DOI:10.16450/j.cnki.issn.1004-6801.2023.06.023

基于功率谱的美声发声特征提取*

张 凯[1], 王舒蕾[1], 齐婷婷[2], 张义民[1]
(1.沈阳化工大学装备可靠性研究所 沈阳,110042) (2.沈阳音乐学院戏剧影视学院 沈阳,110818)

摘要 为了更好地研究歌唱中科学发声的特点,提出了一种基于功率谱的美声发声特征提取方法。首先,利用Burg法功率谱估计分别对正确和错误的美声信号进行分析;其次,针对功率谱曲线差异较大的地方,运用最小二乘法进行了函数多项式的拟合;最后,提取其多项式系数作为特征。采集了某音乐学院3名美声老师和5名美声新生共400条女高音信号并进行分析,结果表明,在功率谱曲线的5 kHz和10 kHz处,2种信号有着较大的差异,经过上述方法提取特征后,根据其样本的箱式图即可以明显区别正确与错误发声,识别率可达100%。相比之下,如果直接运用反向传播(back propagation,简称BP)神经网络识别功率谱信号,其识别率仅为95.23%。该研究成果从振动理论的角度对美声发声的辅助训练提供了技术支撑。

关键词 音频信号;特征提取;美声唱法;女高音;功率谱
中图分类号 TP391;TN912.3

引 言

美声唱法由于音色清脆高亢、灵活多变及音量较大[1],对于歌唱者的发声技巧要求较多,且美声唱法的共鸣是“所有腔体共同运作达到整体效果的展现”。相比于其他唱法,美声唱法需要共鸣腔体以及骨骼都参与共鸣,即要求身体的各个器官放在一起共同产生共鸣。其他唱法参与共鸣的器官相对较少,发声的位置也有所不同[2],导致美声初学者在头腔、口腔、胸腔和咬字等方式上相对于其他唱法出现的问题较多。目前,在声乐领域的教学中,基本是通过老师的言传身教来纠正学生歌唱技巧上的错误。为了更深入研究美声发声的特点,笔者利用美声发声信号的功率谱去评价初学者的发音状态,从振动理论的角度比较发音的异同,从功率谱中提取美声发声的信号特征。

国内外学者围绕美声发声原理开展了相关研究。文献[3-5]从声门振动和空气动力学的角度对声音信号进行了分析。Mayr[6]利用长期平均频谱(long-term average spectrum,简称LTAS)和功率谱对美声男高音的生理和声学特征进行了研究,比较了假音和胸腔音的差异。Souza[7]通过对女高音的共振峰分析比较,得到音高的变化会导致基频和共振峰的不同。Hasan等[8]使用经验模态分解(empirical mode decomposition,简称EMD)方法对歌曲的清音和浊音进行能量估计,以观察学习者歌声中的差异和错误。Zysk等[9]设计了一套声音记录程序,利用频谱特征对女高音的头部和胸部音域表演进行分类。Barlow等[10]根据平均元音谱(average vowel spectra,简称AVS)和长期平均谱对歌手在古典和现代风格之间的声乐作品的差异进行了量化。

国内学者的研究主要集中在美声唱法与民族唱法、流行唱法的融合与对比领域[11-13],但针对声音信号特点进行研究的文献较少。钱一凡等[14]针对标准元音提取了其基频、共振峰和各通道振幅,比较不同元音的声学特征,分析得知不同的元音发声与身体不同部位的共鸣有关。

大部分关于发声信号的研究采用傅里叶变换的方法,将原时域信号转化为频域信号。然而,频域信号仅对变换后信号的实部进行对比,忽略了相频信息。另外,对美声唱法样本的采集主要集中在美声与通俗唱法的对比上,但是通俗唱法从发声特点上与美声唱法存在明显差异,难以突出美声声音信号的特殊性。

针对上述问题,笔者利用功率谱的估计对信号进行研究,即从能量的观点对信号进行分析,保留频谱法所丢掉的相位信息。同时,从美声初学者与歌唱技巧成熟的美声老师中提取样本并进行对比研

