

病理嗓音发声系统的非对称建模研究

陶智 曾晓亮 顾玲玲 张晓俊 吴迪 薛隆基

(苏州大学物理与光电·能源学部, 苏州, 215006)

摘要: 为了在病理嗓音识别中为特征参数选择提供依据, 提出声带非对称力学建模仿真病变声带并进行分析研究。依据声带的分层结构和组织特性, 建立声带力学模型, 耦合声门气流, 求取模型输出的声门源激励波形。采用遗传粒子群-拟牛顿结合优化算法(Genetic particle swarm optimization based on quasi-Newton method, GPSO-QN)将模型输出的声门源和实际目标声门波相匹配, 提取优化模型参数。仿真实验结果表明, 该声带模型能产生与实际声门源相一致的声门波形, 同时也证明了左右声带生理组织间的非对称性是产生病理嗓音的重要原因。

关键词: 病理嗓音; 非对称; 声带力学模型; 遗传粒子群-拟牛顿结合优化算法

中图分类号: TN912.3 **文献标志码:** A

Research on Asymmetric Model of Vocal Fold in Pathology Voice

Tao Zhi, Zeng Xiaoliang, Gu Lingling, Zhang Xiaojun, Wu Di, Xue Longji

(College of Physics, Optoelectronics and Energy, Soochow University, Suzhou, 215006, China)

Abstract: To provide the basis for parameter selection of pathological voice recognition, an asymmetric modeling method is proposed to simulate diseased vocal fold. According to the layered structure and tissue properties of the vocal fold, a mechanical model is set up to produce the voice source with the straight airflow expelled by lungs. An inversion procedure adopting genetic particle swarm optimization based on quasi-Newton method (GPSO-QN) is developed to adjust the parameters of the vocal fold model and to reproduce the targeted voice source. Experimental results show that the vocal fold mechanical model can produce the voice source that is consistent with the target. In addition, the optimized parameter sets show that the asymmetries of two opposing vocal folds result in the pathology voice.

Key words: pathology voice; asymmetry; vocal fold mechanical model; GPSO-QN

引 言

肺部呼出的气流造成两侧声带的相互振动, 因此声门出现周期性的开闭, 由此产生一系列周期性的气流脉冲, 这种气流脉冲则为“声门波”。声带组织结构的病变则会使其振动变得无规律, 导致声门波出现明显的变化, 在嗓音方面体现为嘶音或是呼吸音。临床上常采用电子喉镜的方式用于观测这种无规律振动从而辅助声带疾病的诊断, 但喉镜检查带有侵入性, 且过多依赖医师的主观判断^[1]。因此, 在病

理嗓音的识别研究中,寻求客观检测方法成为关键。已有研究主要采用声学分析法,即利用声学特征参数结合模式识别的方法来辅助病理嗓音识别分类^[2],尽管该方法能实现远距离、快速且非侵入地诊断,但其采用的声学信号仅能提供临床声带疾病的症状信息,无法提供病变声带的无规律振动信息^[3],使得所用声学特征参数与声带实际组织结构无关,因此对导致嗓音疾病的声带组织结构的改变需进行更加深入的研究。

病变常常发生于单侧声带,从而导致了两侧声带张力的不平衡,而有关研究表明这种张力不平衡是造成声带无规律振动的主要原因^[4],因此构建声带模型有助于观测非对称声带的无规律振动特征。已有研究提出了大量声带建模方法,其中最简单且最经典的用于描述声带振动特征的物理模型为二质量块模型(Two-mass model, 2MM),它最早由 Ishizaka 和 Flanagan 提出^[5],后经 Steinecke 和 Herzel 进行简化^[6]。声带模型的振动需要肺部呼出气流的驱使,常用的描述气流运动的方程为纳维-斯托克斯方程,文献^[7]利用该方程实现了简化 2MM 模型的振动,同时也利用该模型分析了因声带非平衡张力导致的无规律振动。

