

## Research Article

# Pathological Voice Source Analysis System Using a Flow Waveform-Matched Biomechanical Model

Xiaojun Zhang <sup>1,2</sup>, Lingling Gu,<sup>2</sup> Wei Wei <sup>2</sup>, Di Wu,<sup>2</sup> Zhi Tao <sup>1,2</sup> and Heming Zhao<sup>1</sup>

<sup>1</sup>School of Electronic and Information Engineering, Soochow University, Suzhou 215000, China

<sup>2</sup>College of Physics, Optoelectronics and Energy, Soochow University, Suzhou 215000, China

Correspondence should be addressed to Wei Wei; [weiwei0728@suda.edu.cn](mailto:weiwei0728@suda.edu.cn) and Zhi Tao; [taoz@suda.edu.cn](mailto:taoz@suda.edu.cn)

Received 30 March 2018; Accepted 24 May 2018; Published 2 July 2018

Academic Editor: Liwei Shi

Copyright © 2018 Xiaojun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Voice production occurs through vocal cord and vibration coupled to glottal airflow. Vocal cord lesions affect the vocal system and lead to voice disorders. In this paper, a pathological voice source analysis system is designed. This study integrates nonlinear dynamics with an optimized asymmetric two-mass model to explore nonlinear characteristics of vocal cord vibration, and changes in acoustic parameters, such as fundamental frequency, caused by distinct subglottal pressure and varying degrees of vocal cord paralysis are analyzed. Various samples of sustained vowel /a/ of normal and pathological voices were extracted from MEEI (Massachusetts Eye and Ear Infirmary) database. A fitting procedure combining genetic particle swarm optimization and a quasi-Newton method was developed to optimize the biomechanical model parameters and match the targeted voice source. Experimental results validate the applicability of the proposed model to reproduce vocal cord vibration with high accuracy, and show that paralyzed vocal cord increases the model coupling stiffness.

## 1. Introduction

Vocal cord vibration interrupts the straight airflow expelled by the lungs into a series of pulses that act as the excitation source for voice and sound. Denervation or organic diseases of vocal cords, such as paralysis and polyps, can cause irregular vibration with consequential changes, manifested as breathy or hoarse voice. These diseases generally affect one side of vocal structure, causing significant imbalance in bilateral vocal cord tension [1, 2]. Irregular vibration of the vocal cords corresponding to a variety of voice disorders can be observed with electronic laryngoscope to assist diagnosing vocal cord disease. However, laryngoscopy examination is invasive, and the outcomes are relatively subjective. Acoustic analysis can complement and in some cases replace the other invasive methods, which based on direct vocal fold observation [3, 4].

Clinical diagnosis and pathological voice classification using objective methods is an important issue in medical evaluation. Previous studies have mainly combined acoustic

parameters with pattern recognition algorithms to assist diagnosis of pathological voice [5, 6]. However, the selected voice signal parameters are not directly linked with the actual physical structure, and vocal structural changes that cause vocal voice disorders require further study.

Nonlinear dynamics theory has provided a new avenue for dynamical system related research, for example, methods combining nonlinear theory with spectral analysis have been successfully applied to EEG and ECG signal analysis. It has also been extended to study voice signals [7, 8].

Nonlinearity inherent in the vocal system can cause irregular voice behavior, as indicated by harmonics, bifurcation, and low-dimensional chaos in high-speed recording of vocal cord vibration signals [9, 10]. The degree of pathological vocal fold is closely related to the nonlinear vibration of the vocal cords. [11]. Therefore, traditional analysis of acoustic parameters may not be accurate, but nonlinear dynamics theory has been shown to have good applicability in characterizing such signals [12]. Time frequency shape analysis based on embedding phase space plots and nonlinear dynamics

methods can be used to evaluate the vocal fold dynamics during phonation [13]. Nonlinear models can also simulate various vocal sound phenomena and have been used for dynamic prediction of disordered speech associated with larynx pathology [14–16]. Many physical modeling methods for glottal excitation have been proposed, and the corresponding model parameters have been utilized to study various voice disorders. The two-mass (IF) model is the most well-known classical physical model of the vocal cords, first proposed by Ishizaka and Flanagan and simplified by Steinecke and Herzel (SH model), to study vibration characteristics of the vocal cords. Xue combined the work of Steinecke and Herzel with Navier-Stokes equations and analyzed irregular vibrations caused by tension imbalance in bilateral vocal cord, as well as sound effects [17]. Recently, Sommer modified the asymmetric vocal contact force of the SH model based on Newton's third law [18]. However, a comprehensive nonlinear analysis for the modified SH model remains incomplete.

Although physical modeling has enormous potential in speech synthesis and voice analysis, the large number of model parameters and the complexity of model optimization to match observational data have prevented its practical application [19]. Döllinger used the Nelder–Mead algorithm to minimize the error between experimental curves obtained from high-speed glottography sequences and curves generated with the two-mass model (2MM) [20]. However, this is an invasive method because an endoscope is required to record vocal cord vibrations during phonation. Gómez computed biomechanical parameters based on the power spectral density of the glottal source to improve detection of voice pathology [21].

Other researchers have used genetic algorithms to optimize model parameters to match recorded glottal area, trajectory, and glottal volume wave and have shown the possibility of model inversion [22, 23]. Tao extracted the physiologically relevant parameters of the vocal fold model from high-speed video image series [24].

The complex optimization process and large number of parameters mean the matching result can be unstable. Thus, finding the important tuning parameters and selecting appropriate optimization algorithms are still important issues to be resolved for physical modeling applications, and simulations for asymmetric vocal cords also require further study.

This paper designed a pathological voice source analysis system using an optimized model to study the dynamics of asymmetric vocal cords. Incorporating spectral analysis, and bifurcation and phase diagrams, this paper investigates the impact of structural change of the vocal cord on its vibration and fundamental frequency. Sound effects due to lung pressure are also studied. An optimized SH model combined with particle swarm and quasi-Newton methods (GPSO-QN) is proposed to determine biomechanical model parameters. Parameter adjustments and changing the oscillation mode of the model allow normal and paralyzed voice sources to be simulated. Differences between optimized model parameters are analyzed to assist in identifying the source of vocal paralysis.

## 2. Method

**2.1. Symmetric Vocal Model.** Vocal cords are two symmetrical membranous anatomical structures located in the throat. Airflow out of the trachea and lungs continuously impacts the vocal cords and causes vibration. The vibration behavior modulates the airflow to generate glottal pulses [25]. Based on the elastic and dynamic properties of the vocal cords, each fold is represented by two coupled oscillators with two masses, three springs, and two dampers, where the quality of the mass and spring constants denote vocal quality and tension, respectively. Figure 1 shows the simplified two-mass (SH) model, which can be expressed as

$$\begin{aligned} \dot{x}_{1\alpha} &= v_{1\alpha}, \\ \dot{v}_{1\alpha} &= -\frac{1}{m_{1\alpha}} (F_{1\alpha} + I_{1\alpha} - r_{1\alpha}v_{1\alpha} - k_{1\alpha}x_{1\alpha} - k_{c\alpha}(x_{1\alpha} - x_{2\alpha})), \\ \dot{x}_{2\alpha} &= v_{2\alpha}, \\ \dot{v}_{2\alpha} &= -\frac{1}{m_{2\alpha}} (I_{2\alpha} - r_{2\alpha}v_{2\alpha} - k_{2\alpha}x_{2\alpha} - k_{c\alpha}(x_{2\alpha} - x_{1\alpha})), \end{aligned} \quad (1)$$