* NSFC-辽宁联合基金资助项目(U1708254);国家重点研发计划专项资助项目(2019YFB2004400)
收稿日期:2021-06-14;修回日期:2021-09-20

究。因为美声初学者的发音近似美声，所以更适合对美声发音的规范性进行系统评价。

1 基于功率谱发声信号特征提取步骤

笔者对美声声音信号的特征提取主要分为以下步骤：①对声音信号进行采集；②对采集到的声音信号进行端点检测处理，去除无用的语音段；③对处理后的信号做Burg法功率谱分析；④将得到的功率谱进行局部二次回归平滑处理。

1.1 声音信号样本的采集与端点检测

对5名美声初学者和3名美声老师进行女高音信号的采集、筛选和分类。录音时要求发音人在相同录音环境下依次清唱出基础元音/a/，/i/和/u/，在录制的声音样本中选取发声时长在3～5 s的语音信号，最终得到老师的发音样本50条（设定为正确发声信号）和学生的错误发音样本350条。美声老师分别对学生的样本进行错误分析，指出发声存在的问题，总结出“口腔没打开”、“咬字位置不正确”等一系列错误原因。为了便于分析，下面只讨论发声为/a/的分析结果，并不影响其统计规律。

由于采集到的美声信号中存在无效的静音段和噪声段，会对功率谱分析和特征提取存在一定程度的干扰，增加运算量，因此需要对声音信号进行端点检测，确定其起点和终点，以便提高计算效率。笔者采用一种基于短时能量和谱质心特征进行端点检测的方法[15]，其方法步骤如下。

首先，对语音信号中的每一帧提取短时能量，设$x_i(n)$(n=1～N)为第i帧信号，长度为N，该帧的能量$E(i)$为

$$E(i)=\frac{1}{N}\sum_{n=1}^{N}\left|x_i(n)\right|^2 \tag{1}$$

其次，提取该帧的谱质心。设第i帧的谱质心C_i为

$$C_i=\frac{\sum_{k=1}^{N}(k+1)X_i(k)}{\sum_{k=1}^{N}X_i(k)} \tag{2}$$

其中：$X_i(k)$(k=1～N)为第i帧的离散傅里叶变换；N为帧长度。

最后，估计短时能量和谱质心特征序列的阈值，设M_1和M_2分别为2个局部最大值的位置，则阈值T为

$$T=\frac{WM_1+M_2}{W+1} \tag{3}$$

其中：W为笔者设置的参数，W越大，阈值就越靠近M_1。

经过上述阈值化处理，可以得到一段标记语音段的阈值化序列，将该序列代入原始信号中，就可获得语音段在原始信号中开始和结束的位置。

1.2 Burg法功率谱估计

将完成端点检测的信号进行Burg法功率谱分析。在对随机信号的分析中，可以利用自回归（autoregressive model，简称AR）模型进行功率谱估计。其中，Burg法无需对自相关函数进行估算，而是用已知序列$x(n)$求出反射系数，再利用Levinson递推算法，由反射系数来计算回归模型参数，以得到较好的谱估计结果。

利用Burg法估计AR模型参数，首先要确定式(4)所示的初始条件，其次根据序列$x(n)$求出式(5)所示的自相关函数σ_0^2

$$e_0(n)=b_0(n)=x(n) \tag{4}$$

$$\sigma_0^2=\frac{1}{N}\sum_{n=0}^{N-1}x^2(n) \tag{5}$$

令k=1，计算AR模型的反射系数K_k

$$K_k=-\frac{2\sum_{n=p}^{N-1}\left[e_{p-1}(n)b_{p-1}(n-1)\right]}{\sum_{n=p}^{N}\left[e_{p-1}^2(n)+b_{p-1}^2(n-1)\right]} \tag{6}$$