尽管物理建模方法在语音合成和嗓音分析等方面有着广泛的应用前景,但只有将物理模型匹配实际观测数据才有应用价值^[8,9]。已有研究通过声带实际振动波形结合智能优化算法优化所建声带模型的模型参数,文献^[10]通过 Nelder-Mead 单纯形法优化声带模型缩小了高速摄像记录的实际声门面积序列与模型输出的声门面积序列之间的误差,由于使用了侵入性的高速摄像技术,该方法对患者造成一定的不便。文献^[11]基于声门波的能量谱密度提取了病变声带的模型参数,但选用的一质量块模型无法完整地体现实际声带振动特征。也有研究通过遗传算法来匹配声门面积波形^[12]、声带临界振动波形^[13]等优化声带力学模型,但所涉及的模型参数众多,反演模型参数的步骤复杂,使得反演结果时好时坏。因此选取合适的智能优化算法,并一定程度地减少优化模型参数的数量在优化声带模型方面体现的尤为重要。

本文采用简化 2MM 模型(Steinecke-Herzel model, SH)来描述实际声带,通过引入非对称参数 β 来仿真病变声带,并利用相图技术研究了实际声带组织结构的非对称性导致的张力不平衡对声带振动造成的不同程度的影响。同时,本文提出遗传粒子群-拟牛顿优化算法(Genetic particle swarm optimization algorithm based on quasi-Newton method, GPSO-QN)对声带 SH 模型进行优化,提取了正常声带和病变声带的优化模型参数。实验表明,本文提出的智能优化算法使声带 SH 模型能准确再现实际声带振动特征,优化模型参数适用于辅助病理嗓音的识别分类。

1 声带非对称模型

1.1 声带 SH 模型

声带由两个左右对称的喉头粘膜结构组成,从肺部产生的气流经由气管输出不断作用在声带上,造成了声带的规律振动^[14]。根据声带的张力性质及组织结构特性,采用两个耦合弹簧振子表示单侧声带,声门下压提供空气动力使得两侧声带模型产生振动,声带 SH 模型如图 1 所示。

假设作用在两侧声带的气流压力满足伯努利定律,射流在声门最窄处形成,同时忽略声门上、下气流间的相互作用、弹力的非线性作用及声门处的粘性损失,弹簧振子的运动方程表示为

$$\begin{cases} \dot{x}_{1a} = v_{1a} \\ \dot{v}_{1a} = -\frac{1}{m_{1a}}(F_{1a} + I_{1a} - r_{1a}v_{1a} - k_{1a}x_{1a} - k_a(x_{1a} - x_{2a})) \\ \dot{x}_{2a} = v_{2a} \\ \dot{v}_{2a} = -\frac{1}{m_{2a}}(I_{2a} - r_{2a}v_{2a} - k_{2a}x_{2a} - k_a(x_{2a} - x_{1a})) \end{cases} \quad (1)$$

式中:

$$F_{1\alpha} = LdP_s \left[1 - \Omega(a_{\min}) \left(\frac{a_{\min}}{a_1} \right)^2 \right] \Omega(a_1); F_{2\alpha} = 0;$$

$$\Omega(x) = \begin{cases} \tanh \left[50 \left(\frac{x}{x_0} \right) \right] & x > 0 \\ 0 & x \leq 0 \end{cases};$$

$$\Theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}; I_{i\alpha} = -\Theta(-a_i) \frac{c_{i\alpha} a_i}{m_{i\alpha} 2L};$$

$$a_i = a_{0i} + L(a_{il} + a_{ir}); a_{\min} = \min(a_1 + a_2);$$

$$U_g = \sqrt{\frac{2P_s}{\rho}} \cdot a_{\min} \cdot \Omega(a_{\min}).$$

其中 $i=1, 2$ 分别为上、

下质量块, $\alpha=l, r$ 分别为左、右侧声带; P_s 为声门下压; $x_{i\alpha}, v_{i\alpha}$ 分别为质量块振动位移及速度; $m_{i\alpha}, k_{i\alpha}, k_{c\alpha}, r_{i\alpha}$ 分别为质量块重量、弹簧弹性系数、耦合弹性系数与阻尼常数; L, d 为声带长度与下层质量块厚度; $c_{i\alpha} = 3k_{i\alpha}$ 为声带碰撞时的附加弹性系数; $F_{1\alpha}, I_{i\alpha}$ 分别为伯努利压力及声带碰撞所产生的冲击力; a_i 为声门间隙面积; U_g 为声门气流体速度。

