring constant, $k_{c\alpha}$ is a coupling constant between m_1 and m_2, $c_{i\alpha}$ is an additional spring constant during vocal fold contact, $x_{i\alpha}$ is a distance of each mass from midline, and F_i stands for a pressure force acting on each mass in the glottis. A vocal fold tension Q is defined by the following equation.

$$k_{i\alpha} = Q_\alpha k_{i\alpha 0}, \quad k_{c\alpha} = Q_\alpha k_{c\alpha 0}, \quad c_{i\alpha} = Q_\alpha c_{i\alpha 0}, \quad m_{i\alpha} = m_{i\alpha 0}/Q_\alpha. \quad (2)$$

A full description of the SH model and a detailed setting of the parameter values can be found in Ref. [4].

Parameters Estimation

Conversion of physical unit of glottal waves

In the high speed recording of the glottal waves, the unit of the amplitude is in pixel. On the other hand, the unit of the amplitude of the glottal waves generated by the SH model is in cm. The physical length, to which 1 pixel corresponds, changes from one data to another, because it is difficult to set the distance between the endoscope and the vocal folds always the same for all patients (see Fig.1). Therefore, it is necessary to transform the value of pixel into physical value. Supposing that the glottal length of each patient is equal to the standard value, amplitude of the glottal waves in the high-speed image was converted from pixel to cm. The standard glottis length was assumed to be 1.4 cm for male subjects and 0.7 cm for female subjects.

Cost function

The parameters of the SH model are estimated in a way that the glottal waves generated by the model becomes similar to the ones of the digital high-speed data. As the vibration parameters, Q_α, P_S, and a_{rest} are estimated. Q_α stands for right or left vocal fold tension, P_S represents a subglottal pressure, and a_{rest} is a prephonatory glottal area. The cost function Γ for the optimization is defined by the following equations:

$$\begin{aligned} \Gamma(Q_l, Q_r, P_S, a_{rest}) &= w_f\left(\frac{|f_{lc} - f_{lt}|}{|f_{lc}|} + \frac{|f_{rc} - f_{rt}|}{|f_{rt}|}\right) \\ &+ w_S\left(\frac{|S_{lc} - S_{lt}|}{|S_{lc}|} + \frac{|S_{rc} - S_{rt}|}{|S_{rt}|}\right) \\ &+ w_c\frac{|C_c - C_t|}{|C_c|} \quad (3) \\ n &= \begin{cases} c: & \text{High-speed data}, \\ t: & \text{SH model}, \end{cases} \\ w_f &= w_S = 1/7,\ w_c = 5/7, \end{aligned}$$

where $f_{\alpha n}$ is the fundamental frequency of the glottal waves, $S_{\alpha n}$ is the power of the fundamental frequency, and C_n is the closed phase of the glottal waves.

Initial values

Since the cost function Γ is non-convex, the optimization procedure requires suitable initial values, which are obtained as follows. By the first order approximation, the tension parameter Q_α is roughly estimated as

$$\tilde{Q}_\alpha = f_{\alpha c}\, 2\pi\sqrt{m_{1\alpha 0}/k_{1\alpha 0}}, \quad (4)$$

using the fundamental frequency $f_{\alpha c}$ of the high-speed data. An interval $Q_\alpha \in [\tilde{Q}_\alpha - 0.2, \tilde{Q}_\alpha + 0.2]$, which is extended around the estimated tension value $\tilde{Q}_\alpha$, is divided into small pieces with a width of 0.02. A search space for the other parameters is also set as $P_S \in \{6, 8, 10, \ldots, 46\}$ and $a_{rest} \in \{0.02, 0.04, 0.06, \ldots, 0.2\}$. For all combinations of the parameter values (Q_l, Q_r, P_S, a_{rest}), 20 sets of parameter values that give the minimal cost function Γ are collected.

By using these 20 sets as the initial values, the parameters are optimized by the Nelder-Mead algorithm. The parameter set that gives the minimum cost Γ among the 20 sets of the optimized parameter values is considered as our final estimation result.

Analysis of glottal waves

The glottal waves generated by the SH model and the digital high-speed camera data were compared by using the following indexes:

- Frequency difference between right and left vocal folds,
- Phase difference between right and left vocal folds,
- Amplitude difference between right and left vocal folds,
- Closed phase,
- Normalized amplitude quotient (NAQ) [6] ($NAQ \in [0.1, 0.2]$ for normal subjects).

Results

The parameters were estimated for the digital high-speed camera data by using the method explained in section 4.

Table 1: Parameters estimated for patient A before and after surgery.