在Levinson关系式的$a_k(i)$(i=1～k－1)中，分别代入p阶AR模型反射系数和p－1阶AR模型反射系数，计算a_{ki}(i=1～k－1)、前向预测误差$e_k(n)$和后向预测误差$b_k(n)$，分别为

$$a_k(i)=a_{k-1}(i)+K_ka_{k-1}(k-i) \tag{7}$$

$$e_k(n)=e_{p-1}(n)+K_pb_{p-1}(n-1) \tag{8}$$

$$b_k(n)=b_{p-1}(n)+K_pe_{p-1}(n) \tag{9}$$

根据$\sigma_k^2=\left(1-K_p^2\right)\sigma_{k-1}^2$计算出$\sigma_k^2$，令$k=k+1$。重复上述步骤，直至预计的阶数为止，以求出所有阶的AR模型参数。

Burg估计算法的递推过程建立在已知序列的基础上，很好地避免了对于序列自相关函数的计算，与其他算法相比，有着较好的频率分辨率[16]。

1.3 局部二次回归平滑

笔者使用局部二次回归平滑对Burg法得到的功率谱进行平滑处理。局部二次回归平滑就是使用

二次多项式作为局部多项式的回归拟合，是一种用于局部回归分析的非参数方法。

在对信号进行二次回归平滑时，首先要确定拟合点的数量和位置，再以拟合点为中心，确定 k 个最邻近的点，通过权重函数计算这些点的权重。其中，对权重的计算要先确定区间内的点到拟合点的 x 轴的距离，找到区间内的最大值，然后对其他距离做归一化处理。归一化函数表达式为

$$w_i(x_0)=W\left(\frac{|x_0-x_i|}{\Delta x_0}\right) \tag{10}$$

使用三次指数函数对权重进行转化，三次函数表达式为

$$W(u)=(1-u^3)^3 \tag{11}$$

接下来对区间内的散点进行局部二次回归拟合，考虑到离拟合点的远近不同，点的取值对拟合线的影响也不同，故在定义损失函数时，应率先降低近的点与拟合线的误差，即对最小二乘法加上权重。加权最小二乘法的表达式为

$$J(a,b)=\frac{1}{N}\sum_{i=1}^{N}w_i(y_i-ax_i-b)^2 \tag{12}$$

对区间内的样本进行多项式拟合后，不断重复拟合过程，得到不同区间内的加权回归曲线，最后通过对回归曲线中心的连接，便可生成完整的平滑曲线。

1.4　BP 神经网络

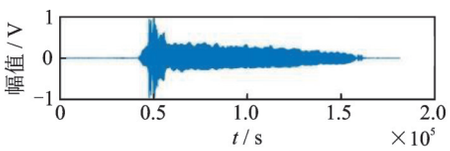
笔者选取 BP 神经网络用于美声特征的分类。BP 神经网络作为一种多层的前馈神经网络，由输入层、隐藏层和输出层组成。本研究对 BP 神经网络设置 2 个隐藏层：第 1 个隐藏层包含 10 个神经元，使用线性函数作为激活函数；第 2 个隐藏层包含 2 个神经元，使用对数 S 形转移函数作为激活函数。所选样本数据为平滑处理后的信号功率谱特征值，最后选择梯度下降自适应学习率的反向传播算法作为训练函数来训练 BP 神经网络。

2　实验数据采集与分析

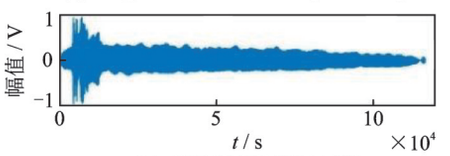
采集某音乐学院 5 名女高音新生和 3 名老师的美声发声信号共 400 条，利用 Matlab 软件对经过预处理的美声信号进行 Burg 功率谱估计，对比正确样本与错误样本之间功率谱形态走势的区别，对与正确功率谱图像差距较大的地方做函数图像的拟合，并提取谱图的特征参数，最后比较科学美声发声和错误美声发声之间功率谱曲线与参数的差距。