根据正常声带的生理结构特性,文献[5]为 SH 模型给出了标准模型参数。 $m_{1\alpha} = 0.125 \text{ g}, m_{2\alpha} = 0.025 \text{ g}, k_{1\alpha} = 0.08 \text{ N/cm}, k_{2\alpha} = 0.008, k_{c\alpha} = 0.025, r_{1\alpha} = r_{2\alpha} = 0.02, d_1 = 0.25 \text{ cm}, d_2 = 0.05 \text{ cm}, a_{01} = a_{02} = 0.05 \text{ cm}^2, l = 1.4 \text{ cm}, \rho = 0.00113 \text{ g/cm}^3, P_s = 800 \text{ Pa}$ 。式(1)可采用标准四阶龙格-库塔方法解得。

1.2 病变声带模型及振动分析

单侧声带病变导致病理嗓音的情况较为多见,而声带组织结构的非对称性导致声带张力的不平衡,这种张力不平衡最终导致无规律的声带振动,从而产生病理嗓音。因此,病变声带模型可以假设为单侧声带病变,另一侧声带正常,通过调节单侧声带的模型参数研究病变声带的振动情况。

考虑到声带组织结构的非对称性,引入非对称参数 $\beta (0.4 < \beta \leq 1)$ 。假设左侧声带发生病变,右侧声带正常,则左侧声带模型参数变化为

$$\bar{m}_l = m_l / \beta \quad \bar{k}_l = k_l \cdot \beta \quad \bar{k}_c = k_c \cdot \beta \quad \bar{c}_l = c_l \cdot \beta \quad (2)$$

式中: β 表示两侧声带物理结构非对称程度, β 值越小,左右声带越不对称。当声带病变导致结构参数变化到超过临界值时,系统振动会出现倍周期分岔现象。本文通过非对称参数 β 来表示声带病变程度,因此可以将 β 作为系统的分岔参数。本文选取声带模型左下侧质量块振动位移 x_{1l} 和右下侧质量块振动位移 x_{1r} 作为一组状态变量, $x_{1l}-x_{1r}$ 的状态轨迹可用于描述该模型系统的振动行为。如图 2 为 β 取不同值 ($\beta=1, 0.8, 0.6, 0.53, 0.45$) 时的 $x_{1l}-x_{1r}$ 相图。从图 2 中可以看出,当 β 取 0.8 时,相轨迹为一个极限环,表明分岔现象并没发生;直到 β 取到 0.53 或更小时,相轨迹出现倍周期极限环,表明系统产生了倍