where

$$\begin{aligned} F_{1\alpha} &= \frac{LdP_1}{m_{1\alpha}}, \\ I_{i\alpha} &= -\Theta(-a_i) \frac{c_{i\alpha}}{m_{i\alpha}} \frac{a_i}{2L}, \\ \Theta(x) &= \begin{cases} 1, & x > 0 \\ 0, & x < 0, \end{cases} \\ a_i &= a_{0i} + L(x_{il} + x_{ir}), \\ a_{\min} &= \min(a_1, a_2), \end{aligned} \quad (2)$$

index  $i = 1, 2$  denotes the upper and lower mass, respectively;  $\alpha = l, r$  denotes the left and right parts, respectively;  $P_s$  is the subglottal pressure;  $x_{i\alpha}$  and  $v_{i\alpha}$  are the displacement and corresponding velocity of the masses, respectively;  $m_{i\alpha}$ ,  $k_{i\alpha}$ ,  $k_{c\alpha}$ , and  $r_{i\alpha}$  represent the mass, spring constant, coupling constant, and damping constant, respectively;  $L$ ,  $d$ , and  $a_{0i}$  represent the vocal cord length, thickness of mass  $m_{1\alpha}$ , and rest area, respectively;  $c_{i\alpha} = 3k_{i\alpha}$  is an additional spring constant for handling collision;  $a_i$  is the glottal area;  $F_{1\alpha}$  and  $I_{i\alpha}$  are the Bernoulli force and restoring force due to vocal cord collision, respectively; and  $P_1$  is the pressure on the lower masses.

Using aerodynamic analysis, pressure drops at the glottal entrance and viscous loss within the glottis is ignored.

In contrast to the IF model, Bernoulli flow exists below the narrowest glottis gap only, with a jet region above the contraction where pressure is considered to be constant [26]. From Bernoulli's equation,

$$P_s = P_1 + \frac{\rho}{2} \left( \frac{U_g}{a_1} \right)^2 = P_0 + \frac{\rho}{2} \left( \frac{U_g}{a_{\min}} \right)^2, \quad (3)$$

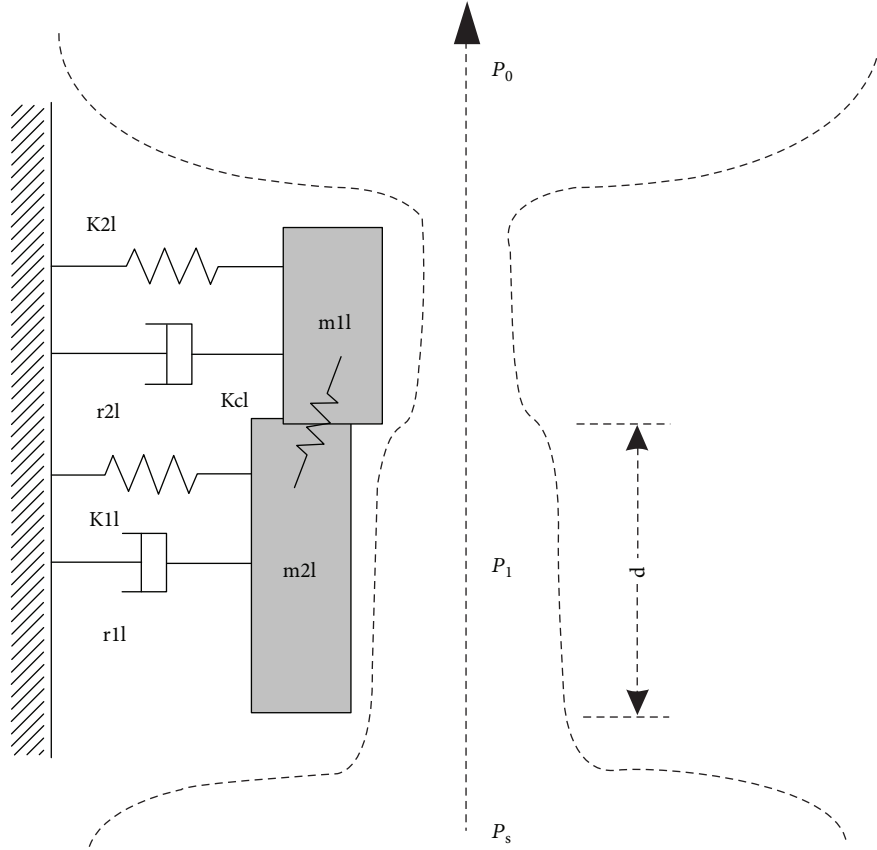


FIGURE 1: Schematic of the Herzel and Steinecke model.

where  $P_0$  is the supraglottal pressure,  $U_g$  is volume flow velocity (glottal waveform), and  $\rho$  is air density. We ignore channel coupling, that is,  $P_0 = 0$ , and consider that Bernoulli pressure exists only when the glottis is open. Therefore,

$$P_1 = P_s \left[ 1 - \Omega(a_{\min}) \left( \frac{a_{\min}}{a_1} \right)^2 \right] \Omega(a_1), \quad (4)$$

$$U_g = \sqrt{2P_s/\rho} a_{\min} \Theta(a_{\min}), \quad (5)$$

where

$$\Omega(x) = \begin{cases} \tanh [50(x/x_0)], & x > 0 \\ 0, & x < 0, \end{cases} \quad (6)$$

with the units centimeters, grams, and milliseconds, respectively.

The standard parameters of this model are  $m_{1\alpha} = 0.125$ ,  $m_{2\alpha} = 0.025$ ,  $k_{1\alpha} = 0.08$ ,  $k_{2\alpha} = 0.008$ ,  $k_{c\alpha} = 0.025$ ,  $r_{1\alpha} = r_{2\alpha} = 0.02$ ,  $P_s = 0.008$ ,  $d = 0.25$ ,  $a_{01} = a_{02} = 0.05$ , and  $L = 1.4$ . These parameters are used by the symmetric model to simulate vocal cord vibration, solving the differential equations using the standard fourth order Runge-Kutta method with initial conditions  $x_{1\alpha}(0) = 0.01$ ,  $x_{2\alpha}(0) = 0.01$ ,  $v_{1\alpha}(0) = 0$ , and  $v_{2\alpha}(0) = 0$ , as shown in Figure 2. Displacement of

upper and lower masses and glottal airflow waveforms are cyclical, and a fixed phase difference exists for the displacement waveform (see Figure 2(a)).

**2.2. Asymmetric Vocal Cord Model.** Vocal polyps and paralysis often occur in one side of the vocal cords. Asymmetric vocal cords cause tension imbalance, and overcritical imbalance may cause irregular vibration. Without loss of generality, we assume the left vocal cord is normal, that is, unchanged parameters, and lesions occur only on the right vocal cord. This imbalance is represented by an asymmetry parameter  $\beta$  ( $0.4 < \beta \leq 1$ ), and right vocal parameters can be expressed as

$$\begin{aligned} \tilde{m}_{ir} &= \frac{m_{ir}}{\beta}, \\ \tilde{k}_{ir} &= \beta k_{ir}, \\ \tilde{k}_{cr} &= \beta k_{cr}, \\ \tilde{c}_{ir} &= \beta c_{ir}. \end{aligned} \quad (7)$$

Small  $\beta$  means a high degree of asymmetry and leads to more complex vocal cord vibration. Consequently, subharmonic performance is enhanced and chaos occurs. Bifurcation diagrams and phase portraits can be used to describe the impact of  $\beta$  changes on the vocal system.