	Before Operation	After Operation
Q_l [g/ms^2]	2.089	1.857
Q_r [g/ms^2]	2.061	1.993
P_S [g/cm·ms^2]	41.7	35.0
a_{rest} [cm^2]	0.0606	0.0198

Table 2: Parameters estimated for patient B before and after surgery.

	Before Operation	After Operation
Q_l [g/ms^2]	2.031	2.254
Q_r [g/ms^2]	2.000	1.899
P_S [g/cm·ms^2]	34.6	29.5
a_{rest} [cm^2]	0.0997	0.0479

Fig. 3 shows the glottal waveform of patient A after the operation, where left (red solid line) and right (blue dotted line) glottal waveforms are simultaneously drawn. The model simulation (lower graph) is compared with the high-speed data (upper graph). Tables 1 and 2 show the estimated parameter values of patients A and B. Table 3 shows the indexes of the previous subsection applied to the estimation results of patient A. Tables 1 and 2 shows that the glottal area a_{rest} becomes smaller after the operation than before. It seems that the estimated glottal area shows the effect of the operation, since the injection medialization technique intended to narrow the glottal area. The subglottal pressure P_S, which is lowered after the operation, implies that the phonation effort has been reduced. Table 3 shows that the frequency and the closed phase of the high-speed data are well reproduced in the model simulation. The NAQ becomes closer to the normal range after the operation. This might indicate that the voice quality has been improved by the operation. On the other hand, for the phase and amplitude differences, there exist some gaps between the high-speed camera data and the model simulation.

Table 3: Analysis of glottal waves for patient A before and after surgery.

Before Operation	SH model	Camera Data
Frequency (left) [Hz]	290	290
Frequency (right) [Hz]	290	290
Closed phase [%]	14.8	14.5
Phase difference [%]	1.24	6.20
Amplitude difference [%]	0.70	14.3
NAQ	0.285	0.260
After Operation	SH model	Camera Data
Frequency (left) [Hz]	288	288
Frequency (right) [Hz]	287	287
Closed phase [%]	25.5	32.3
Phase difference [%]	6.46	11.8
Amplitude difference [%]	10.1	8.72
NAQ	0.249	0.212

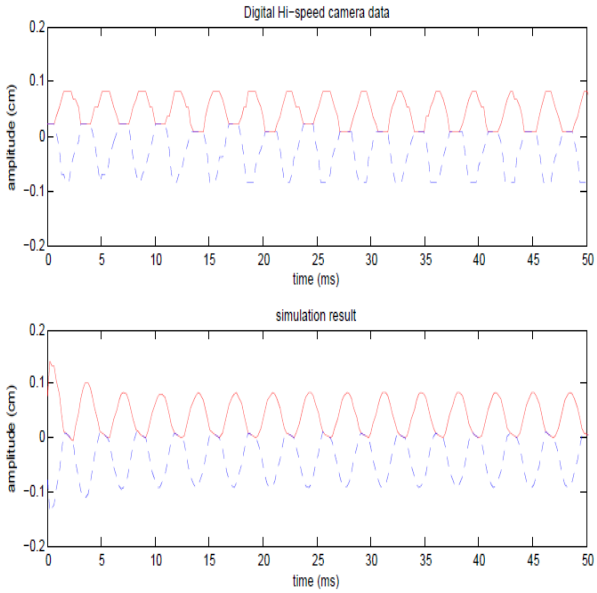

Figure 3: Simultaneous drawing of left (red solid line) and right (blue dotted line) glottal waveforms. The model simulation (lower graph) is compared with digital high-speed recording data (upper graph).

Conclusions

In this paper, vibration parameters of the SH model have been estimated for digital high-speed recording of vocal fold atrophy. The results imply that the effect of the voice surgery is appropriately reflected in the estimated parameter values. Future studies include automatic determination of the weighting coefficient of the cost function Γ as well as application of the present method to the data of other patients.

Acknowledgements

This study was supported by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

References

[1] M. Kimura, *High-speed digital recording of abnormal vocal fold oscillations and its surgical treatment,* PhD Thesis, University of Tokyo, 2007.

[2] K. Ishizaka and J. L. Flanagan, Bell. Syst. Tech. J., Vol. 51, pp. 1233-1268, 1972.

[3] K. Ishizaka and K. Isshiki, J. Acoust. Soc. Am., Vol. 60, pp. 1193-1198, 1976.

[4] I. Steinecke and H. Herzel, J. Acoust. Soc. Am., vol. 97, pp. 1249-1259, 1995.

[5] M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, and U. Eysholdt, of the vocal folds," IEEE Trans. on Biomed. Engin., Vol. 49, pp. 773-781, 2002.

[6] P. Alku, T. Backstrom, and E. Vilkmam, J. Acoust. Soc. Am., Vol.112, pp701-710, 2002.