



(12)发明专利申请

(10)申请公布号 CN 110473563 A

(43)申请公布日 2019.11.19

(21)申请号 201910763627.5

(22)申请日 2019.08.19

(71)申请人 山东省计算中心(国家超级计算济南中心)

地址 250014 山东省济南市历下区科院路19号

(72)发明人 李晔 宫晓飞 张杰 张鹏

(74)专利代理机构 济南圣达知识产权代理有限公司 37221

代理人 黄海丽

(51)Int.Cl.

G10L 21/0208(2013.01)

G10L 21/0216(2013.01)

G10L 25/03(2013.01)

G10L 25/45(2013.01)

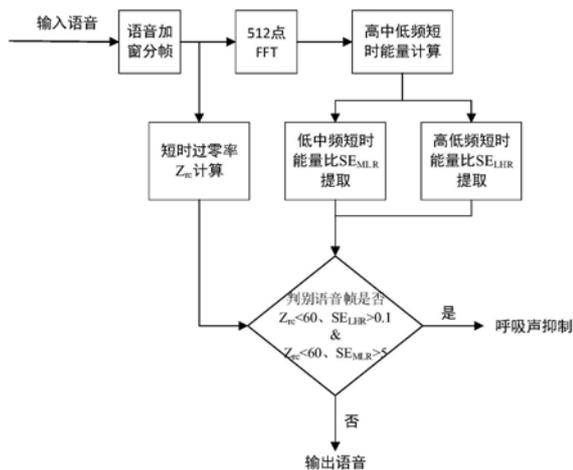
权利要求书1页 说明书6页 附图1页

(54)发明名称

基于时频特征的呼吸声检测方法、系统、设备及介质

(57)摘要

本公开公开了基于时频特征的呼吸声检测方法、系统、设备及介质,包括:获取待检测的语音信号;所述待检测的语音信号是由飞行在高空的飞机上的工作人员发出的,且传输给地面接收员的语音信号;对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号;对每一帧语音信号进行特征提取,提取短时过零率、高低频短时能量比和低中频短时能量比;如果当前帧语音信号的短时过零率大于第一设定阈值,且高低频短时能量比大于第二设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。合利用短时过零率、短时能量、高低频短时能量比等特征参数,检测传输语音中的呼吸声,并对检测到的呼吸声进行后处理,改善听觉效果。



1. 基于时频特征的呼吸声检测方法,其特征是,包括:

获取待检测的语音信号;所述待检测的语音信号是由飞行在高空的飞机上的工作人员发出的,且传输给地面接收员的语音信号;

对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号;

对每一帧语音信号进行特征提取,提取短时过零率、高低频短时能量比和低中频短时能量比;

如果当前帧语音信号的短时过零率大于第一设定阈值,且高低频短时能量比大于第二设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

2. 如权利要求1所述的方法,其特征是,还包括:

如果当前帧语音信号的短时过零率小于第一设定阈值,且低中频短时能量比小于第三设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

3. 如权利要求1所述的方法,其特征是,呼吸声进行消除采用置零法或削减法。

4. 如权利要求1所述的方法,其特征是,对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号,具体包括:按8kHz频率采样,每20ms,也就是160个语音样点构成一帧。

5. 如权利要求3所述的方法,其特征是,置零法是指将当前帧全部采样点进行置零操作。

6. 如权利要求3所述的方法,其特征是,削减法,即对当前帧幅值进行削减;将呼吸声帧的每个样点的幅值削弱为原来的1/8,约为9dB。

7. 基于时频特征的呼吸声检测系统,其特征是,包括:

获取模块,其被配置为获取待检测的语音信号;所述待检测的语音信号是由飞行在高空的飞机上的工作人员发出的,且传输给地面接收员的语音信号;

分帧处理模块,其被配置为对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号;

特征提取模块,其被配置为对每一帧语音信号进行特征提取,提取短时过零率、高低频短时能量比和低中频短时能量比;

呼吸声第一检测模块,其被配置为如果当前帧语音信号的短时过零率大于第一设定阈值,且高低频短时能量比大于第二设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

8. 如权利要求7所述的系统,其特征是,还包括:

呼吸声第二检测模块,其被配置为如果当前帧语音信号的短时过零率小于第一设定阈值,且低中频短时能量比小于第三设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

