



(12)发明专利

(10)授权公告号 CN 109727608 B

(45)授权公告日 2020.07.24

(21)申请号 201711010103.6

G10L 25/93(2013.01)

(22)申请日 2017.10.25

G10L 15/16(2006.01)

(65)同一申请的已公布的文献号

G10L 15/14(2006.01)

申请公布号 CN 109727608 A

G10L 15/04(2013.01)

G10L 15/06(2013.01)

(43)申请公布日 2019.05.07

G10L 15/08(2006.01)

A61B 5/00(2006.01)

(73)专利权人 香港中文大学深圳研究院

地址 518057 广东省深圳市南山区虚拟大

学园区粤兴二道10号

(56)对比文件

CN 103928023 A,2014.07.16,

(72)发明人 李丹 刘媛媛

审查员 薛彦杰

(74)专利代理机构 深圳新创友知识产权代理有

限公司 44223

代理人 江耀纯

(51)Int.Cl.

G10L 25/51(2013.01)

G10L 25/66(2013.01)

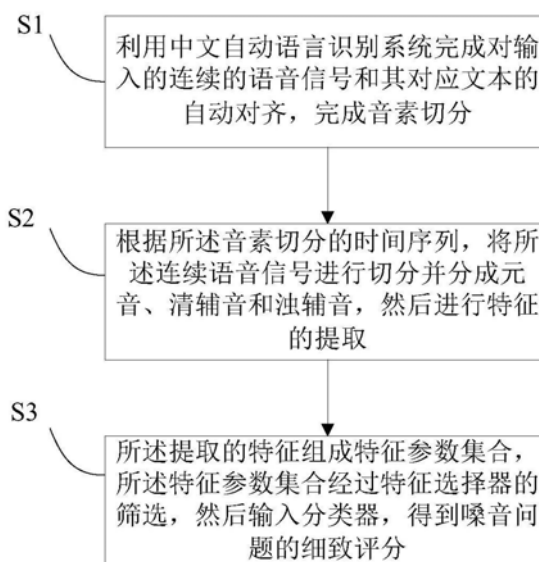
权利要求书2页 说明书10页 附图5页

(54)发明名称

一种基于中文语音的病态嗓音评估系统

(57)摘要

本发明提供一种基于中文语音的病态嗓音评估系统,包括如下步骤:利用中文自动语言识别系统完成对输入的连续的语音信号和其对应文本的自动对齐,完成音素切分;根据所述音素切分的时间序列,将所述连续语音信号进行切分并分成元音、清辅音和浊辅音,然后进行特征的提取;所述提取的特征组成特征参数集合,所述特征参数集合经过特征选择器的筛选,然后输入分类器,得到嗓音问题的细致评分。该方法使用大规模正常语音数据库训练自动语音识别系统,按照不同音素的发声特点对切分的语音序列进行分类,分别针对不同音素类型设计特征参数,最终得到针对嗓音问题的客观整体评分及各细致评分,为临床诊断和康复治疗提供重要参考。



1. 一种基于中文语音的病态嗓音评估系统,其特征在于,包括语音输入模块、语音评估模块、评估结果输出模块,

所述语音输入模块用于接受实时录音或音频,并将所述实时录音或音频传输给语音评估模块;

所述语音评估模块用于对所述实时录音或音频进行评估,并将评估结果传输给评估结果输出模块;对所述实时录音或音频进行评估包括采用基于中文语音的病态嗓音评估方法,所述方法包括:S1:对所述实时录音或音频的连续的语音信号和其对对应文本的自动对齐,完成音素切分;S2:根据所述音素切分的时间序列,将所述连续的语音信号进行切分并分成元音、清辅音和浊辅音,然后对所述元音提取第一类特征、第二类特征和第四类特征的参数;对所述浊辅音提取第二类特征、第三类特征、第四类特征的参数;对所述清辅音提取第三类特征、第四类特征的参数;所述第一类特征包括:基频抖动,振幅抖动,信号噪声比,音高,能量强度;所述第二类特征包括:语音中断,无声信号比值;所述第三类特征包括:梅尔倒谱系数;所述第四类特征包括:高低频能量比,基频变化量,信号噪声比变化量,连续能量差值,连续能量差值变化量,归一化能量;S3:对所述元音提取所述第一类特征、所述第二类特征和所述第四类特征的参数、对所述浊辅音提取所述第二类特征、所述第三类特征、所述第四类特征的参数、对所述清辅音提取所述第三类特征、所述第四类特征的参数组成特征参数集合,所述特征参数集合经过特征选择器的筛选,然后输入分类器,得到嗓音问题的细致评分,所述细致评分作为所述评估结果;

所述结果输出模块用于将所述评估结果输出。

2. 如权利要求1所述的基于中文语音的病态嗓音评估系统,其特征在于,所述基于中文语音的病态嗓音评估方法是使用正常语音数据库训练基于深度神经网络-隐马尔可夫模型。

3. 如权利要求1所述的基于中文语音的病态嗓音评估系统,其特征在于,所述基于中文语音的病态嗓音评估方法采用了说话人自适应技术。

4. 如权利要求1所述的基于中文语音的病态嗓音评估系统,其特征在于,所述基于中文语音的病态嗓音评估方法使用Viterbi算法将所述连续的语音信号和其对对应文本的音素序列强制对齐以完成音素切分工作。

5. 如权利要求1所述的基于中文语音的病态嗓音评估系统,其特征在于,所述第四类特征中:

所述高低频能量比HiLo的计算公式为:
$$Hilo = 20 * \log_{10} \left(\frac{E_{hi} - E_0}{E_{lo} - E_0} \right)$$
; 其中, E_{hi} 为6000HZ以上频段的能量, E_{lo} 为6000HZ以下频段的能量, E_0 为背景噪声能量;

所述基频变化量var_F0的计算公式为:
$$\text{var_F0} = \text{var} \left(\frac{F0(1:n)}{\text{median}(F0(1:n))} \right)$$
; 其中, F0为一帧信号的基频, n为一段连续语音的总帧数目, var是调用MATLAB中的var函数;

所述信号噪声比变化量var_HNR的计算公式为:
$$\text{var_HNR} = \text{var} \left(\frac{HNR(1:n)}{\text{median}(HNR(1:n))} \right)$$
; 其中, HNR为信噪比, n为一段连续语音的总帧数目;