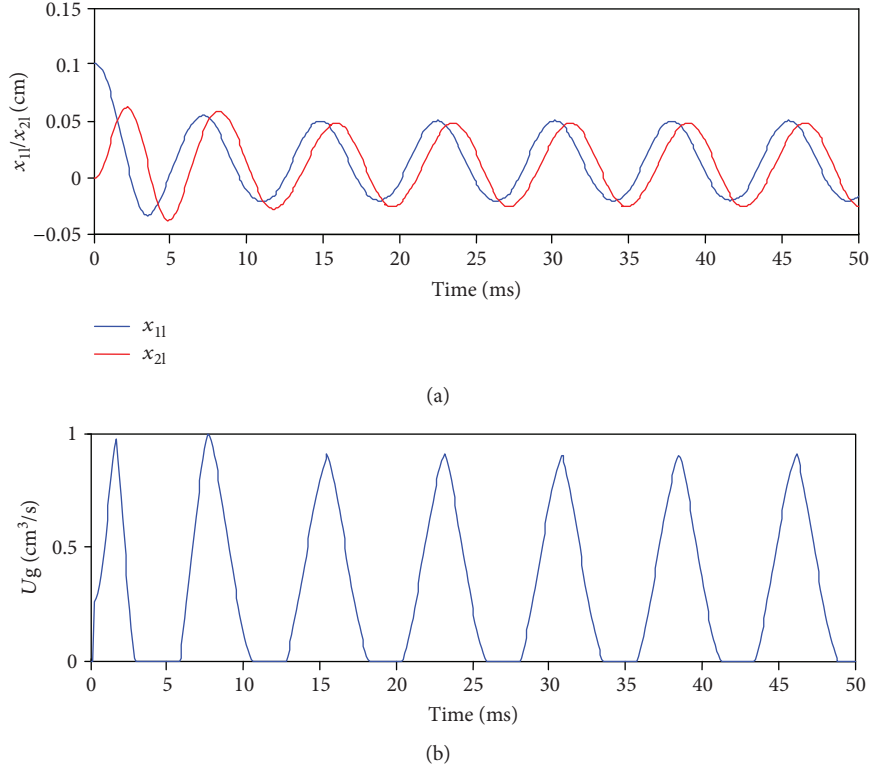


FIGURE 2: Simulation of the standard symmetric model showing oscillation of (a) left lower and upper masses ( $x_{1l}$  and  $x_{2l}$ , resp.), and (b) glottal volume flow velocity  $U_g$ .

When the vocal cords are asymmetric, contact forces are modified as

$$\begin{aligned} I_{ir} &= -\Theta(-a_i) \frac{c_{ir}}{m_{ir}} \frac{a_i}{L} \left( \frac{1}{\beta + 1} \right), \\ I_{il} &= -\Theta(-a_i) \frac{c_{il}}{m_{il}} \frac{a_i}{L} \left( \frac{\beta}{\beta + 1} \right). \end{aligned} \quad (8)$$

**2.3. Analysis of Vocal Vibration.** Vibration characteristics of the asymmetric two-mass model were analyzed with respect to time, frequency, and phase. The vocal mechanism of clinical pathological voice was also investigated with respect to physical simulation. As discussed above, we assumed the left vocal cord was normal, and lesions occurred only in the right vocal cord. Clinical observation of vocal cord physiological characteristics suggested  $0.4 < \beta \leq 1$  was an appropriate range and subglottal pressure was fixed at 0.8 kPa.

Figure 3 shows displacement of the lower bilateral mass for  $\beta = 0.45, 0.53, 0.6, 0.8,$  and  $1$ . Vocal cords on both sides were structurally symmetrical for normal voice, and the vibrational waveforms on both sides coincided completely. Duration of the vocal opening and closing once is defined as one pitch period, and there exists one maximum value of  $x_{ir}$  in such a period.

Asymmetric vocal cord vibrations are significantly more complex. When the degree of asymmetry was relatively small ( $\beta = 0.8$ ), right vocal amplitude was slightly larger than the left side, and the phase was relatively advanced.

As the degree of asymmetry increases, right vocal amplitude also increases with left amplitude remaining essentially unchanged. Consequently, phase difference increases, and the extrema ratio of both sides is no longer 1:1. Figure 3(d) shows the extrema ratio changes to 1:3, and quasiperiodic or irregular oscillations appear, leading to irregular airflow velocity.

Before and after bifurcation, evolution of the dynamical systems in phase space can be described with phase diagrams of the displacement of bilateral vocal cord vibration in the  $x_{1l} - x_{1r}$  plane. Figure 4 shows that when  $\beta = 0.8$ , no bifurcation occurs, and the phase trajectory is a limit cycle. As  $\beta$  reduces to 0.53, asymmetry increases, bifurcation appears, and the phase trajectory becomes a complicated period doubling limit cycle. However, when  $\beta = 0.45$ , the phase trajectory geometry simplifies, which is consistent with the results in the time domain.

Considering the cases with fixed subglottal pressure (0.8 kPa) and  $\beta = 0.45, 0.53, 0.6, 0.8,$  and  $1$ , we compared Fourier spectra corresponding to  $x_{1l}, x_{1r}, U_g$ , and the natural frequencies obtained from an eigenvalue analysis of the system. Figures 5(a)–(e) show two vertical dashed lines that represent the two natural frequencies of the left vocal cord, and dash-dotted lines represent those of the right vocal cord.

When  $\beta = 1$  (Figure 5(a)), the healthy phonation case and the bilateral folds have the same natural frequency. This phonation frequency is approximately 145 Hz, located between the two eigenfrequencies of the left (or right) side.

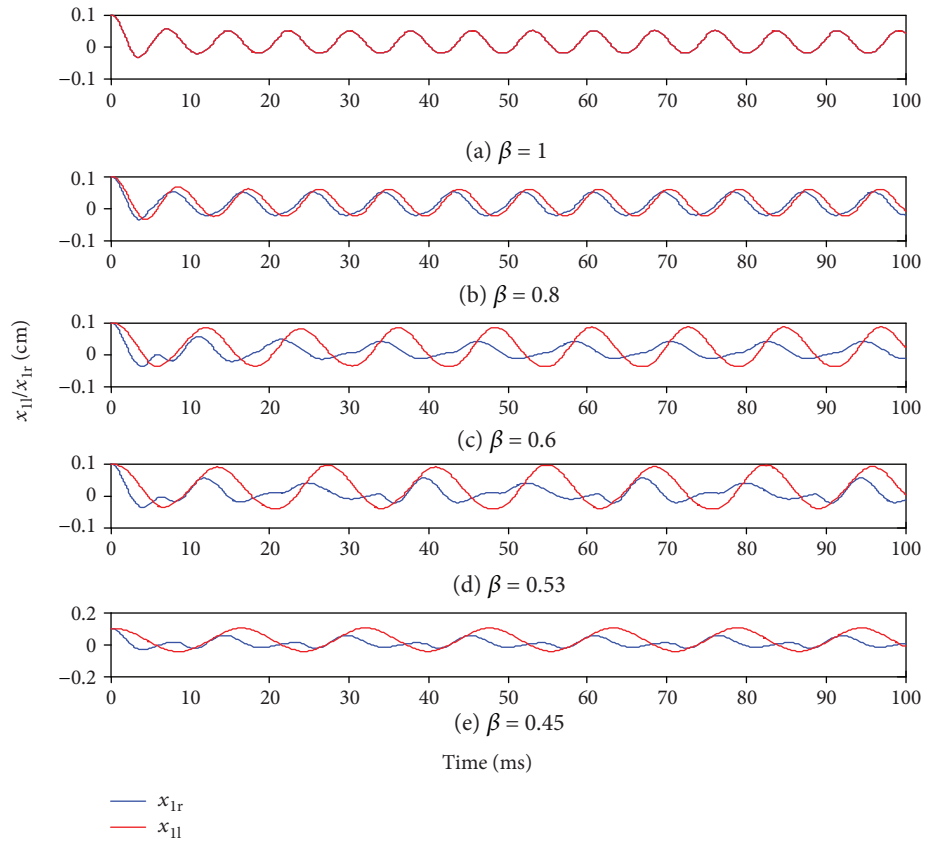


FIGURE 3: Mass displacements of the lower left and right sides.