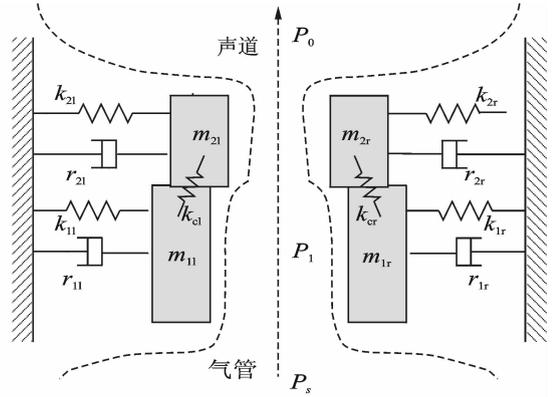


图 1 声带 SH 模型结构示意图

Fig. 1 Structure diagram of vocal fold SH model

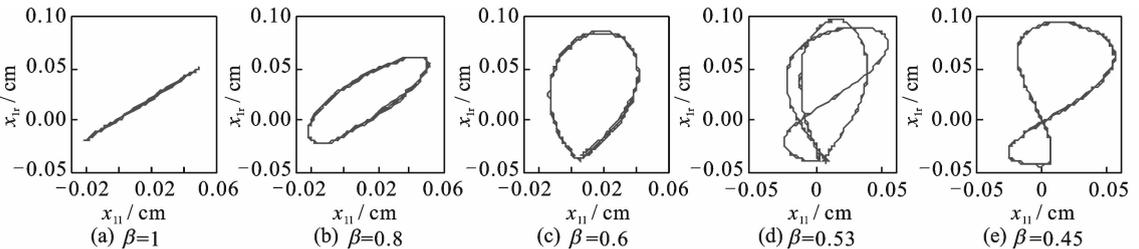


图 2 $x_{1l}-x_{1r}$ 平面的相空间轨迹

Fig. 2 Trajectories of phase space on plane of $x_{1l}-x_{1r}$

周期分岔。

2 模型物理参数优化

2.1 实际喉声源估计

本文实际喉声源的提取采用 Alku 提出的自适应迭代滤波算法(Iterative adaptive inverse filtering, IAIF)^[15]。采用最小均方误差的线性预测方法对声道滤波器进行估计,利用估计得出的声道滤波器对实际语音去除声道影响,得到实际声门波。算法原理如图 3 所示,首先语音信号 $s(n)$ 经过一阶滤波器 $H(z)$ 去除口唇辐射影响,然后用声门脉冲的 K_1 ($K_1=1\sim 3$) 阶滤波器 $H_g(z)$ 估计声门共振对嗓音的影响,滤波可消除声门对嗓音的影响。之后可建立 K_2 阶的全极点滤波器 $H_v(z)$ 作为声道模型 $F_v(z)$ (K_2 受采样率影响),声道对嗓音的影响则可通过滤波消除,此处则得到了声门激励源 U_g 的第 1 次估计。根据声门激励源 U_g 再次构建声门脉冲滤波器和声道滤波器,重复上述操作。进行 2 次迭代之后,可获得精确的实际喉声源。图 4 为采用上述算法得到的元音/a:/的声门气流波形。

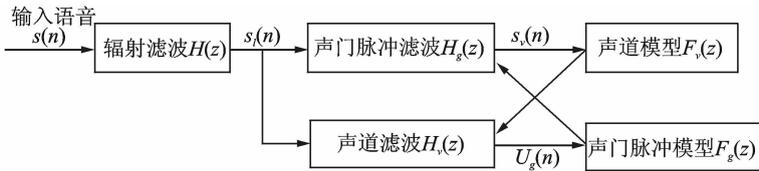


图 3 声门气流 U_g 估计的自适应迭代滤波过程

Fig. 3 Procedure of IAIF for estimation of U_g

2.2 优化算法设计及目标函数

根据上述分析,病变声带的组织结构参数发生变化,产生复杂且无规律的振动,因此优化的模型参数应包括声带质量、阻尼和张力等系数。左右声带之间的非对称性是造成病理嗓音的重要原因。同时,已有研究表明,声门下压也是影响发声的重要因素。设定模型优化参数 $\Phi = [m_w, k_w, k_c, r_w, \beta, P_s]$,各参数搜索范围为 $[m_i, k_i, k_c, r_i] \in [0.001, 0.500]$, $\beta \in [0.4, 1.0]$, $P_s \in [0.001, 0.050]$ 。

目标函数的确定是优化设计的重要组成部分,应针对优化任务、问题等特点进行具体分析。本文的病变声带模型对声门波形在时域和频域造成影响,考虑模型声门波和实际喉激励源的时域特征参数和频域特征参数^[16],包括开商(Open quotient, OQ),闭合商(Closing quotient, CIQ)速度商(Speed quotient, SQ),归一化振幅商(Normalized amplitude quotient, NAQ)和基频 F_0 。为了描述实际嗓音声门波形和模型声门波形之间的误差,定义目标函数为

$$Y(\Phi) = \omega_1 \left(\frac{|OQ - OQ^*|}{OQ} + \frac{|CIQ - CIQ^*|}{CIQ} + \frac{|SQ - SQ^*|}{SQ} + \frac{|NAQ - NAQ^*|}{NAQ} \right) + \omega_2 \frac{|F_0 - F_0^*|}{F_0} \quad (3)$$