所述连续能量差值D3_Energy的计算公式为：
$$D3_Energy = \frac{\sum_{i=1}^3 E(t+i) - \sum_{i=1}^3 E(t-i)}{3 * \left(\max_{j=t-3:t+3} E(j) - \min_{j=t-3:t+3} E(j) \right)}$$
；其

中，E(t)为第t帧的时域能量；

所述连续能量差值变化量var_D3的计算公式为：
$$\text{var_D3} = \text{var} \left(\frac{D3_Energy(1:n)}{\text{median}(D3_Energy(1:n))} \right) ;$$

其中，n为一段连续语音的总帧数；

所述归一化能量E_nor的计算公式为：
$$E_nor = 20 * \log_{10} \left(\frac{E_{\text{voice}}}{E_{\text{noise}}} \right)$$
；其中，E_voice为语音信号的能量，E_noise为静音部分的底噪能量。

6.如权利要求1所述的基于中文语音的病态嗓音评估系统，其特征在于，所述细致评分针对待评估的嗓音问题，所述待评估的嗓音问题包括如下嗓音问题：

Q1:总体感觉上一段嗓音的病变的严重程度；

Q2:感觉到嗓音异常、不规则、不清晰；

Q3:感觉到嗓音里有呼气或漏气现象；

Q4:感觉到发音过度用力；

Q5:发音过程中，存在突然不受控制的停顿；

Q6:感觉到突然不受控制的基频或声调变化；

Q7:发音过程中存在快速连续的爆裂声；

Q8:感觉到有规律有周期的声调或响度的变化；

Q9:嗓音中存在明显的基频或声调与说话人性别、年龄和文化教育背景不符合的异常；

Q10:发音音量明显与说话内容不符合的过大或者过小。

7.如权利要求1所述的基于中文语音的病态嗓音评估系统，其特征在于，所述细致评分得出针对嗓音问题的严重程度，所述严重程度包括：正常或轻微、中度、严重。

一种基于中文语音的病态嗓音评估系统

技术领域

[0001] 本发明涉及嗓音检测技术领域,具体涉及一种基于中文语音的病态嗓音评估方法。

背景技术

[0002] 嗓音是人类语言沟通的载体。发声系统在声带振动的驱使下发出声音,并通过咽喉、口腔组成的通道传播出来。声音承载不同的信息,是人与人之间沟通必不可少的手段。发声系统发出的声音可以通过波形信号来描述,称为嗓音信号。当发声器官处于正常状态时,声带的振动带有明显的周期规律,咽喉口腔组成的传播通道也有规律的变化,因而产生的嗓音信号也很有规律可循。

[0003] 实际生活中,嗓音不只被用作相互交流信息,而且还用于歌唱,表演,表达情感等等目的。嗓音疾病会对我们的日常交流造成困难,降低沟通效率。嗓音障碍的主要原因有过度用嗓,发音不当,或者发声器官感染疾病。嗓音障碍在嗓音信号上有着明显的体现,例如嗓音的声调、音量、共振峰出现不规则特性,或者嗓音的音质、音色等特征出现明显与说话人的年龄、性别、母语以及文化背景不同的异常变化。

[0004] 在对嗓音障碍的临床诊断和治疗过程中,对嗓音信号进行人工感知评价是一个标准而且重要的无创方法。尽管人工感知评价的评价标准和操作步骤已经发展多年,但是由于缺少足够的有经验的言语治疗师,人工感知评价无法满足现代社会中的需求。基于数字信号处理技术的自动嗓音评估,对于嗓音障碍的临床诊断与治疗有着重要的实际应用价值。其优势包括非侵入性,可重复性,客观性,可靠性等。传统的嗓音分析方法一般只是分析以固定发音位置产生的长元音,这使声学信号分析变得简单,但与连续语言相比,长元音存在以下限制:

[0005] (1) 研究表明,嗓音信号质量受到语言因素的影响,如辅音-元音转换,而这些语言因素在长元音中不能体现,故发长元音时的嗓音信号不能代表复杂的自然语言交流时的嗓音信号;

[0006] (2) 长元音发声有诸多限制,如持续保持音量、音高和发声器官位置,这使长元音音频的获得不如自然连续语音简便灵活;

[0007] (3) 另外,也有研究表明,言语治疗师使用长元音进行嗓音感知评价的可靠性不如使用连续语音。

[0008] 目前对病态嗓音的客观评价研究多集中在对嗓音信号质量进行整体的评价,方法有回归分析或分类;但实际上病态嗓音的临床表现多样,有声音嘶哑、发声费力、咽干、失声、颤音等,仅仅评价嗓音问题的整体严重程度不能为临床辨别具体的嗓音问题提供帮助。目前很少有相关研究工作能够细致到量化嗓音的个别特定方面,如声嘶、气息音等。我们认为这些细致的特性研究对于临床治疗是有关键作用的。

发明内容

[0009] 本发明为了解决现有技术中没有一种方法可以细致评价病态嗓音的问题,提供一种基于中文语音的病态嗓音评估系统。

[0010] 为了解决上述问题,本发明采用的技术方案如下所述:

[0011] 一种基于中文语音的病态嗓音评估方法,包括如下步骤:利用中文自动语言识别系统完成对输入的连续的语音信号和其对对应文本的自动对齐,完成音素切分;根据所述音素切分的时间序列,将所述连续语音信号进行切分并分成元音、清辅音和浊辅音,然后进行特征的提取;所述提取的特征组成特征参数集合,所述特征参数集合经过特征选择器的筛选,然后输入分类器,得到嗓音问题的细致评分。

[0012] 优选地,所述中文自动语言识别系统是使用正常语音数据库训练基于深度神经网络-隐马尔可夫模型的中文语音识别系统。

[0013] 优选地,所述中文自动语言识别系统采用了说话人自适应技术。

[0014] 优选地,所述中文自动语言识别系统使用Viterbi算法将连续的语音信号和其对对应文本的音素序列强制对齐以完成音素切分工作。

[0015] 优选地,所述步骤S2中的所述特征包括:第一类特征:基频抖动,振幅抖动,信号噪声比,音高,能量强度;第二类特征:语音中断,无声信号比值;第三类特征:梅尔倒谱系数;第四类特征:高低频能量比,基频变化量,信号噪声比变化量,连续能量差值,连续能量差值变化量,归一化能量。