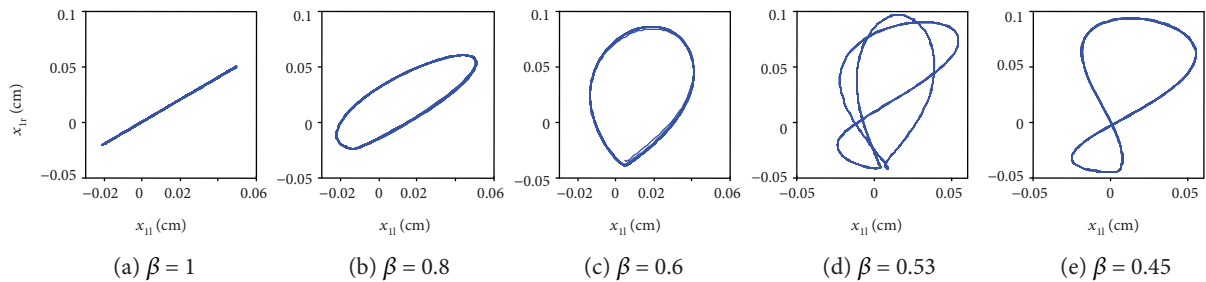


FIGURE 4: Phase space portrait in the  $x_{1l} - x_{1r}$  plane.

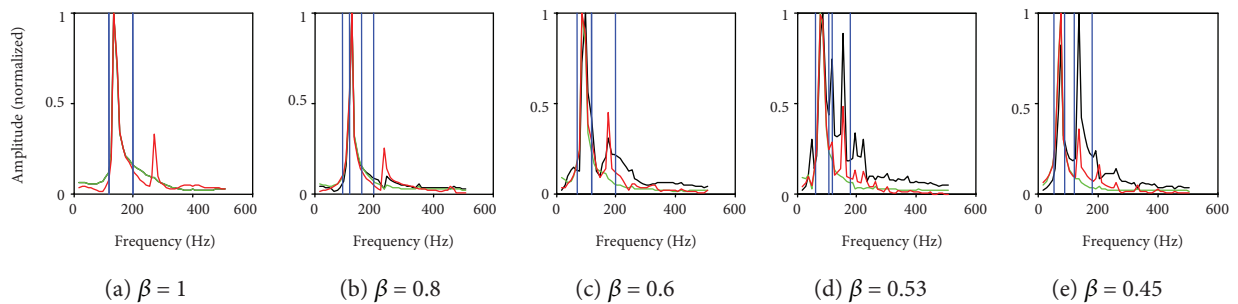


FIGURE 5: Fourier spectra corresponding to displacement of the two lower masses and the normalized glottal volume flow rate with  $P_s = 0.8$  kPa. Volume flow rate and left and right mass displacement are represented by the red, black, and green lines, respectively. Vertical dash lines represent the two left vocal cord natural frequencies, and the dash-dot lines represent those of the right vocal cord.

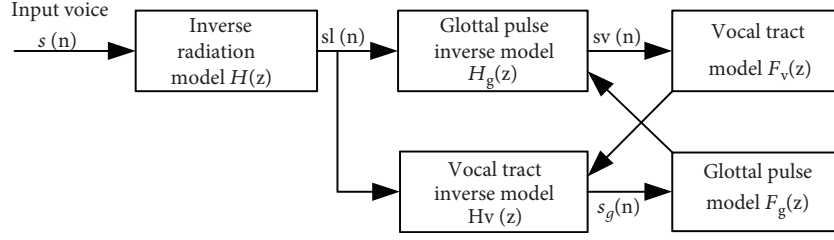


FIGURE 6: Estimation of the glottal pulse  $s_g(n)$  by iterative filtering.

As  $\beta$  reduces, the eigenfrequencies do not coincide again and more complex vibratory behaviors are observed. Figure 5(b) shows that for less asymmetry,  $\beta = 0.8$ , although the intrinsic frequency changes, there is relatively little effect on the frequency spectrum. Figure 5(c) shows that when  $\beta = 0.6$ , a frequency approximately 190 Hz with relatively small amplitude appears between the two eigenfrequencies of the left normal folds. Figure 5(d) shows that when  $\beta = 0.53$ , the overlapped frequencies of the preexisting overtone separate and a small overtone frequency appears between them at 110 Hz. Figure 5(e) shows that when  $\beta = 0.45$ , the overtone between the second eigenfrequency of the right fold and the first left fold disappears. However, the amplitude of the overtone frequency becomes nearly as large as the pitch frequency.

Thus, the fundamental frequency is mainly dependent on the pathological vocal cords, while the normal folds mainly influence the overtone.

### 3. Model Parameter Optimization

We propose an optimization process to find appropriate parameters for the biomechanical model that can accurately simulate normal and paralyzed voice sources. First, inverse filtering is implemented to reduce the channel effect on the speech signal, and glottal flow is extracted. Glottal flow is separately parameterized in time and frequency domains to reduce computational complexity. Then, an optimization algorithm is employed to optimize SH model parameters to obtain a simulated glottal flow. Finally, minimizing error between the parameters of the simulated and extracted glottal flows allows the model to accurately reproduce the particular voice source, and corresponding vocal parameters can also be obtained.

**3.1. Estimation of the Glottal Source.** Reconstruction of the glottal source is based on the adaptive version of iterative inverse filtering developed by Alku [27]. The voice trace,  $s$ , may be considered as the output of a generation model,  $f_g$ , excited by a train pulse,  $\delta$ , whose output is modeled by the vocal tract transfer function,  $f_v$ , to yield voice at the lips,  $s_l$ , which is radiated as  $s$ , where  $r$  is the radiation model, that is,  $*$  means convolution of signals,

$$s = \left( (\delta * f_g) * f_v \right) * r = (f_g * f_v) * r = s_l * r \quad (9)$$

Figure 6 shows the inverse filtering procedure. The radiation effect is first removed by  $H(z)$ , and the resulting radiation compensated voice,  $s_l(n)$ , is filtered by  $H_g(z)$  to reconstruct the deglottalized voice,  $s_v(n)$ , from which the estimate of  $F_v(z)$  may be derived. The vocal tract inverse model fed with the  $F_v(z)$  filter parameters was used to remove the influence of the vocal tract from  $s_l(n)$ , producing a first estimate of the glottal pulse,  $s_g(n)$ . Another iteration was started with the new estimated  $H_g(z)$  loaded by  $F_g(z)$ , and the cycle repeated 2 or 3 times to obtain a good estimation of the glottal source.

The glottal flow will be defined as

$$u_g(n) = s_g(n) - \bar{s}_g(n). \quad (10)$$

An example of the glottal flow estimation from inverse filtering is shown in Figure 7.

**3.2. Objective Function Vocal Cord.** Since the asymmetric SH model influences oscillations in both time and frequency domains, the glottal flow,  $u_g$ , and simulated waveforms,  $U_g$ , were also parameterized within those domains for comparison frequency,  $F_0$ , and time quotients based on the Lijencrants-Frant model were calculated, including speed quotient (SQ), the ratio of the glottal opening to closing time open quotient (OQ), the ratio of the open time to the fundamental period; closing quotient (CIQ), the closing time divided by the fundamental period; and normalized amplitude quotients (NAQ), the ratio of amplitude quotients (maximum amplitude divided by corresponding maximum negative peak of its first derivative) to the fundamental period.

To describe the error between normal target glottal flow and simulated waveforms, the objective function,  $FY$ , was defined as

$$FY = \omega_1 \left( \frac{|OQ - OQ'|}{OQ} + \frac{|SQ - SQ'|}{SQ} + \frac{|CIQ - CIQ'|}{CIQ} + \frac{|NAQ - NAQ'|}{NAQ} \right) + \omega_2 \frac{|F_0 - F_0'|}{F_0}, \quad (11)$$

where “'” means the parameters are derived from the simulation waveform.