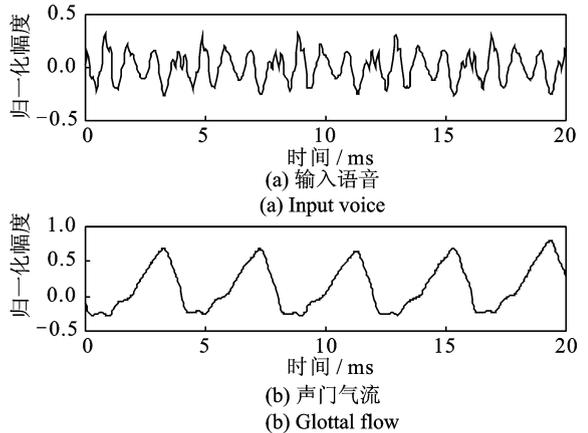


图 4 元音/a:/的声门气流波形

Fig. 4 Glottal wave curves for vowel /a:/

实验结果表明,权重 ω_1 和 ω_2 分别取 0.1 和 0.6 时,匹配效果最好。当目标函数 Y 取得全局最小值时,所建声带模型能仿真实际声带的振动。

本文确定的目标函数是非凸的,会存在很多局部最小值,演化算法在处理全局最优问题时具有较好的鲁棒性。常用演化算法包括遗传算法和粒子群优化算法,遗传算法^[17]是依据生物进化思想,采用选择、变异和交叉操作获取全局最优值的一种算法,而粒子群算法^[18,19]是利用种群中所有个体与最佳个体不断繁殖,通过迭代操作获取最优解的一种优化算法,两者在处理不同问题时有各自特点^[20],因此本文依据他们各自的优点,将两种算法进行结合。同时由于本文的目标函数问题是非线性问题,而拟牛顿算法常用于求解线性优化问题,但该算法对初始值敏感,容易产生次优解,因此本文拟将前两种算法融合拟牛顿算法,提出 GPSO-QN 算法优化声带 SH 模型。算法流程如图 5 所示。

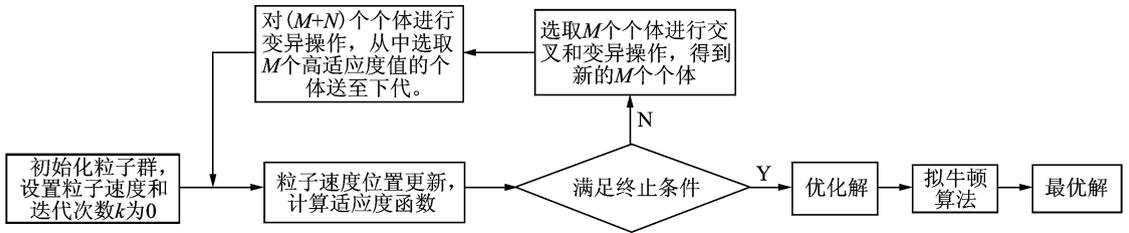


图 5 GPSO-QN 算法的流程图

Fig. 5 Flow chart of GPSO-QN

3 仿真实验结果及分析

本文从 MEEI 嗓音数据库中挑选了 4 例正常嗓音和 4 例病理嗓音,用于声带模型的优化。8 个样本分别编号为 1~8 号;在优化后模型参数的统计学分析方面,本文从数据库中挑选了 22 例正常嗓音和 44 例病理嗓音,具体样本数据如表 1 所示。选取的嗓音信号的采样率均为 25 kHz。在 GPSO-QN 优化算法面,设定初始种群的粒子数为 30,总的迭代次数不超过 400 次。学习因子 $c_1=c_2=2$,权重系数 ω 在 $[0.5, 0.9]$ 范围内取定。

表 1 嗓音数据统计情况

Tab. 1 Experimental data statistics of various voice

嗓音种类	男			女		
	人数	年龄均值	标准差/年	人数	年龄均值	标准差/年
正常	11	37.1	7.9	11	38.6	10.6
声带麻痹	11	36.5	8.7	11	38.5	7.6
声带息肉	11	35.1	6.7	11	36.7	8.5