[0016] 优选地,对所述元音提取所述第一类特征、第二类特征和第四类特征的特征参数;对所述浊辅音提取第二类特征、第三类特征、第四类特征的特征参数;对所述清辅音提取第三类特征、第四类特征的特征参数。

[0017] 优选地,所述第四类特征中:所述高低频能量比Hilo的计算公式为:

$$Hilo = 20 \cdot \log_{10} \left(\frac{E_{hi} - E_0}{E_{lo} - E_0} \right);$$
 其中, E_{hi} 为1200HZ以上频段的能量, E_{lo} 为1200HZ以下频段的能量, E_0

为背景噪声能量;所述基频变化量var_F0的计算公式为: $var_F0 = var \left(\frac{F0(1:n)}{\text{median}(F0(1:n))} \right);$ 其

中, $F0$ 为一帧信号的基频, n 为一段连续语音的总帧数目, var 是调用MATLAB中的var函数;所述信号噪声比变化量var_HNR的计算公式为: $var_HNR = var \left(\frac{HNR(1:n)}{\text{median}(HNR(1:n))} \right);$ 其中,

HNR为信噪比, n 为一段连续语音的总帧数目;所述连续能量差值D3_Energy的计算公式为:

$$D3_Energy = \frac{\sum_{i=1}^3 E(t+i) - \sum_{i=1}^3 E(t-i)}{3 \cdot \left(\max_{j=t-3:t+3} E(j) - \min_{j=t-3:t+3} E(j) \right)};$$
 其中, $E(t)$ 为第 t 帧的时域能量;

[0018] 所述连续能量差值变化量var_D3的计算公式为: $var_D3 = var \left(\frac{D3_Energy(1:n)}{\text{median}(D3_Energy(1:n))} \right);$

其中, n 为一段连续语音的总帧数;所述归一化能量 E_{nor} 的计算公式为: $E_{nor} = 20 \cdot \log_{10} \left(\frac{E_{voice}}{E_{noise}} \right);$

其中, E_{voice} 为语音信号的能量, E_{noise} 为静音部分的底噪能量。

[0019] 优选地,所述步骤S3中所述细致评分针对待评估的嗓音问题,所述待评估的嗓音

问题包括如下嗓音问题:Q1:总体感觉上一段嗓音的病变的严重程度;Q2:感觉到嗓音异常、不规则、不清晰;Q3:感觉到嗓音里有呼气或漏气现象;Q4:感觉到发音过度用力;Q5:发音过程中,存在突然不受控制的停顿;Q6:感觉到突然不受控制的基频或声调变化;Q7:发音过程中存在快速连续的爆裂声;Q8:感觉到有规律有周期的声调或响度的变化;Q9:嗓音中存在明显的基频或声调与说话人性别、年龄和文化教育背景不符合的异常;Q10:发音音量明显与说话内容不符合的过大或者过小。

[0020] 优选地,所述细致评分得出针对嗓音问题的严重程度,所述严重程度包括:正常或轻微、中度、严重。

[0021] 本发明还提供一种基于中文语音的病态嗓音评估系统,包括语音输入模块、采用如上任一所述方法的语音评估模块、评估结果输出模块,所述语音输入模块用于接受实时录音或音频,并将所述实时录音或音频传输给语音评估模块;所述语音评估模块用于对所述实时录音或音频进行评估,并将评估结果传输给评估结果输出模块;所述结果输出模块用于将所述评估结果输出。

[0022] 本发明的有益效果为:提供一种基于中文语音的病态嗓音评估系统,该方法使用大规模正常语音数据库训练,按照不同音素的发声特点对切分的语音序列进行分类,针对不同的语音序列设计不同的特征参数,最终得到嗓音问题的细致评分,嗓音问题的客观评分,为临床诊断和康复治疗提供重要参考。

附图说明

[0023] 图1是本发明实施例1中基于中文语音的病态嗓音评估方法的步骤图。

[0024] 图2是本发明实施例1中基于中文语音的病态嗓音评估方法的演示图。

[0025] 图3是本发明实施例1中基于深度神经网络-隐马尔可夫模型的中文语音识别系统。

[0026] 图4是本发明实施例1中评分分类器方法的示意图。

[0027] 图5是本发明实施例1中没有声嘶语音中元音的HiLo值分布直方图。

[0028] 图6是本发明实施例1中有严重声嘶语音中元音的HiLo值分布直方图。

[0029] 图7是本发明实施例1中没有气息音语音中元音的HiLo值分布直方图。

[0030] 图8是本发明实施例1中有严重气息音语音中元音的HiLo值分布直方图。

[0031] 图9是本发明实施例1中音量正常语音中元音的E_nor值分布直方图。

[0032] 图10本发明实施例1中有中度、严重低音量语音中元音的E_nor值分布直方图。

[0033] 图11是本发明实施例3中基于中文语音的病态嗓音评估系统的示意图。

具体实施方式

[0034] 下面结合附图通过具体实施例对本发明进行详细的介绍,以使更好的理解本发明,但下述实施例并不限制本发明范围。另外,需要说明的是,下述实施例中所提供的图示仅以示意方式说明本发明的基本构思,附图中仅显示与本发明中有关的组件而非按照实际实施时的组件数目、形状及尺寸绘制,其实际实施时各组件的形状、数量及比例可为一种随意的改变,且其组件布局形态也可能更为复杂。

[0035] 实施例1

[0036] 在嗓音信号处理领域和言语治疗领域,病态嗓音的客观分析已经引起多方面的关注。例如,MDVP是一个言语治疗专家所常用的嗓音信号分析软件系统。MDVP提供了22个声学信号参数来描述嗓音质量。这些参数被言语治疗师用来作为依据进行嗓音评估。LingWAVES是另一个商用软件系统,用来辅助医生对嗓音问题作出临床诊断。LingWAVES可以进行基本的声学信号分析,例如频谱特征分析,声调分析,音量分析等等。近年不少研究学者采用一个免费软件Praat来进行声学信号分析,但是Praat并不是一个专门针对病变嗓音分析的工具,也无法对病变种类与严重程度给出判断。