9. 一种电子设备,包括存储器和处理器以及存储在存储器上并在处理器上运行的计算机指令,所述计算机指令被处理器运行时,完成权利要求1-6任一项方法所述的步骤。

10. 一种计算机可读存储介质,用于存储计算机指令,所述计算机指令被处理器执行时,完成权利要求1-6任一项方法所述的步骤。

## 基于时频特征的呼吸声检测方法、系统、设备及介质

### 技术领域

[0001] 本公开涉及语音信号处理技术领域,特别是涉及基于时频特征的呼吸声检测方法、系统、设备及介质。

### 背景技术

[0002] 本部分的陈述仅仅是提到了与本公开相关的背景技术,并不必然构成现有技术。

[0003] 随着通信技术的快速发展,用户对语音质量的要求越来越高。在很多特殊的通讯场合,除了正常传输的语音之外,往往会受到一些背景噪声或特殊声音的干扰。例如,当飞行员在高空飞行时,大气压力较地面标准大气压力要高出几倍,飞行员与地面通信时需要进行大量的换气,导致话音信号中会含有大量较强的呼吸声,呼吸声经过语音编解码后会收听感觉非常差,且会对地面接收人员的听力造成损伤。

[0004] 在实现本公开的过程中,发明人发现现有技术中存在以下技术问题:

[0005] 对呼吸声的检测不够精准,容易将非呼吸声的语音信号也被误检测出来,影响地面人员对语音信号的捕捉;

[0006] 对呼吸声的检测过程复杂,需要提取复杂的信号特征,检测过程浪费时间,难以保证地面通信员对语音信号捕捉的实时性要求。

### 发明内容

[0007] 为了解决现有技术的不足,本公开提供了基于时频特征的呼吸声检测方法、系统、设备及介质;综合利用短时过零率、短时能量、高低频短时能量比等特征参数,检测传输语音中的呼吸声,并对检测到的呼吸声进行后处理,改善听觉效果。

[0008] 第一方面,本公开提供了基于时频特征的呼吸声检测方法;

[0009] 基于时频特征的呼吸声检测方法,包括:

[0010] 获取待检测的语音信号;所述待检测的语音信号是由飞行在高空的飞机上的工作人员发出的,且传输给地面接收员的语音信号;

[0011] 对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号;

[0012] 对每一帧语音信号进行特征提取,提取短时过零率、高低频短时能量比和低中频短时能量比;

[0013] 如果当前帧语音信号的短时过零率大于第一设定阈值,且高低频短时能量比大于第二设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0014] 作为一个或多个实施例,所述方法,还包括:

[0015] 如果当前帧语音信号的短时过零率小于第一设定阈值,且低中频短时能量比小于第三设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0016] 第二方面,本公开还提供了基于时频特征的呼吸声检测系统;

[0017] 基于时频特征的呼吸声检测系统,包括:

[0018] 获取模块,其被配置为获取待检测的语音信号;所述待检测的语音信号是由飞行

在高空的飞机上的工作人员发出的,且传输给地面接收员的语音信号;

[0019] 分帧处理模块,其被配置为对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号;

[0020] 特征提取模块,其被配置为对每一帧语音信号进行特征提取,提取短时过零率、高低频短时能量比和低中频短时能量比;

[0021] 呼吸声第一检测模块,其被配置为如果当前帧语音信号的短时过零率大于第一设定阈值,且高低频短时能量比大于第二设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0022] 作为一个或多个实施例,所述系统,还包括:

[0023] 呼吸声第二检测模块,其被配置为如果当前帧语音信号的短时过零率小于第一设定阈值,且低中频短时能量比小于第三设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0024] 第三方面,本公开还提供了一种电子设备,包括存储器和处理器以及存储在存储器上并在处理器上运行的计算机指令,所述计算机指令被处理器运行时,完成第一方面所述方法的步骤。

[0025] 第四方面,本公开还提供了一种计算机可读存储介质,用于存储计算机指令,所述计算机指令被处理器执行时,完成第一方面所述方法的步骤。

[0026] 与现有技术相比,本公开的有益效果是:

[0027] 本公开综合利用语音帧与带有呼吸声的语音帧之间的短时过零率、短时能量、高低频短时能量比等特征参数之间的差别,检测出语音中的呼吸声,并对呼吸声进行处理,从而对带有呼吸声的语音帧的抑制效果。

## 附图说明

[0028] 构成本申请的一部分的说明书附图用来提供对本申请的进一步理解,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。

[0029] 图1为第一个实施例的方法流程图。

## 具体实施方式

[0030] 应该指出,以下详细说明都是示例性的,旨在对本申请提供进一步的说明。除非另有指明,本文使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解的相同含义。