3.1 正常及病变声带模型优化结果

图 6 为正常嗓音(1~4 号)激励源在时域上的匹配结果。从图 6 中可看出,采用该优化算法能够获得良好的仿真效果。尽管 3 号匹配准确率相对较低,但模型输出的声门波的基频和实际激励源基频仍保持一致。

病理嗓音(5~8 号)激励源的时域匹配结果如图 7 所示,匹配的效果同正常嗓音相比较是有所降低,但从图 7 可看出,主要误差是波形振幅偏差,而真正反映声门波特征的是开启时间,闭合时间等时间类参数,振幅类参数影响较小,因此对于病理嗓音,采用该优化算法可认为匹配效果也相对较好。

3.2 正常及病变声带模型优化参数的统计学分析

为了进一步分析优化所得的模型参数在描述正常嗓音和病理嗓音的优劣性,本文给出了不同嗓音

的部分优化模型参数的盒图,如图 8 所示。正常和病理嗓音的弹性系数、阻尼和质量块质量系数的差异性不大,而耦合弹簧弹性系数 k_c 和非对称参数 β 却能较好表征正常嗓音和病理嗓音。

独立样本 t 检验用 t 分布理论来推导差异发生的概率,从而比较两独立样本平均数之间的差异是否显著。本文对优化后模型参数进行 t 检验,设定置信区间为 95%。当推断概率 $P < 0.05$ 时,表明两组样

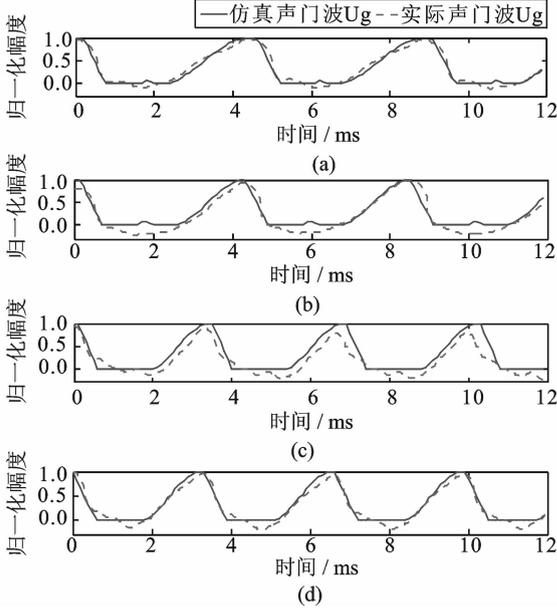


图 6 正常嗓音激励源的时域匹配结果

Fig. 6 Time-domain(TD) simulation results of glottal excitation for healthy voice

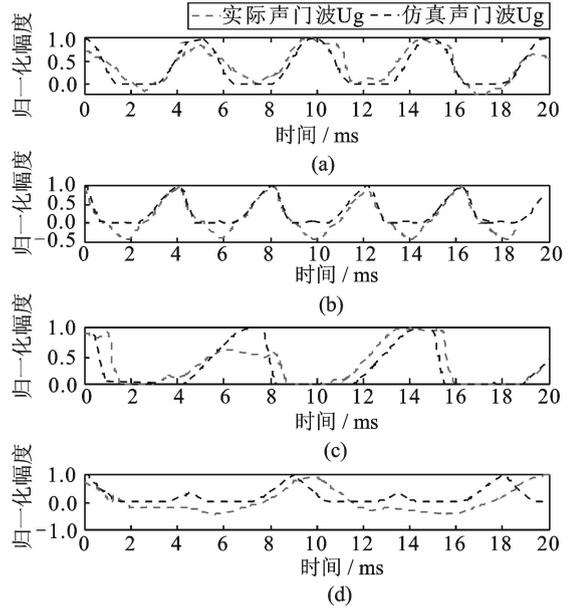


图 7 病理嗓音激励源的时域匹配结果

Fig. 7 Time-domain(TD) simulation results of glottal excitation for pathology voice