[0037] 从信号处理的角度,相关的研究主要关注如何提取有效的特征参数来描述病变嗓音信号的不规则特性。最常用的特征参数包括基频抖动(jitter),振幅抖动(shimmer)和信号噪声比(HNR)。基频抖动,振幅抖动以及他们相关的类似特征参数,在时间域上描述了嗓音信号在声调和音量的变动程度。通常来说,病态嗓音由于不够稳定,基频抖动和振幅抖动数值会比较大。信号噪声比计算嗓音信号周期成分与加性噪声成分之间的能量比。病态嗓音的信号噪声比通常都比较低。其他相关的声学特征参数包括:pitch-period entropy (PPE),glottis quotient (GQ),voice turbulence index (VTI),和soft phonation index (SPI)等。另一方面,梅尔倒谱系数(Mel-Frequency Cepstral Coefficients,MFCC),是语音识别和说话人识别中常用的特征。该特征也被尝试用于病态嗓音的检测与分类。人工神经网络(Artificial neural network,ANN)和混合高斯模型(Gaussian mixture model,GMM)被用来作为嗓音分类工具。如前所述,分析的语音信号主要集中在长元音,近几年因为机器学习的发展,在病理语音信号研究方面也开始转向连续语音信号。这些研究虽然使用连续语言信号,但在声学信号分析中并没有细致区分不同语言内容的声音信号。对于一段连续语音输入,通常首先按时间帧提取上述特征参数,然后对所有帧级特征向量做统计分析,比如取某特征在所有时间帧上的平均值或标准方差或中位数,所得统计值被视为整段连续语音信号的特征,并作为后续的回归或分类问题的输入。

[0038] 总结现有研究工作,我们认为存在以下不足:

[0039] 上述常用特征参数如jitter,shimmer,主要集中描述相对稳定的发音状态,这些特征参数可以用来鉴别正常嗓音和病态嗓音,但是无法进一步细致检测出病变种类和病变程度。类似Jitter,shimmer参数更适合从稳定长元音中提取,自然连续语音在不同的音素之间快速转换,其声学特征也在不断变化,如描述声带振动的基频和声道变化的倒谱系数。连续语音中的这些变化因素不仅使一些传统的特征参数(如jitter,shimmer)失效,且连续语音所有帧级特征的统计结果也无法消除语言内容变化的影响。

[0040] 这些工作在检测和分类时大多采用统计数据模型,如混合高斯模型。当训练数据量足够且训练数据与实际测试数据匹配时,统计方法能够有效准确的描述不同类别嗓音的特征分布。但是病态嗓音的原因与种类很多,声学特征分布并不规则,且很多类别的训练数据并不足够,因此这些统计方法无法取得应有效果。

[0041] 目前很少有相关研究工作能够细致到量化嗓音的个别特定方面,如声嘶、气息音等。我们认为这些细致的特性研究对于临床治疗是有关键作用的。

[0042] 针对现有技术的不足,本辅助嗓音评估工具根据语言内容切分语音信号,针对不同类语音信号设计或寻找适合的特征参数。除了对嗓音问题进行整体评价,本设计还针对不同嗓音问题训练不同的评价模型,如声嘶(Roughness),咽干(Glottal fry),发声费力

(Strain)等,以实现对自然连续语音信号进行整体和细致的多维嗓音问题评分,为嗓音问题的诊断和治疗过程提供重要参考。

[0043] 如图1所示,一种基于中文语音的病态嗓音评估方法,包括步骤1-3,具体如下所示:

[0044] 1.利用中文自动语言识别系统完成对输入的连续的语音信号和其对应文本的自动对齐,完成音素切分;

[0045] 如图2所示,将利用中文自动语言识别系统完成对输入连续语音和其对应文本的自动对齐,从而实现对连续语音进行音素切分的目的。虽然使用中需要对病态嗓音语音进行识别,但缺乏大规模病态嗓音连续语音数据库。

[0046] 如图3所示,本方法使用大规模正常语音数据库训练基于深度神经网络-隐马尔可夫模型(DNN-HMM)的中文语音识别系统。在图3中,1为连续语音信号的声学特征向量 0_t ,2为深度神经网络DNN,1作为2的输入,3为深度神经网络输出的后验概率向量 $(P(a_j|0_t))$, a_j 表示一个音素的HMM状态,4是HMM链。为提高中文语音识别系统对病态嗓音语音的识别正确率,在训练声学模型时采用了说话人自适应技术(Speaker Adaptation Training, SAT)。

[0047] 在临床嗓音评估诊断中,病患会按言语治疗师指导做不同类型的发声,通常包含长元音、阅读短文和自由交谈。因为阅读的短文内容已知,对这部分的连续语音,其文本也可作为语音识别系统的输入。语音识别系统内部Viterbi算法将文本中的音素序列和语音信号强制对齐以完成音素切分工作。

[0048] 2.根据所述音素切分的时间序列,将所述连续语音信号进行切分并分成元音、清辅音和浊辅音,然后进行特征的提取;

[0049] 分别提取4个子类的特征,第一类特征:基频抖动(jitter),振幅抖动(shimmer),信号噪声比(HNR),音高(pitch),能量强度(intensity);第二类特征:语音中断(voice break),无声信号比值(unvoiced fraction);第三类特征:梅尔倒谱系数(MFCC);第四类:自设计参数包括:高低频能量比:HiLo;基频变化量:var_F0;信号噪声比变化量:var_HNR;连续能量差值:D3_Energy;D3_Energy变化量:var_D3;归一化能量。根据音素切分时间序列,将连续语音信号进行切分并分成元音、清辅音和浊辅音,针对这几类音素语音信号提取不同的特征参数。对元音提取上述第一、二和四类特征参数;对浊辅音提取第二、三、四类特征参数;对清辅音提取第三、四类特征参数。本方法提出的第四类特征参数介绍如下:

[0050] (1)HiLo,高低频能量比,公式如下,其中 E_{hi} 为高频段(6000HZ以上)能量, E_{lo} 为低频段(6000HZ以下)能量, E_0 为背景噪声能量。实验数据显示,对于声音嘶哑(Roughness)和有气息音(Breathiness)的语音信号。

$$[0051] \quad \text{Hilo} = 20 * \log_{10} \left(\frac{E_{hi} - E_0}{E_{lo} - E_0} \right)$$

[0052] (2) var_F0,基频变化量,公式如下,其中F0为一帧信号的基频,n为一段连续语音的总帧数目,var是调用MATLAB中的var函数。

$$[0053] \quad \text{var_F0} = \text{var} \left(\frac{F0(1:n)}{\text{median}(F0(1:n))} \right)$$