2.1　信号的 Burg 功率谱估计

声音信号端点检测时域波形如图 1 所示。首先对采集到的美声信号进行端点检测，原始信号的时域波形见图 1(a)，去除多余的静音段和噪声段，得到无干扰的声信号时域波形见图 1(b)。



(a) 信号原始时域图
(a) Original time domain diagram of signal



(b) 信号端点检测时域图
(b) Time domain diagram of signal endpoint detection

图 1　声音信号端点检测时域波形

Fig.1　Time domain waveform of sound signal endpoint detection

将预处理后的信号带入 25 阶 AR 模型，美声发声信号功率谱曲线如图 2 所示，得到正确美声信号功率谱和 3 种具有代表性的、不同错误类型的美声信号功率谱。根据图中功率谱整体的波动和走势情况，可将功率谱划分为 3 个能量区，如图 2 中竖线所示。其中：0～6 kHz 为第 1 能量区；6～11 kHz 为第 2 能量区；11～15 kHz 为第 3 能量区。

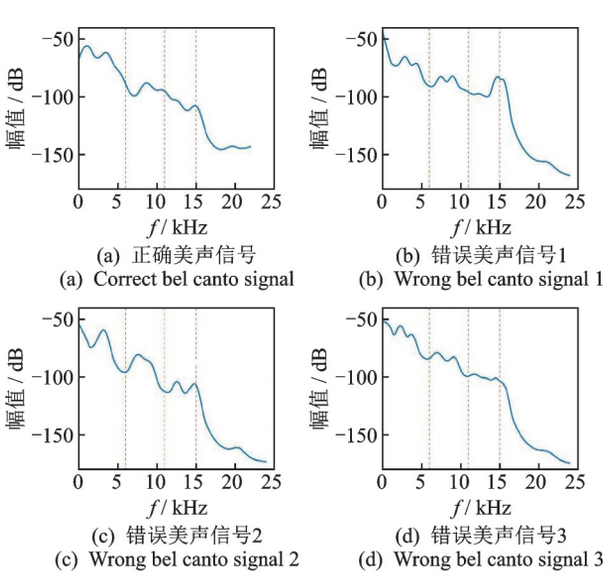


图 2　美声发声信号功率谱曲线

Fig.2　Power spectrum curve of bel canto signal

由功率谱曲线可以看出，高音信号的功率谱整体均呈下降趋势。由图2(a)的标准美声信号功率谱中可以发现，1，6和15 kHz处均存在明显的峰值变化，6 kHz处有明显的下降趋势，波谷平均深度为−100 dB，与波峰有着40 dB的落差。曲线从10 kHz开始平稳下降且无较大波动，在15 kHz处下降速度加快，曲线陡峭，至17 kHz处降至最低点−140 dB。

在错误美声信号的功率谱中，图2(b)所示的错误样本1存在着“口腔没有打开、气息没有用上”的错误，其功率谱在6 kHz处的波谷相对较浅，与左侧波峰的落差仅有20 dB，而在15 kHz处的曲线呈明显上升趋势的波动，持续约1 kHz后加速下降至最低点。由图2(c)所示的错误样本2可以看出，曲线在1，6和11 kHz处均有波谷产生，且波动幅度较大，曲线相对不稳定，存在“咬字位置不对”的错误，在15 kHz处变陡加速下降。由图2(d)所示的错误样本3可以看出，曲线整体无较大波动，几乎呈平稳态势下降，直至15 kHz处曲线变陡并下降至最低点，存在“口腔发声位置错误”的问题。

从能量区的分割上可以看出，错误样本曲线在每个能量区中均有不同幅度的波动；而正确样本曲线只有在进入第2能量区后有一处波谷，从第2能量区中部至第3能量区结束之间的图像下降匀速，无明显起伏特征。