[0031] 需要注意的是,这里所使用的术语仅是为了描述具体实施方式,而非意图限制根据本申请的示例性实施方式。如在这里所使用的,除非上下文另外明确指出,否则单数形式也意图包括复数形式,此外,还应当理解的是,当在本说明书中使用术语“包含”和/或“包括”时,其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

[0032] 实施例一,本实施例提供了基于时频特征的呼吸声检测方法;

[0033] 如图1所示,基于时频特征的呼吸声检测方法,包括:

[0034] 获取待检测的语音信号;所述待检测的语音信号是由飞行在高空的飞机上的工作人员发出的,且传输给地面接收员的语音信号;

[0035] 对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号;

[0036] 对每一帧语音信号进行特征提取,提取短时过零率、高低频短时能量比和低中频短时能量比;

[0037] 如果当前帧语音信号的短时过零率大于第一设定阈值,且高低频短时能量比大于第二设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0038] 作为一个或多个实施例,所述方法,还包括:

[0039] 如果当前帧语音信号的短时过零率小于第一设定阈值,且低中频短时能量比小于第三设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0040] 作为一个或多个实施例,呼吸声进行消除采用置零法或削减法。

[0041] 作为一个或多个实施例,对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号,具体包括:

[0042] 按8kHz频率采样,每20ms,也就是160个语音样点构成一帧。

[0043] 应理解的,置零法是指将当前帧全部采样点进行置零操作;

[0044] 将呼吸声帧进行置零,公式如下:

$$[0045] \quad S_{\text{out}} = \begin{cases} 0, & Z_{\text{rc}} > 60, SE_{\text{LHR}} > 0.1 \\ 0, & Z_{\text{rc}} < 60, SE_{\text{MLR}} > 5 \\ S_{\text{in}}, & \text{其他} \end{cases}$$

[0046] 应理解的,削减法,即对当前帧幅值进行削减;

[0047] 将呼吸声帧的每个样点的幅值削弱为原来的1/8,约为9dB,公式如下:

$$[0048] \quad S_{\text{out}} = \begin{cases} \frac{1}{8} S_{\text{in}}, & Z_{\text{rc}} > 60, SE_{\text{LHR}} > 0.1 \\ \frac{1}{8} S_{\text{in}}, & Z_{\text{rc}} < 60, SE_{\text{MLR}} > 5 \\ S_{\text{in}}, & \text{其他} \end{cases}$$

[0049] 其中, $S_{\text{in}}$ 与 $S_{\text{out}}$ 分别为输入的语音帧和输出的语音帧的能量。

[0050] 应理解的,短时过零率 $Z_{\text{rc}}$ 的计算过程为:

$$[0051] \quad Z_{\text{rc}} = \sum_{i=0}^{N-1} | \text{sgn}[s_w(i)] - \text{sgn}[s_w(i-1)] | \quad (1)$$

[0052] 其中, $\text{sgn}[\ ]$ 是符号函数,即

$$[0053] \quad \text{sgn}[x] = \begin{cases} 1, & (x \geq 0) \\ -1, & (x < 0) \end{cases}$$

[0054] 其中, $i$ 表示第 $i$ 帧。

[0055] 应理解的,高低频短时能量比的计算过程为:

$$[0056] \quad SE_{\text{LHR}} = \frac{E_{\text{OL}}}{E_{\text{OH}}}$$

[0057] 其中,低频短时能量 $E_{\text{OL}}$ 为,

[0058]  $E_{0L} = \sum_{n=0}^{128} RE_w^2(i) + IM_w^2(i);$

[0059] 其中,高频短时能量 $E_{0H}$ 为,

[0060]  $E_{0H} = \sum_{n=128}^{256} RE_w^2(i) + IM_w^2(i)$

[0061] 应理解的,低中频短时能量比 $SE_{MLR}$ 的计算过程为:

[0062]  $SE_{MLR} = \frac{E_{0M}}{E_{0L}}$

[0063] 其中,中频短时能量 $E_{0M}$ 为,

[0064]  $E_{0M} = \sum_{n=60}^{80} RE_w^2(i) + IM_w^2(i)$

[0065] 低频短时能量 $E_{0L}$ 为,

[0066]  $E_{0H} = \sum_{n=0}^{50} RE_w^2(i) + IM_w^2(i)$

[0067] 进一步地,所述第一设定阈值为:60。

[0068] 进一步地,所述第二设定阈值为:0.1。

[0069] 进一步地,所述第三设定阈值为:5。

[0070] 本发明的特点主要是针对特殊环境下的语音通信中,呼吸声的存在使得经过语音编解码后的语音收听感觉非常差,会对地面接收人员的听力造成损伤。因此提出了一种基于时频特征的呼吸声检测算法,综合利用短时过零率、短时能量、高低频语音能量比等特征参数,检测传输语音中的呼吸声,并对检测到的呼吸声进行后处理,从而减少呼吸声在语音通话过程中对人体产生的不适,改善语音通话质量。