[0054] (3) var_HNR,信噪比变化量,公式如下,其中HNR为信噪比,n为一段连续语音的总帧数目。对于一些病理嗓音语音信号,不仅信噪比(HNR)更低,且HNR的变化更大,可以体现

在var_HNR上。

$$[0055] \quad \text{var_HNR} = \text{var}\left(\frac{\text{HNR}(1:n)}{\text{median}(\text{HNR}(1:n))}\right)$$

[0056] (4) D3_Energy, 连续能量差值, 公式如下, 其中E(t)为第t帧的时域能量。D3_Energy描述的是t时刻前后3帧语音信号能量的变化。

$$[0057] \quad \text{D3_Energy} = \frac{\sum_{i=1}^3 E(t+i) - \sum_{i=1}^3 E(t-i)}{3 * \left(\max_{j=t-3:t+3} E(j) - \min_{j=t-3:t+3} E(j) \right)}$$

[0058] (5) var_D3, D3_Energy变化量, 公式如下, 其中n为一段连续语音的总帧数。

$$[0059] \quad \text{var_D3} = \text{var} \left(\frac{\text{D3_Energy}(1:n)}{\text{median}(\text{D3_Energy}(1:n))} \right)$$

[0060] (6) E_nor, 归一化能量, 公式如下, 其中E_{voice}为语音信号的能量, E_{noise}为静音部分的底噪能量。E_nor可用于衡量语音信号能量相对于环境噪声的大小, 利于判断说话音量偏大或偏小(Loudness High/Low)。

$$[0061] \quad E_nor = 20 * \log_{10} \left(\frac{E_{\text{voice}}}{E_{\text{noise}}} \right)$$

[0062] 3. 所述提取的特征组成特征参数集合, 所述特征参数集合经过特征选择器的筛选, 然后输入分类器, 得到嗓音问题的细致评分。

[0063] 如图2所示, 特征筛选器共有10个, 分别针对待评估的10类嗓音问题, 见附件表1。

[0064] 表1嗓音问题列表

名称	具体含义
Overall Severity	总体感觉上一段嗓音的病变的严重程度
Roughness	感觉到嗓音异常、不规则、不清晰
Breathiness	感觉到嗓音里有呼气或漏气现象
Strain	感觉到发音过度用力
Phonation breaks	发音过程中, 存在突然不受控制的停顿
Pitch breaks	感觉到突然不受控制的基频或声调变化
Glottal fry	发音过程中存在快速连续的爆裂声
Tremor	感觉到有规律有周期的声调或响度的变化
Pitch: High/Low	嗓音中存在明显的基频或声调与说话人性别、年龄和文化教育背景不符合的异常
Loudness: High/Low	发音音量明显与说话内容不符合的过大或者过小

[0066] 对应的分类器共有10个, 分别针对待评估的10类嗓音问题训练。每一对特征选择

器和分类器构成一组评分模型,针对10类嗓音问题中的特定一种训练得到。特征提取模块输出的大规模特征参数集合先经过特征选择器,经过筛选后的参数将输入分类器,最终得到相应嗓音问题的评分。在特征选择上,除了综合常用算法,如LASSO(least absolute shrinkage and selection operator),mRMR(minimum redundancy maximum relevance)和LLBFS(local learning-based feature selection),还会特别参考设计者对各类嗓音问题进行信号分析的经验。实验表明,特定的嗓音问题在特定的声学特征参数上有表现,如声音嘶哑会引起语音信号HiLo参数增大,颤音的var_F0和var_HNR较大。参考临床言语治疗师对嗓音问题的评价,本设计使用4类分类器,其输出可表示嗓音正常或轻微、中度、严重嗓音问题。由于是多类分类问题,本方法选择多层神经网络(MLP,Multi-layer Perceptron)作为分类器。训练分类器所用嗓音数据含专业言语治疗师针对各项嗓音问题的主观评分,此主观评分作为训练MLP的参考标准。

[0067] 如图4所示,以声嘶(Roughness)为例说明每个评分分类器原理,对每个输入特征参数向量,MLP输出一个4维的one-hot向量,表示嗓音声嘶的严重程度,见表2。其他9种嗓音问题的评估也同理。

[0068] 表2嗓音问题的严重程度

[0069]	0001	正常(不存在声嘶问题)
	0010	轻微(存在轻微声嘶)
	0100	中度(存在中度声嘶)
	1000	严重(存在严重声嘶)

[0070] 下面举例说明自设计参数在实验中的效果:

[0071] (1) 关于HiLo值

[0072] 如图5所示,没有声嘶语音中元音的HiLo值分布直方图,图中横坐标为HiLo值,纵坐标为样本数量。

[0073] 如图6所示,有严重声嘶语音中元音的HiLo值分布直方图,图中横坐标为HiLo值,纵坐标为样本数量。

[0074] HiLo值偏高,如图5对比图6为没有声嘶对比有严重声嘶语音的HiLo值分布直方图,以及图7对比图8为没有气息音和有严重气息音语音HiLo值分布直方图。

[0075] 如图7所示,没有气息音语音中元音的HiLo值分布直方图,图中横坐标为HiLo值,纵坐标为样本数量。

[0076] 如图8所示,有严重气息音语音中元音的HiLo值分布直方图,横坐标为HiLo值,纵坐标为样本数量。

[0077] (2) 关于var_F0值:实验数据显示,有严重嗓音问题语音的var_F0值偏高,如表3对比表4为没有颤音和有严重颤音语音的var_F0分布。

[0078] 表3无颤音语音中元音的var_F0值

[0079]	语音序号	1	2	3	4	5	6	7	8	9
	var_F0(*10 ⁻³)	0.91	31.11	22.32	33.34	0.16	6.93	38.89	34.83	12.34

[0080] 表4有严重颤音语音中元音的var_F0值

[0081]	语音序号	1	2	3	4	5	6	7	8	9
	var_F0(*10 ⁻³)	91.47	150.47	241.33	49.23	20.37	39.53	69.50	119.30	63.11

[0082] (3) 关于var_HNR:即一些病理嗓音的语音信号其var_HNR值更大,如表5对比表6为无颤音和有严重颤音语音信号var_HNR值分布所示。

[0083] 表5无颤音语音中元音的var_HNR值

语音序号	1	2	3	4	5	6	7	8	9
var_HNR (*10 ⁻³)	4.07	6.07	7.47	3.67	2.00	4.38	8.16	5.65	6.73