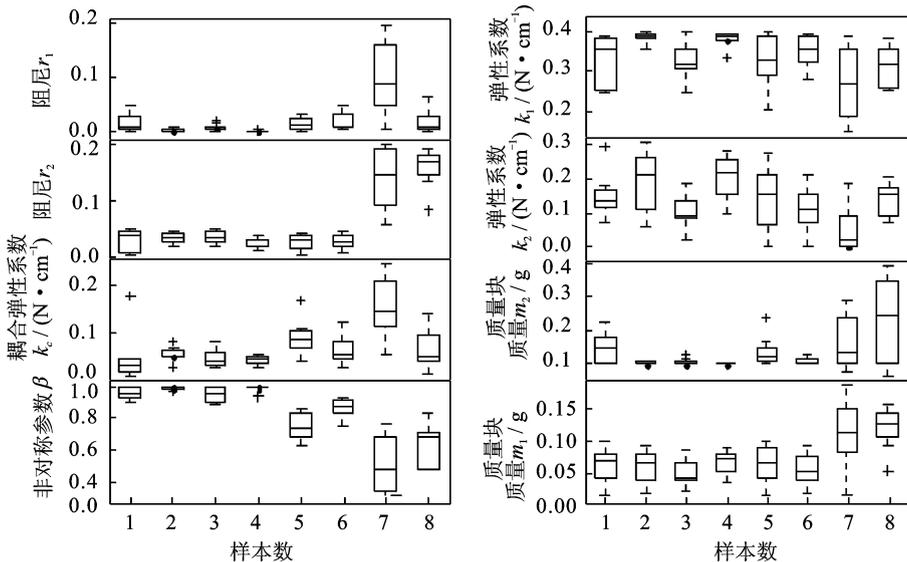


图 8 正常声带模型和病变声带模型参数的盒图分布

Fig. 8 Box-plot of model optimization parameters for normal and diseased vocal fold

本之间的差异性显著,该值越小,差异越显著。表 2 给出了各模型参数的双侧 P 值,由表 2 可看出,非对称参数 β 和耦合弹簧弹性系数 k_c 在正常和病变声带模型间的差异性较为显著,尤其是非对称参数,差异性最为显著。

表 2 正常声带模型和病变声带模型间各优化参数差异性统计表

Tab. 2 Statistical results of model optimization parameters' difference between normal and diseased vocal fold

	r_1	r_2	k_c	β	k_1	k_2	m_1	m_2
差异性	0.584	0.352	0.025	0.000	0.225	0.424	0.281	0.324

4 结束语

本文建立了声带非对称模型来表征病变声带并对其振动进行了分析,同时提出了一种智能优化算法用于优化声带模型各部分参数,包括质量块质量参数、弹簧弹性系数、耦合弹性系数、阻尼常数、非对称参数及声门下压。研究表明,该声带力学模型能仿真非平衡张力导致的无规律振动,采用相图技术对振动行为进行分析研究发现,随着非对称性的增加,系统的振动出现了倍周期分岔现象。本文采用实际嗓音声门波,对所建声带力学模型的参数进行优化,优化后的模型能准确仿真实际病理嗓音激励源,通过对优化后模型参数的统计学分析,本文发现,耦合弹性系数和非对称参数能较好地辅助区分正常和病理嗓音,同时也证实了声带间的非对称性是导致病理嗓音的重要原因。不同声带疾病导致的临床嗓音现象具有相似性,本文的理论仿真能为这些嗓音现象的产生提供很好的解释,因此采用建立声带模型结合优化算法仿真实际声带振动的方法进行病理嗓音的识别细分将具有广阔的发展前景。

参考文献:

- [1] Gavidia-Ceballos L, Hansen J H L, Kaiser J F. Vocal fold pathology assessment using AM autocorrelation analysis of the Teager energy operator[C]//4th IEEE International Conference. Philadelphia, Pennsylvania; Spoken Language, 1996, 2: 757-760.
- [2] Zhang Y, Jiang J J, Biazzo L, et al. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis[J]. *Journal of Voice*, 2005, 19(4): 519-528.
- [3] Eysholdt U, Rosanowski F, Hoppe U. Vocal fold vibration irregularities caused by different types of laryngeal asymmetry [J]. *European Archives of Oto-Rhino-Laryngology*, 2003, 260(8): 412-417.
- [4] Tokuda I, Herzel H. Detecting synchronizations in an asymmetric vocal fold model from time series data[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2005, 15(1): 013702.
- [5] Ishizaka K, Flanagan J L. Synthesis of voiced sounds from a two-mass model of the vocal cords[J]. *Bell System Technical Journal*, 1972, 51(6): 1233-1268.
- [6] Steinecke I, Herzel H. Bifurcations in an asymmetric vocal-fold model[J]. *The Journal of the Acoustical Society of America*, 1995, 97(3): 1874-1884.
- [7] Xue Q, Mittal R, Zheng X, et al. A computational study of the effect of vocal-fold asymmetry on phonation[J]. *The Journal of the Acoustical Society of America*, 2010, 128(2): 818-827.
- [8] Takemoto H, Mokhtari P, Kitamura T. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method[J]. *The Journal of the Acoustical Society of America*, 2010, 128(6): 3724-3738.
- [9] Klapuri A. Multipitch analysis of polyphonic music and speech signals using an auditory model[J]. *Audio, Speech, and Language Processing*, IEEE Transactions on, 2008, 16(2): 255-266.
- [10] Dollinger M, Hoppe U, Hettlich F, et al. Vibration parameter extraction from endoscopic image series of the vocal folds[J]. *Biomedical Engineering*, IEEE Transactions on, 2002, 49(8): 773-781.
- [11] Gómez-Vilda P, Fernández-Baillo R, Rodellar-Biarge V, et al. Glottal source biometrical signature for voice pathology detection[J]. *Speech Communication*, 2009, 51(9): 759-781.
- [12] Pinheiro A P, Stewart D E, Maciel C D, et al. Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling[J]. *Digital Signal Processing*, 2012, 22(2): 304-313.

- [13] Schwarz R, Döllinger M, Wurzbacher T, et al. Spatio-temporal quantification of vocal fold vibrations using high-speed video-endoscopy and a biomechanical model[J]. The Journal of the Acoustical Society of America, 2008, 123(5): 2717-2732.
- [14] Titze I R. Principles of voice production[M]. Utah: National Center for Voice and Speech, 2000.
- [15] Alku P. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering[J]. Speech Communication, 1992, 11(2): 109-118.
- [16] Alku P, Böckström T, Vilkman E. Normalized amplitude quotient for parametrization of the glottal flow[J]. The Journal of the Acoustical Society of America, 2002, 112(2): 701-710.
- [17] 邢英, 何炳发. 改进遗传算法在双弯曲反射面天线设计中的应用[J]. 数据采集与处理, 2013, 28(4): 520-523.
Xing Ying, He Bingfa. Improved genetic algorithm application to design of doubly curved reflector antenna[J]. Journal of Data Acquisition and Processing, 2013, 28(4): 520-523.
- [18] Kennedy J. Encyclopedia of machine learning[M]. New York: Springer, 2010: 760-766.
- [19] 张玲华, 姚绍芹, 解伟超. 基于自适应粒子群优化径向基函数神经网络的语音转换[J]. 数据采集与处理, 2015, 30(2): 336-343.
Zhang Linghua, Yao Shaoqin, Xie Weichao. Voice conversion based on adaptive particle swarm optimization radial basis function neural network[J]. Journal of Data Acquisition and Processing, 2015, 30(2): 336-343.
- [20] Shi X H, Liang Y C, Lee H P, et al. An improved GA and a novel PSO-GA-based hybrid algorithm[J]. Information Processing Letters, 2005, 93(5): 255-261.

作者简介:



陶智(1970-),男,教授,研究方向:语音信号处理、太阳能光伏, E-mail: taoz@suda.edu.cn.



曾晓亮(1991-),男,硕士研究生,研究方向:语音信号处理。



顾玲玲(1989-),女,硕士研究生,研究方向:语音信号处理。



张晓俊(1981-),男,实验师,研究方向:语音信号处理。



吴迪(1980-),男,讲师,研究方向:语音信号处理、人工神经网络智能计算。



薛隆基(1994-),男,本科,研究方向:语音信号处理。