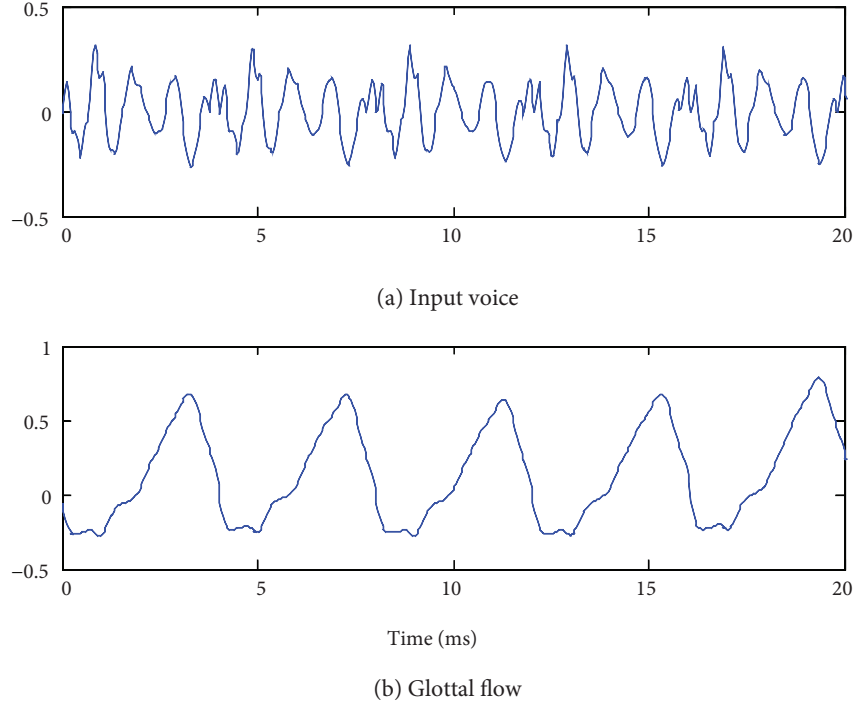


FIGURE 7: Example from vowel /a/ for a normal speaker.

Traditional perturbation analyses have shown instability of pathological vocal sound. The resultant objective function is defined as:

$$\begin{aligned}
 FY_p = \omega_3 FY + \omega_4 \left( \frac{|J_{OQ} - J_{OQ}'|}{J_{OQ}} + \frac{|J_{SQ} - J_{SQ}'|}{J_{SQ}} \right. \\
 \left. + \frac{|J_{CIQ} - J_{CIQ}'|}{J_{CIQ}} + \frac{|J_{NAQ} - J_{NAQ}'|}{J_{NAQ}} \right. \\
 \left. + \frac{|J_{F_0} - J_{F_0}'|}{J_{F_0}} \right), \quad (12)
 \end{aligned}$$

where variables with superscript denote parameters of the simulated glottal flow.

If the time-based quotients are equally weighted, the effect of frequency and time parameters on  $F$  are the same, and their differential impact on  $F_p$  is equivalent to the original parameters, that is,  $\omega_1 = 0.125$  and  $\omega_2 = \omega_3 = \omega_4 = 0.5$ . When  $F$  or  $F_p$  reaches a global minimum, the corresponding model can accurately reproduce the target glottal waveform.

**3.3. Optimization Algorithm.** Gradient techniques have proven to be inadequate, since the objective function is non-convex and contains many local minima. The evolutionary algorithm has high robustness, and broad applicability for global optimization can deal with complex problems that traditional optimization algorithms cannot solve. Particle swarm optimization (PSO) and genetic algorithm (GA) are

similar but have various strengths in dealing with different problems [28].

Therefore, we combined their advantages. PSO is an evolutionary computation technique based on swarm intelligence and is a community-based optimization tool. The PSO algorithm first initializes a group of random particles with random solutions and then all individuals and the best individuals of groups breed. The optimal solution is found through an iterative process. We added selection and crossover processes similar to GA into PSO, generating a GPSO algorithm.

In contrast, the quasi-Newton method is commonly used for solving nonlinear optimization problems, where the gradient of the objective function at each iteration step is obtained. An objective function can be constructed from the measured gradient to produce superlinear convergence. However, this method is somewhat sensitive to the initial point, and results are mostly local optima. Therefore, we combined the GPSO and quasi-Newton algorithm (GPSO-QN) to optimize the biomechanical model parameters to match the target voice sources.

The masses, spring constants, coupling coefficients, damping constants, and subglottal pressure all need optimization, which can be expressed as a vector  $\Phi = [m_{ia}, k_{ia}, k_{ca}, r_{ia}, P_s]$ . With optimized  $\Phi$  the model should simulate  $U_g$  in good agreement with  $u_g$ .

Previous analysis has shown that asymmetric pathological vocal cords are the leading cause of irregular vibration. Consequently, we took  $\Phi$  and  $\beta$  as matching parameters with the search interval  $[m_{ia}, k_{ia}, k_{ca}, r_{ia}] \in [0.001, 0.5]$ ,  $\beta \in [0.4, 1]$ , and  $P_s \in [0.001, 0.05]$ . Then suitable matching parameters

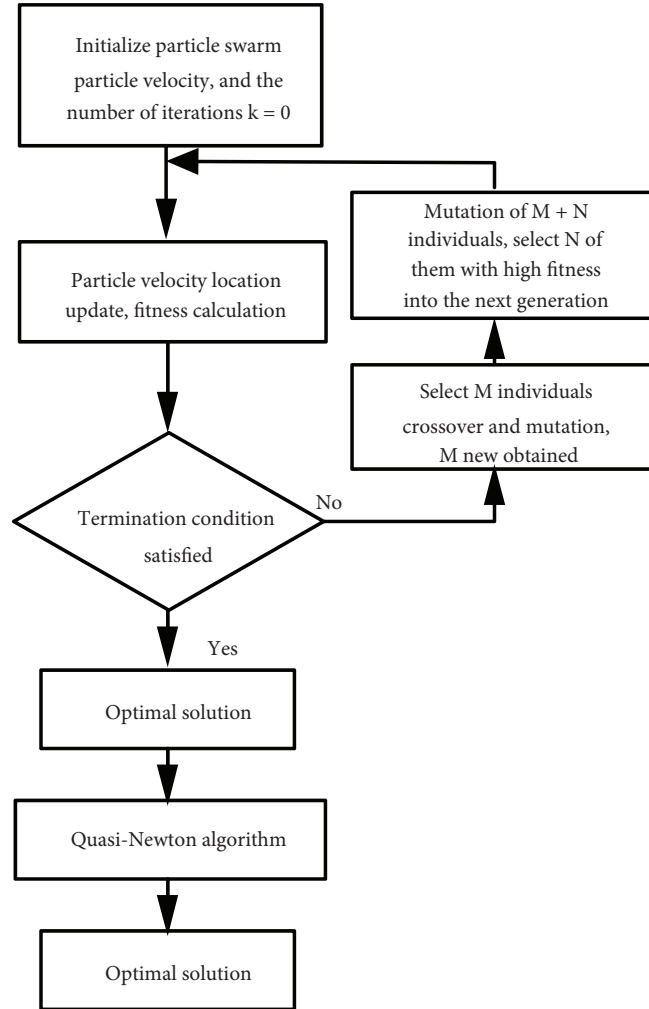


FIGURE 8: Proposed GPSO-QN algorithm structure.

can be obtained using the proposed GPSO-QN algorithm to ensure the optimized model accurately reproduces the glottal waveform.