[0085] 表6有严重颤音语音中元音的var_HNR值

语音序号	1	2	3	4	5	6	7	8	9
var_HNR (*10 ⁻³)	198.30	215.85	718.98	140.82	90.25	116.21	134.84	149.75	215.21

[0087] (4) 关于var_D3:一般病理嗓音会表现出较大的能量变化。D3_Energy也用来计算var_D3。var_D3用于衡量一段语音信号的能量稳定性,通常病态嗓音的var_D3值偏高,如表7对比表8为无颤音和有严重颤音语音信号var_D3值分布所示。

[0088] 表7无颤音语音中元音的var_D3值

语音序号	1	2	3	4	5	6	7	8	9
var_D3 (*10 ⁻⁵)	0.06	0.004	0.003	0.004	0.006	0.003	0.005	0.007	0.003

[0090] 表8有严重颤音语音中元音的var_D3值

语音序号	1	2	3	4	5	6	7	8	9
var_D3 (*10 ⁻⁵)	15	723	2218	5730	1217	825	62	840	306

[0092] (5) 关于E_nor:如图9对比图10为音量正常及有低音量语音的E_nor值分布,可见低音量语音的E_nor值偏低。

[0093] 如图9,音量正常语音中元音的E_nor值分布直方图,其中横坐标为E_nor值,纵坐标为样本数量。

[0094] 如图10,有中度、严重低音量语音中元音的E_nor值分布直方图,其中横坐标为E_nor值,纵坐标为样本数量。

[0095] 实施例2

[0096] 采用本发明所述的方法进行病态嗓音评估的具体实施方式如下所述:

[0097] (1) 建立病态嗓音数据库:此数据库由医院言语治疗师、学者和自愿者病患共同完成。病患规模为230个母语为中文人士,性别、年龄分别平衡。在相对固定的环境下,每个病患录制包括长元音、朗读短文和问题回答的语音信号。录制音频为双声道、16bit、采样率为44.1kHz的wav文件。所收集语音信号在做简单的前期处理(使用speaker diarisation技术删除言语治疗师说话内容,固定倍数放大和降噪)后,交给总共48位专业的言语治疗师做主观评分。对每个病患的录音做10个嗓音问题的主观评分。为提高主观评分的可靠性,最终对所有言语治疗师的主观评分做统计分析,去除不可信高分或低分,求中位值、方差等,最终得到每个病患其嗓音在10个方面的主观评分。

[0098] (2) 中文语音系统训练:使用正常语音的大规模词汇语音自动识别数据库训练一个中文语音自动识别系统。

[0099] (3) 将病态嗓音数据库中每个病患的短文阅读的语音和对应文本输入到中文语音自动识别系统,得到音素时间序列对齐数据。

[0100] (4) 根据步骤(3)中的音素时间序列对齐数据,再对短文阅读的语音信号中的元音、清辅音和浊辅音分别提取特征参数。

[0101] (5) 步骤(4)中提取的所有参数分别经过10个自动评分模块(10个特征参数选择器+10个四类分类器),这10个自动评分模块分别针对10个嗓音问题。在评分模型训练阶段,步骤(1)中得到的主观评分作为模型训练的目标值,用最小均方误差(Least Mean Square)算法使评分模型最优。

[0102] (6) 当自动评分模型最优后,即将此方法植入计算机系统,用于临床测试或学术研究。测试系统,除了模型计算所需的计算单元(NVIDIA GPU),还需要前端录音设备(可用Shure全向麦克风+TASCAM 2*2声卡),音频存储单元(计算机内存),评分结果显示单元(液晶显示屏)。测试开始时,首先请受试者(嘴巴)距离麦克风一定距离(30cm)发长元音(/a/, /i/, /u/),朗读短文,回答两个问题。录制的受试者语音信号先经过前期处理,然后短文朗读部分语音信号将进入计算系统进行音素切分、特征提取、自动评分,最后将在显示单元输出全部10个语音问题的评分结果。

[0103] 实施例3

[0104] 如图5所示,本发明还提供一种基于中文语音的病态嗓音评估系统,包括语音输入模块、采用如实施例1所述方法的语音评估模块、评估结果输出模块。

[0105] 所述语音输入模块用于接受实时录音或音频,并将所述实时录音或音频传输给语音评估模块;所述语音评估模块用于对所述实时录音或音频进行评估,并将评估结果传输给评估结果输出模块;所述结果输出模块用于将所述评估结果输出。

[0106] 本发明所提供的方法和系统具有下述有益效果:

[0107] (1) 基于自动语音识别技术对输入病态嗓音连续语音信号进行音素切分,并按照不同音素的发声特点对切分的语音序列进行分类,如元音和辅音,浊音和清音;

[0108] (2) 针对不同的语音序列设计不同的特征参数,对元音使用传统的jitter, shimmer参数以反映声带震动的规律性,对清音使用MFCC和HiLo参数以声道滤波器的变化和能量在各频段的分布;

[0109] (3) 针对不同嗓音问题训练不同的基于MLP的评分模型,实现对病态嗓音在整体和声嘶等10种嗓音问题的客观评分,为临床诊断和康复治疗提供重要参考。

[0110] 本方法/系统能实现对固定文本内容的连续语音进行自动嗓音评估。若将来建成大规模病态嗓音连续语音数据库,本方法/系统则可推广至任意内容的自然连续语音。现有技术不考虑连续语音中语言内容的变化,对所有语音序列提取相同特征参数。本方法/系统能分别处理不同语言内容类型的语音序列,可实现更细致分析某嗓音问题对不同语音内容发声的影响,为康复治疗提供指导建议。不同于现有技术中对嗓音的单一评价,本方法/系统对嗓音作多维度评估,实现对嗓音问题种类和严重程度的判断。

[0111] 在人力许可下,本方法/系统的音素切分模块也可使用人工标注方法,此方法/系统能得到更高精度的音素时间序列。若有大规模病态嗓音的连续语音数据库,可用来训练自动中文语音识别系统,相比使用正常语音训练的语音识别系统,将得到更可靠的音素切分时间序列。在评分模块,也可采用稀疏估计的方法来量化区分病态嗓音的类别和严重程度。