2.2 信号的曲线拟合与箱式图

基于上述情况，笔者在功率谱曲线区别较大的区间内进行基于最小二乘法的一阶拟合和二阶拟合，得到一元二次曲线方程和一元一次直线方程，再对2种方程的系数取平均值和方差。其中，一元二次方程拟合了3～7 kHz之间功率谱中存在的波谷曲线，由于2种信号在其区间内的变化差距较大，得到的方程在系数上有着较大差别。功率谱曲线一元二次方程拟合系数如表1所示，正确发声信号曲线的一次项系数b大于错误信号，而二次项系数a和常数项c则小于错误信号。

在曲线方程中，二次项系数a代表函数抛物线的开口大小，a的绝对值越大，抛物线的开口越窄。对于2条抛物线$A_1x^2+B_1x+C_1y+D_1=0$和$A_2x^2+B_2x+C_2y+D_2=0$，其开度公式分别为

$$\sigma_1=|4C_1/A_1| \tag{13}$$

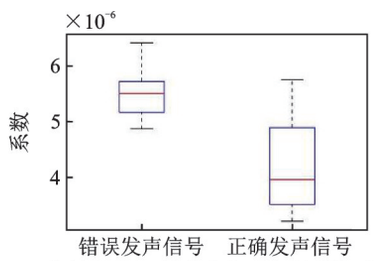
$$\sigma_2=|4C_2/A_2| \tag{14}$$

将正确信号和错误信号的多项式系数分别代入σ_1和σ_2，得到$\sigma_1>\sigma_2$，即正确信号抛物线的开口度要大于错误信号。

表1 功率谱曲线一元二次方程拟合系数

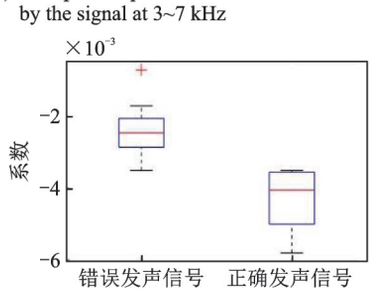
Tab.1 Fitting coefficient of the power spectrum curve by the quadratic equation

发声类型	多项式系数	平均值	方差
正确发声信号	二次项系数a	4.14×10^{-6}	6.51×10^{-13}
	一次项系数b	−0.056	1.10×10^{-4}
	常数项c	93.69	1 177.61
错误发声信号	二次项系数a	5.75×10^{-6}	1.80×10^{-13}
	一次项系数b	−0.067	3.90×10^{-5}
	常数项c	111.83	794.41



(a) 信号3~7 kHz拟合曲线二次项系数a箱式图

(a) Box-plot of quadratic coefficient a of fitting curve by the signal at 3~7 kHz



(b) 信号10~15 kHz拟合直线斜率k箱式图

(b) Box-plot of slope k of fitting straight line by the signal at 10~15 kHz

图3 多项式系数箱式图

Fig.3 Box-plot with polynomial coefficients

再对图中10～15 kHz的下降直线进行拟合，得到了斜截式的一次函数直线方程，功率谱曲线一元一次方程拟合系数如表2所示。可以发现，正确信号的斜率k要小于错误信号，而截距b大于错误信号，即正确信号的倾斜坡度较大，错误信号坡度较为平缓。

表2 功率谱曲线一元一次方程拟合系数

Tab.2 Fitting coefficient of power spectrum curve by linear equation

发声类型	多项式系数	平均值	方差
正确发声信号	斜率k	-4.31×10^{-3}	6.43×10^{-7}
	截距b	−50.627	95.817
错误发声信号	斜率k	-1.78×10^{-3}	8.16×10^{-7}
	截距b	−73.769	169.543

为了更直观地观察数据的离散分布情况，了解数据分布状态，将拟合出的多项式系数进行箱式图分析，如图 3所示。由图 3(a)所示的二次项系数 a 的箱式图可以看出：错误信号的系数整体低于正确信号，其箱式图长度较短，数据多集中分布在很小的范围内；正确信号的箱式图较长，表明数据间差异比较大，方差也大于错误信号。由图 3(b)所示的斜率 k 的箱式图可以看出：正确信号的数据波动较大，但在错误信号中存在一处离群值，导致方差比正确信号的方差大。