To avoid obtaining local minima in a nonconvex search space by direct application of the gradient method, the GPSO algorithm is first applied to provide a rough approximation, and then the QN method is applied to locally optimize the approximate solution, providing the globally optimal result.

Figure 8 shows the parameter optimization process. Selection and crossover process utilizes the Monte Carlo selection rule to choose  $M$  individuals. The termination condition is that the obtained maximum fitness value exceeds a preset threshold or the preset number of iterations is reached.

## 4. Result and Discussion

**4.1. Experimental Parameters.** This paper selected sustained vowel /a/ from the MEEI database [29], numbering the samples 1–8 (4 normal and 4 paralysis voices). Sampling frequency was 25 kHz, and the proposed GPSO-QN algorithm was used to optimize the model parameters with

the number of particles for the initial population set as 30 and the number of generations limited to 400. Learning factors  $c_1$  and  $c_2$  were set=2, and the range of weight coefficient  $\omega$  was set=[0.5, 0.9].

**4.2. Normal Voice Source Matching.** Figure 9 shows the excitation sources (red dashed lines) extracted from the four normal voice samples using the optimized model were accurately simulated. Using sample 3 as an example, Figure 10 shows that the simulated and actual spectra also have good consistency.

**4.3. Paralysis Voice Source Matching.** Figure 11 shows that the model simulated waveforms for paralyzed voice samples (red dashed lines) have significant errors to actual samples, particularly for samples 7 and 8. However, the spectra show good consistency with only magnitude bias, as shown in Figure 12.

**4.4. Difference Analysis of Matching Results.** To investigate the differences between normal and paralysis voice sources, we matched 9 consecutive frames of samples 1–8, and Figure 13 shows the statistical distribution of the optimized



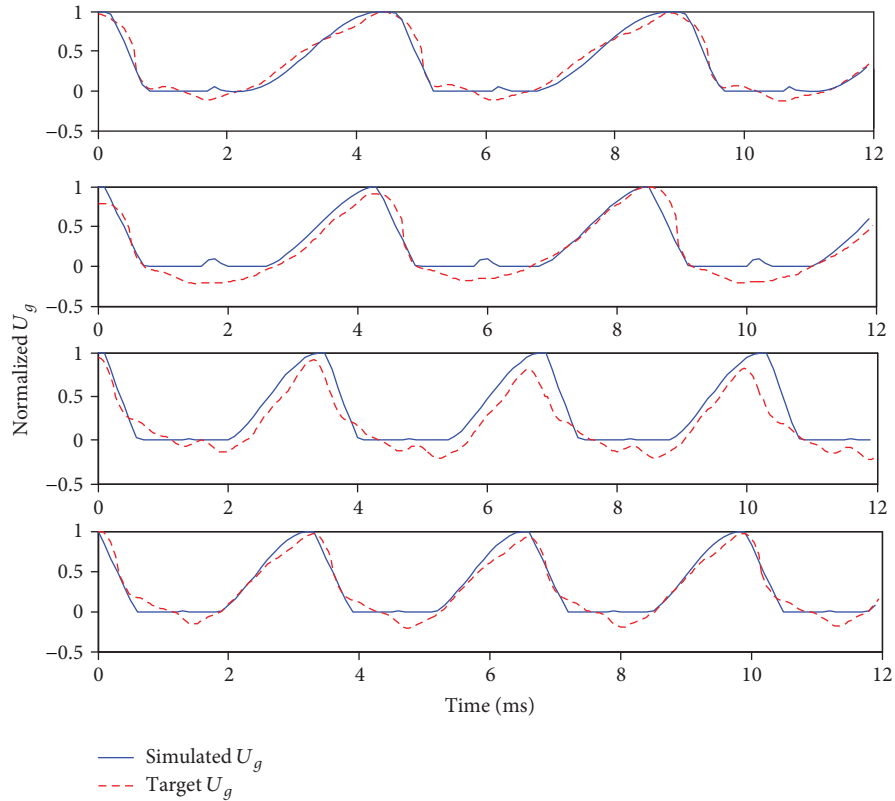


FIGURE 9: Matching result of normal voice source in the time domain.

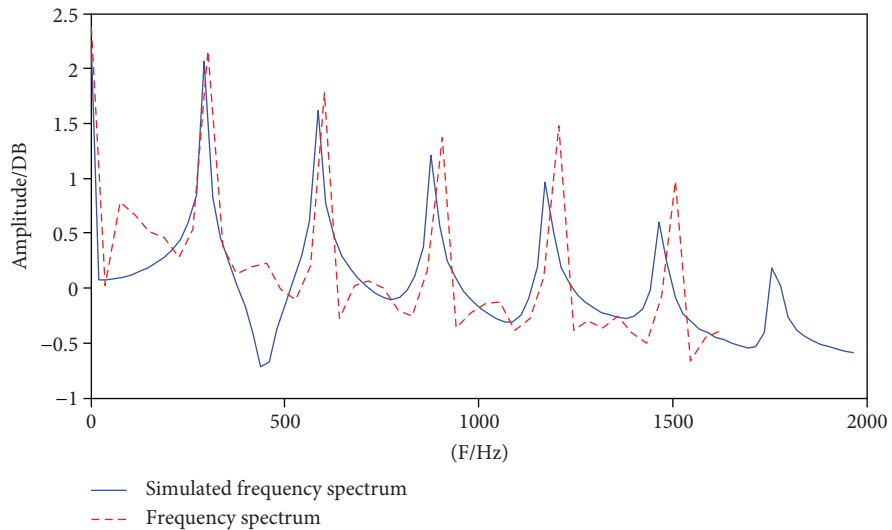


FIGURE 10: Matching result of sample 3 voice source in the frequency domain.

parameters. There were no significant differences between stiffness, quality, and damping of normal and paralysis models. However, the coupling stiffness of paralyzed vocal voice sources is greater than that of normal sources, and significant asymmetry in the paralyzed vocal cords was observed, as shown in the last two rows of Figure 13(b).

Therefore, coupling stiffness and the asymmetry parameter,  $\beta$ , could be used as a basis for classifying normal and paralyzed vocal sources. Figure 14 shows the pathological

voice source analysis system. It is designed and programmed by MATLAB.

### 5. Conclusion

This study analyzed nonlinear characteristics of asymmetric vocal cord motion using an optimized biomechanical model to design a pathological voice source analysis system. A proposed algorithm was employed to optimize the

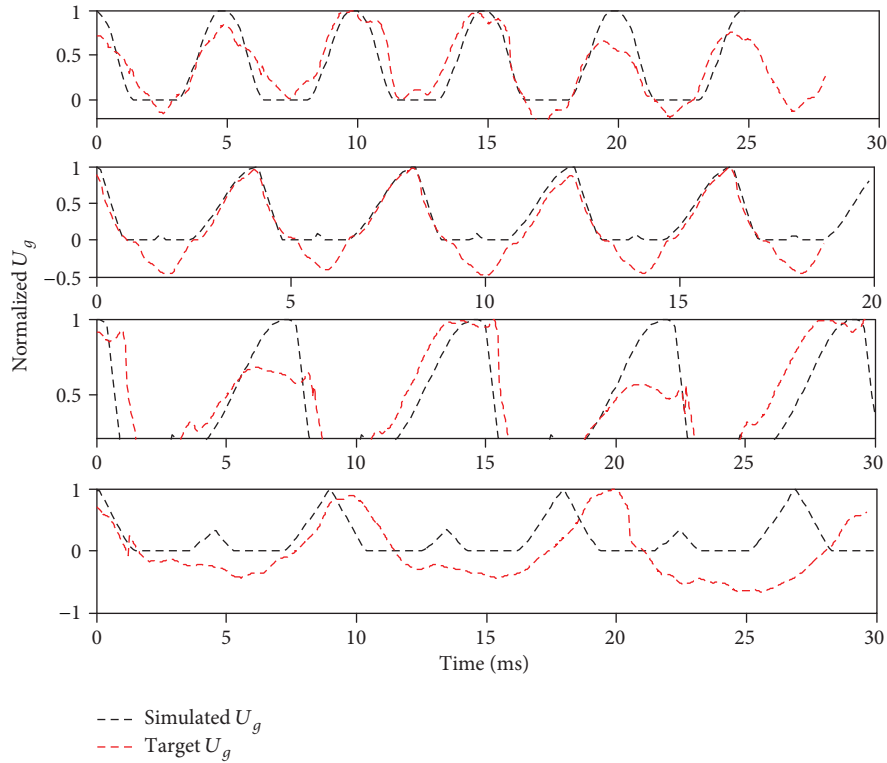


FIGURE 11: Matching result of paralysis voice source in the time domain.