[0112] 以上内容是结合具体的优选实施方式对本发明所作的进一步详细说明,不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的技术人员来说,在不脱离本发明构思的前提下,还可以做出若干等同替代或明显变型,而且性能或用途相同,都应

当视为属于本发明的保护范围。

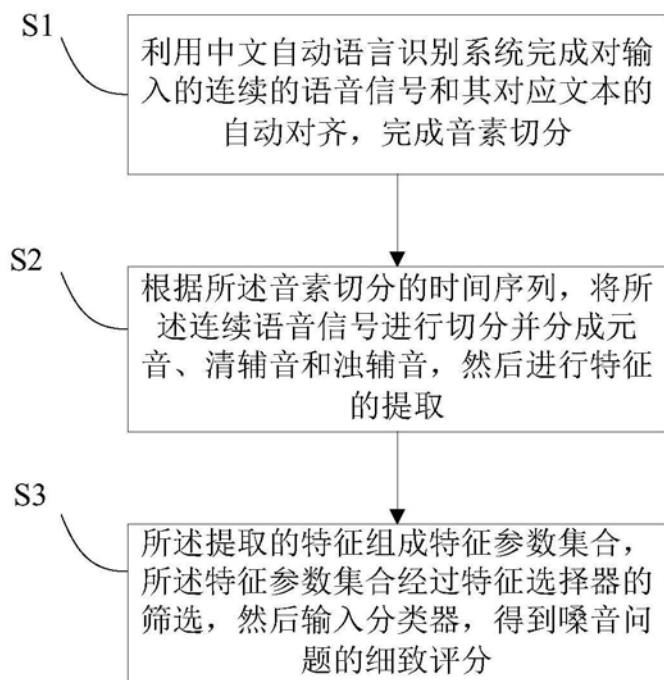


图1

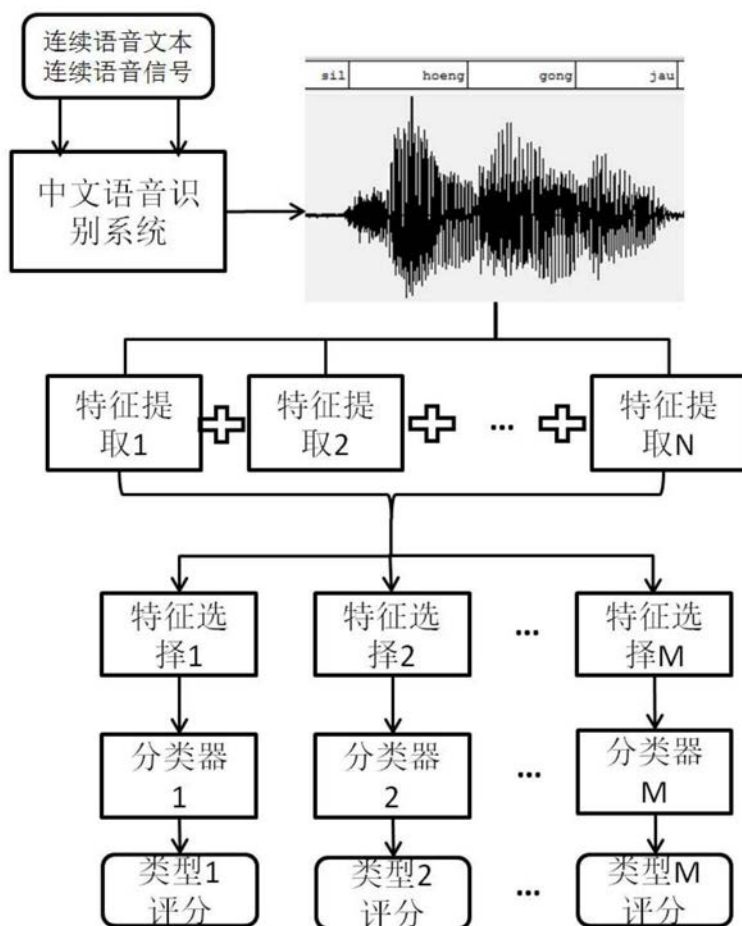


图2

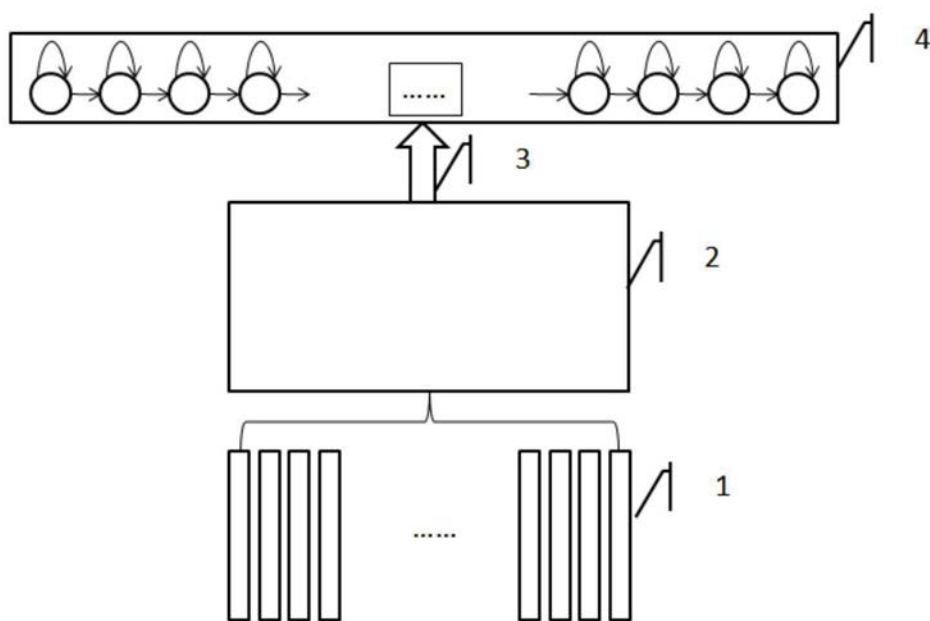


图3

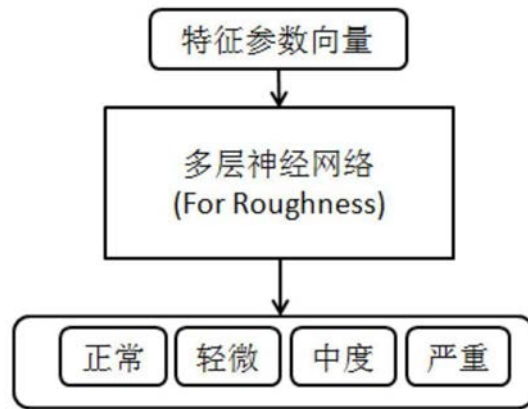