由箱式图可知，在二次项系数箱式图的 5×10^{-6} 处和斜率箱式图的 -3×10^{-3} 处均有明显的分界，可以把正确信号和错误信号按照分界数值直接区分开，故采用阈值法的识别率可达到 100%。

对美声信号的功率谱曲线做特征值统计，如表 3所示。由表可知，错误信号的方差明显大于正确信号，说明错误信号的功率谱数据波动较大，数据分布比较分散，在平均数附近波动较大，且存在较大的上下限差。

表 3　美声信号功率谱统计特征值

Tab.3　Statistical eigenvalues of power spectrum of bel canto signal

发声类别	平均值	标准差	方差	中位数	四分位差	最大值	最小值
正确发声信号	−105.13	28.5	814.96	−103.43	46.23	−32.06	−145.9
错误发声信号	−112.07	39.8	1 472.10	−102.37	75.87	−36.96	−178.7

2.3　基于 BP 网络的神经分类

对 400 条声音信号进行训练集和测试集的划分，其中 75% 的数据作为训练集导入 BP 神经网络中进行训练，使 BP 神经对两类发声信号的特征值有记忆能力；再将剩余的 15% 数据作为测试集，来测试 BP 神经网络的识别正确率。BP 神经网络收敛图如图 4所示，由图可以看出，训练在 120 次左右达到收敛，识别率为 95.23%。

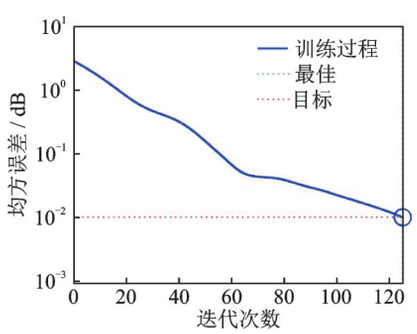


图 4　BP 神经网络收敛图

Fig.4　Convergence diagram of BP neural network

由 BP 神经网络的识别结果可知，相比于利用 BP 神经网络对美声进行分类，基于系数箱式图的阈值法可以更直接地将 2 种类别区分开，且识别率达 100%。因此，采用函数拟合的方法明显优于直接对功率谱特征值进行分类训练的方法。

3　结　论

1) 标准美声唱法的功率谱仅在 6 kHz 左右有一处明显的波谷，下降落差约为 40 dB，其余频率并无较大的波谷产生。在错误的美声唱法中，有些谱线没有明显的波谷，而有些谱线波谷较多，波动幅度较大。对 3～7 kHz 内的波谷曲线和 10～15 kHz 内的下降直线分别做一元二次函数拟合和一元一次函数拟合，可以得出正确信号在拟合的曲线上有着更大的开口度和更深的波谷，在直线上有着更大的倾斜度。在系数箱式图中使用阈值法，可以将 2 种类型的信号直接区分开。

2) 根据功率谱的波动和走势，可将其划分为 3 个能量区。在能量区中，错误样本的曲线波动频率更大，且在区域交界处有波谷；正确样本仅在第 1、第 2 能量区之间有波动，其余区域波动较不明显。

3) 使用美声声音信号功率谱进行 2 种声音信号的 BP 神经网络训练和分类识别，识别正确率可达 95.23%；而使用系数阈值法，可实现对 2 种发声信号的 100% 分类，表明本研究提出的美声发声信号特征阈值法更加有效。

4) 可以利用笔者目前的研究结果建立一套针对美声发声的打分系统，用于评估声乐初学者在发声训练时的标准程度。

参　考　文　献

[1] UM E, ZHENG Y. Misunderstanding analysis and countermeasure research in vocal music teaching of bel canto[J]. Advances in Social Science, Education and Humanities Research, 2018, 300: 773-777.