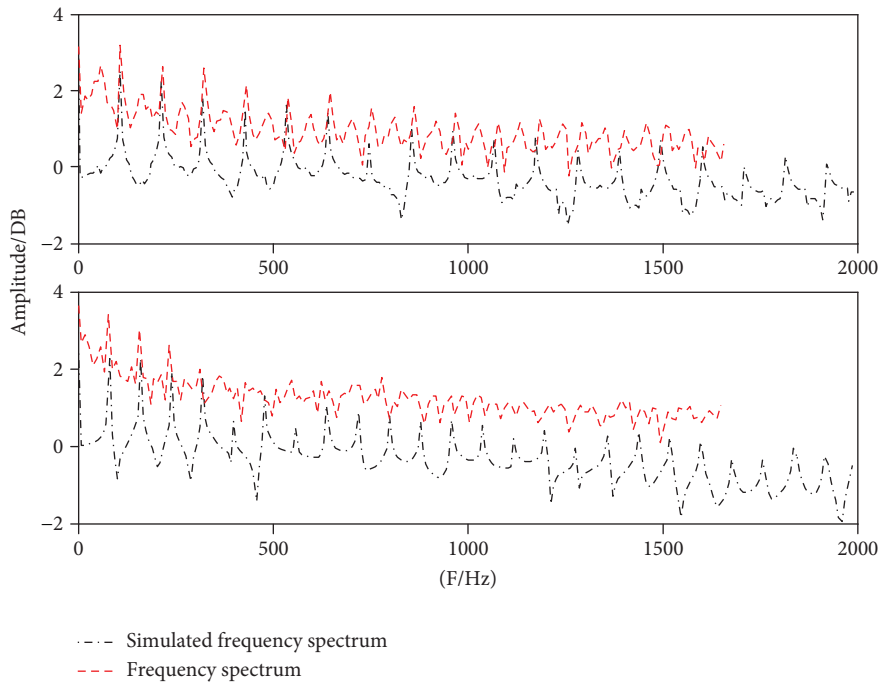


FIGURE 12: Matching results of samples 7 and 8 in the frequency domain.

masses, spring constants, coupling coefficient, damping constants, asymmetry parameter, and subglottal pressure of the mass model.

The proposed biomechanical model accurately simulated irregular vibration caused by unbalanced vocal tension.

Period doubling bifurcation and frequency entrainment were observed in the bifurcation and phase diagrams, and spectrograms.

Vibration system complexity and asymmetry do not have a simple proportional relationship. This study shows that

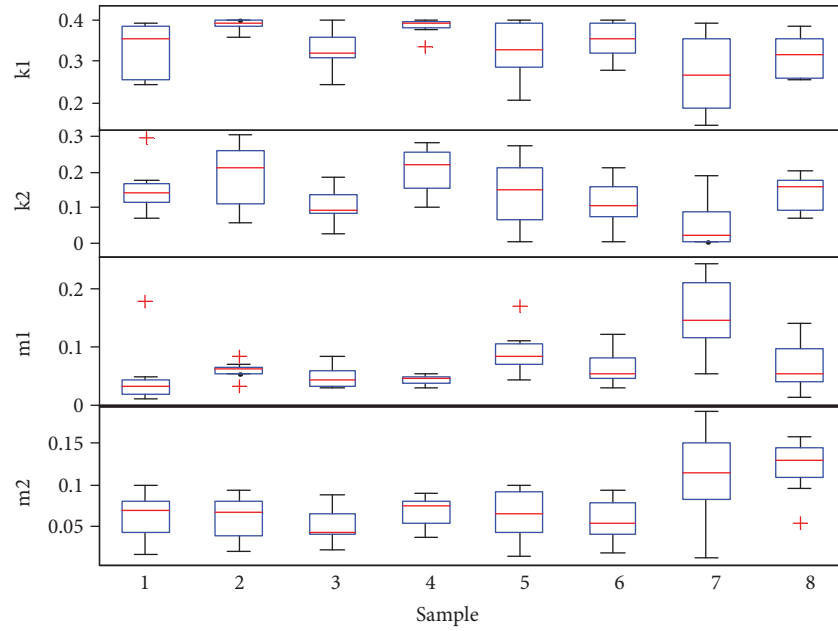


FIGURE 13: Statistical distribution of normal (1-4) and paralyzed samples (5-8).

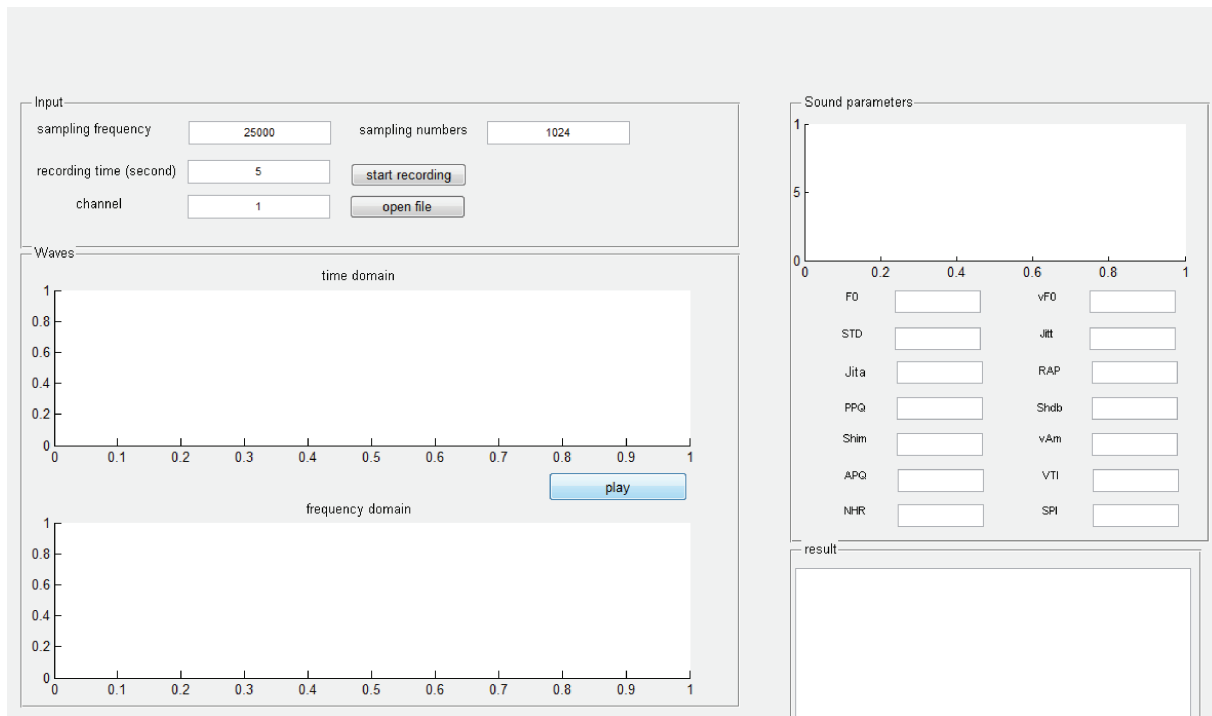


FIGURE 14: Pathological voice source analysis system.

pitch frequency is mainly affected by the asymmetric structure of the vocal cord, whereas the impact of subglottal pressure is relatively small.