图4

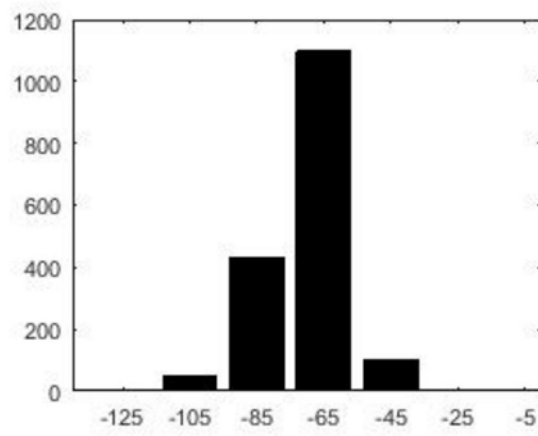


图5

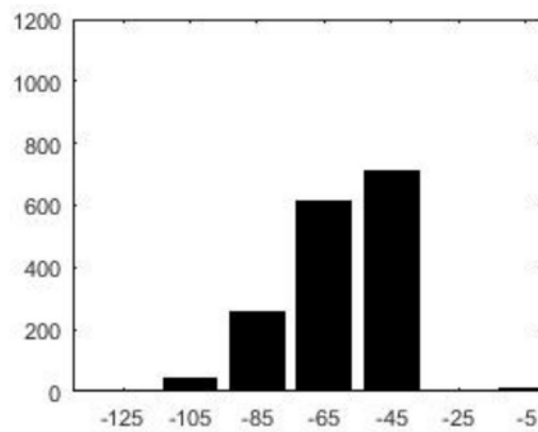


图6

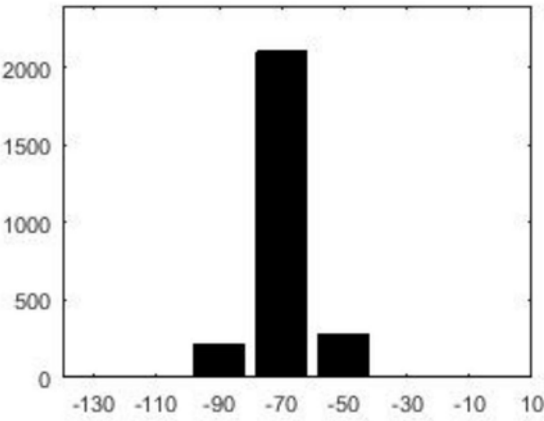


图7

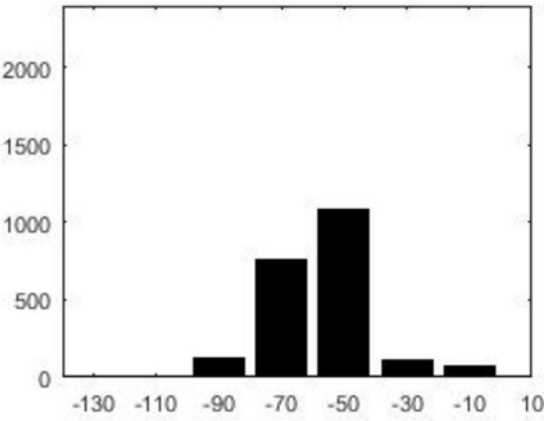


图8

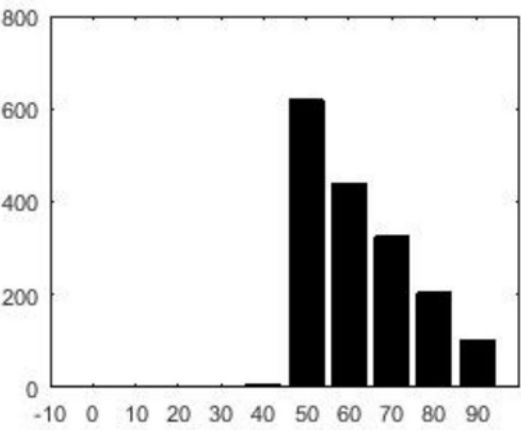


图9

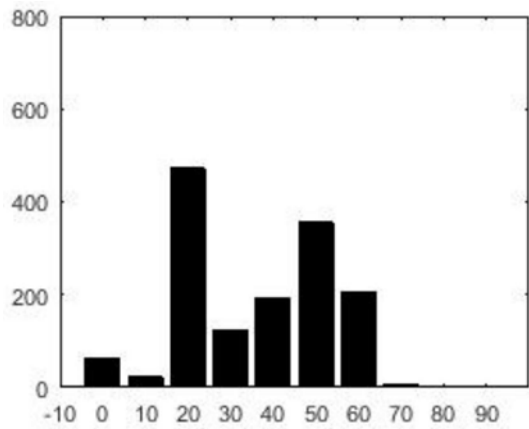


图10

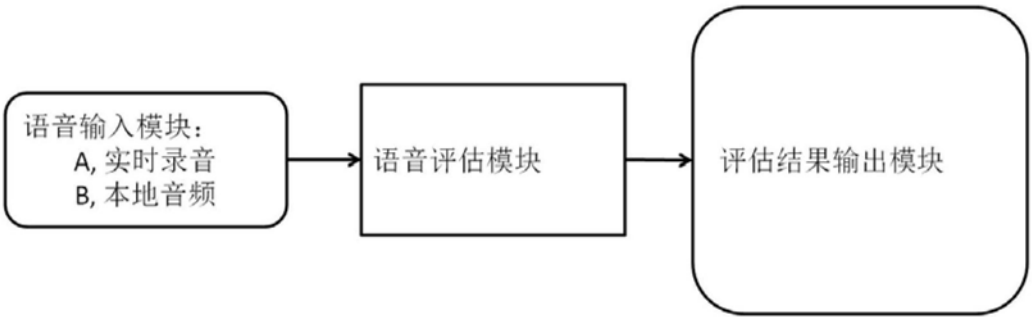


图11