[2] 黄珣. 声乐教学中美声唱法与民族唱法的对比分析[J]. 艺术教育, 2020, 10: 58-61.
HUANG Xun. Comparative analysis of bel canto and national singing in vocal music teaching[J]. Art Education, 2020 ,10: 58-61. (in Chinese)

[3] JOLIVEAU E, SMITH J, WOLFE J. Vocal tract resonances in singing: the soprano voice[J]. Acoustical Society of America, 2004, 116(4): 2434-2439.

[4] MCHENRY M A, EVANS J, POWITZKY E. Effects of bel canto training on acoustic and aerodynamic characteristics of the singing voice [J]. Journal of Voice, 2016, 30(2): 198-204.

[5] CAFFIER P P, NASR A I, RENDON M, et al. Common vocal effects and partial glottal vibration in professional nonclassical singers[J]. Journal of Voice, 2018, 32(3): 340-346.

[6] MAYR A. Investigating the voce faringea: physiological and acoustic characteristics of the bel canto tenor′s forgotten singing practice[J]. Journal of Voice, 2017, 31(2): 13-23.

[7] SOUZA G V, DUARTE J, VIEGAS F, et al. An acoustic examination of pitch variation in soprano singing [J]. Journal of Voice, 2020, 34(4): 41-49.

[8] HASAN T, SHAKARA A. A signal processing approach to music tutor [J]. Journal of Computer Engineering, 2017, 19(6): 13-25.

[9] ZYSK A, BADURA P. An approach for vocal register recognition based on spectral analysis of singing [J]. International Journal of Cognitive and Language Sciences, 2017, 11(2): 207-212.

[10] BARLOW C, LOVETRI J. Closed quotient and spectral measures of female adolescent singers in different singing styles [J]. Journal of Voice, 2010, 24(3): 314-318.

[11] LU N. A tentative discussion on analysis methods of bel canto[C]//2018 International Conference on Culture, Literature, Arts & Humanities. London: Francis Academic Press, 2018: 323-326.

[12] LYU S L, ZHOU L X. The application of acoustic analysis in the study of yugur traditional folk songs[J]. Advances in Computer Science Research, 2017, 82: 61-64.

[13] DING S Y. The application of bel canto in national vocal music[J]. Advances in Social Science, Education and Humanities Research, 2017, 171: 232-236.

[14] 钱一凡, 孔江平. 民歌男高音共鸣的实验研究[C]//第七届中国语音学学术会议暨语音学前沿问题国际论坛. 北京: 中国中文信息学会, 2012: 268-274.

[15] SREEKUMAR K T, GEORGE K K, ARUNRAJ K, et al. Spectral matching based voice activity detector for improved speaker recognition [C]//2014 International Conference on Power Signals Control and Computations. Thrissur: IEEE, 2014: 1-4.

[16] 姚文俊. 自相关法和Burg法在AR模型功率谱估计中的仿真研究[J]. 计算机与数字工程, 2007, 35(10): 32-35.
YAO Wenjun. Research on AR model power spectrum estimation based on the algorithm and Burg algorithm [J]. Computer and Digital Engineering, 2007, 35(10): 32-35. (in Chinese)

第一作者简介: 张凯, 男, 1981年6月生, 博士、副教授、硕士生导师。主要研究方向为振动检测、可靠性及人工智能算法。曾发表《基于分类学习粒子群优化算法的液压矫正机控制》(《机械工程学报》2017年第53卷第18期)等论文。
E-mail:99267502@qq.com

《振动、测试与诊断》参加江苏优秀期刊展

由中国科协、国家新闻出版署主办的“第十八届中国科技期刊发展论坛”于2023年11月29—30日在南京召开。根据江苏省科技期刊学会推荐并经中国科协确定,《振动、测试与诊断》被邀参加“江苏期刊展”,与其他部分江苏优秀期刊亮相大会,共同展示江苏期刊的魅力!