The optimal biomechanical model can accurately reproduce the voice source stream modulated by asymmetric vocal cords. Although the physiological parameters of voice sources were different, the asymmetry and coupling stiffness parameters helped determine paralysis voice sources.

Optimized model simulations will be of great value for understanding clinical hoarse voices corresponding to asymmetric vocal structure and predicting the effect on unilateral vocal disease treatment.

Future work will establish rational sound models for vocal cord polyps and other organic diseases to match real voice sources, assisting in classification of vocal cord diseases.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This project was funded by National Natural Science Foundation of China under Grant no. 61372146 and no. 61271359. The authors are thankful for the support from Collaborative Innovation Centre of Suzhou Nano Science and Technology and Key Joint Laboratory of Jiemei Biomedical Engineering, Soochow University.

## References

- [1] Q. Xue, R. Mittal, X. Zheng, and S. Bielamowicz, "Computational modeling of phonatory dynamics in a tubular three-dimensional model of the human larynx," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1602–1613, 2012.
- [2] A. Granados, M. K. Misztal, J. Brunskog, V. Visseq, and K. Erleben, "A numerical strategy for finite element modeling of frictionless asymmetric vocal fold collision," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 33, no. 2, 2017.
- [3] J. Lohscheller, J. G. Švec, and M. Döllinger, "Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects," *Logopedics Phoniatrics Vocology*, vol. 38, no. 4, pp. 182–192, 2013.
- [4] L. Verde, G. De Pietro, and G. Sannino, "A methodology for voice classification based on the personalized fundamental frequency estimation," *Biomedical Signal Processing and Control*, vol. 42, pp. 134–144, 2018.
- [5] N. Huang, Y. Zhang, W. Calawerts, and J. J. Jiang, "Optimized nonlinear dynamic analysis of pathologic voices with laryngeal paralysis based on the minimum embedding dimension," *Journal of Voice*, vol. 31, no. 2, pp. 249.e1–249.e7, 2017.
- [6] E. Cataldo and C. Soize, "Voice signals produced with jitter through a stochastic one-mass mechanical model," *Journal of Voice*, vol. 31, no. 1, pp. 111.e9–111.e18, 2017.
- [7] P. S. Deshpande and M. S. Manikandan, "Effective glottal instant detection and electroglottographic parameter extraction for automated voice pathology assessment," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 398–408, 2018.
- [8] J. J. Jiang, Y. Zhang, and C. McGilligan, "Chaos in voice, from modeling to measurement," *Journal of Voice*, vol. 20, no. 1, pp. 2–17, 2006.
- [9] L. Lancia, D. Voigt, and G. Krasovitskiy, "Characterization of laryngealization as irregular vocal fold vibration and interaction with prosodic prominence," *Journal of Phonetics*, vol. 54, no. 9, pp. 80–97, 2016.
- [10] M. Zaňartu, D. D. Mehta, J. C. Ho, G. R. Wodicka, and R. E. Hillman, "Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study," *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 326–339, 2011.
- [11] N. Wan, D. D. Peng, M. Sun, and D. Zhang, "Nonlinear oscillation of pathological vocal folds during vocalization," *Science China Physics, Mechanics and Astronomy*, vol. 56, no. 7, pp. 1324–1328, 2013.
- [12] A. P. Pinheiro and G. Kerschen, "Vibrational dynamics of vocal folds using nonlinear normal modes," *Medical Engineering & Physics*, vol. 35, no. 8, pp. 1079–1088, 2013.
- [13] A. P. Pinheiro, D. E. Stewart, C. D. Maciel, J. C. Pereira, and S. Oliveira, "Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling," *Digital Signal Processing*, vol. 22, no. 2, pp. 304–313, 2012.
- [14] Y. Zhang, A. J. Sprecher, Z. X. Zhao, and J. J. Jiang, "Nonlinear detection of disordered voice productions from short time series based on a Volterra–Wiener–Korenberg model," *Chaos, Solitons & Fractals*, vol. 44, no. 9, pp. 751–758, 2011.
- [15] Y. Zhang and J. J. Jiang, "Chaotic vibrations of a vocal fold model with a unilateral polyp," *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1266–1269, 2004.
- [16] S. H. Choi, Y. Zhang, J. J. Jiang, D. M. Bless, and N. V. Welham, "Nonlinear dynamic-based analysis of severe dysphonia in patients with vocal fold scar and sulcus vocalis," *Journal of Voice*, vol. 26, no. 5, pp. 566–576, 2012.
- [17] Q. Xue, R. Mittal, X. Zheng, and S. Bielamowicz, "A computational study of the effect of vocal-fold asymmetry on phonation," *The Journal of the Acoustical Society of America*, vol. 128, no. 2, pp. 818–827, 2010.
- [18] D. E. Sommer, B. D. Erath, M. Zaňartu, and S. D. Peterson, "Corrected contact dynamics for the Steinecke and Herzel asymmetric two-mass model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. EL271–EL276, 2012.
- [19] S. Panchapagesan and A. Alwan, "A study of acoustic-to-articulatory inversion of speech by analysis-by-synthesis using chain matrices and the Maeda articulatory model," *The Journal of the Acoustical Society of America*, vol. 129, no. 4, pp. 2144–2162, 2011.
- [20] M. Dollinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, and U. Eysholdt, "Vibration parameter extraction from endoscopic image series of the vocal folds," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 8, pp. 773–781, 2002.
- [21] P. Gómez-Vilda, R. Fernández-Baillo, V. Rodellar-Biarge et al., "Glottal source biometrical signature for voice pathology detection," *Speech Communication*, vol. 51, no. 9, pp. 759–781, 2009.
- [22] P. J. Hadwin, G. E. Galindo, K. J. Daun et al., "Non-stationary Bayesian estimation of parameters from a body cover model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2683–2696, 2016.
- [23] X. L. Zeng, X. J. Zhang, D. Wu et al., "Parameter inversion method of vocal fold dynamic model in pathological voice classification," *Acta Acustica*, vol. 43, no. 1, pp. 91–100, 2018.
- [24] C. Tao, Y. Zhang, and J. J. Jiang, "Extracting physiologically relevant parameters of vocal folds from high-speed video image series," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 794–801, 2007.

- [25] T. Riede, Z. Li, I. T. Tokuda, and C. G. Farmer, "Functional morphology of the Alligator mississippiensis larynx with implications for vocal production," *The Journal of Experimental Biology*, vol. 218, no. 7, pp. 991–998, 2015.
- [26] Z. Zhang, J. Kreiman, B. R. Gerratt, and M. Garellek, "Acoustic and perceptual effects of changes in body layer stiffness in symmetric and asymmetric vocal fold models," *The Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. 453–462, 2013.
- [27] P. Alku, "Parameterisation methods of the glottal flow estimated by inverse filtering," in *In ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis*, pp. 81–87, International Speech Communication Association, 2003.
- [28] X. H. Shi, Y. C. Liang, H. P. Lee, C. Lu, and L. M. Wang, "An improved GA and a novel PSO-GA-based hybrid algorithm," *Information Processing Letters*, vol. 93, no. 5, pp. 255–261, 2005.
- [29] Massachusetts Eye and Ear Infirmary, *Voice Disorders Database (Version 1.03)*, Kay Elemetrics Corp, Lincoln Park, NJ, 1994.