[0071] 通过实验证明,在一段带有呼吸声的语音当中,呼吸声与传输语音之间的短时过零率存在差别,与此同时,短时能量比相较于短时过零率对呼吸声的影响更为显著,结合过零率和短时能量比参数对正常语音和呼吸声具有更好的区分度。因此,确定短时过零率及短时能量比为判别语音帧中呼吸声的特征参数。

[0072] 首先求出分帧后的语音帧短时过零率 $Z_{rc}$ 等时域参数,并提取出带有呼吸声语音帧参数的值;之后对语音信号样点进行短时傅里叶变换(FFT),求出语音信号的高、中、低频短时能量,并通过高低频短时能量比 $SE_{LHR}$ 与低中频短时能量比 $SE_{MLR}$ 将语音特征表示出来,提取出带有呼吸声语音帧的参数的值。

[0073] 本发明主要是综合利用短时过零率、高低频短时能量比等特征参数,判别出语音帧与带有呼吸声的语音帧,并对带有呼吸声参数的语音帧进行抑制。本发明可以实现呼吸声的检测,并对伴有呼吸声的语音帧进行呼吸声抑制,从而减少呼吸声在语音通话过程中对人体产生的不适,实现更为舒适、清晰通话的目的。

[0074] 实施例二,本实施例还提供了基于时频特征的呼吸声检测系统;

[0075] 基于时频特征的呼吸声检测系统,包括:

[0076] 获取模块,其被配置为获取待检测的语音信号;所述待检测的语音信号是由飞行在高空的飞机上的工作人员发出的,且传输给地面接收员的语音信号;

[0077] 分帧处理模块,其被配置为对待检测的语音信号进行加窗分帧处理,得到若干帧语音信号;

[0078] 特征提取模块,其被配置为对每一帧语音信号进行特征提取,提取短时过零率、高低频短时能量比和低中频短时能量比;

[0079] 呼吸声第一检测模块,其被配置为如果当前帧语音信号的短时过零率大于第一设定阈值,且高低频短时能量比大于第二设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0080] 作为一个或多个实施例,所述系统,还包括:

[0081] 呼吸声第二检测模块,其被配置为如果当前帧语音信号的短时过零率小于第一设定阈值,且低中频短时能量比小于第三设定阈值;则判定当前帧语音信号为呼吸声,对呼吸声进行消除。

[0082] 本公开还提供了一种电子设备,包括存储器和处理器以及存储在存储器上并在处理器上运行的计算机指令,所述计算机指令被处理器运行时,完成方法中的各个操作,为了简洁,在此不再赘述。

[0083] 所述电子设备可以是移动终端以及非移动终端,非移动终端包括台式计算机,移动终端包括智能手机(Smart Phone,如Android手机、IOS手机等)、智能眼镜、智能手表、智能手环、平板电脑、笔记本电脑、个人数字助理等可以进行无线通信的移动互联网设备。

[0084] 应理解,在本公开中,该处理器可以是中央处理单元CPU,该处理器还算是其他通用处理器、数字信号处理器DSP、专用集成电路ASIC,现成可编程门阵列FPGA或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0085] 该存储器可以包括只读存储器和随机存取存储器,并向处理器提供指令和数据、存储器的一部分还可以包括非易失性随机存储器。例如,存储器还可以存储设备类型的信息。

[0086] 在实现过程中,上述方法的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。结合本公开所公开的方法的步骤可以直接体现为硬件处理器执行完成,或者用处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器、闪存、只读存储器、可编程只读存储器或者电可擦写可编程存储器、寄存器等领域成熟的存储介质中。该存储介质位于存储器,处理器读取存储器中的信息,结合其硬件完成上述方法的步骤。为避免重复,这里不再详细描述。本领域普通技术人员可以意识到,结合本文中公开的实施例描述的各示例的单元即算法步骤,能够以电子硬件或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0087] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0088] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以

通过其他方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能的划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另外一点,所显示或讨论的相互之间的耦合或者直接耦合或者通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性、机械或其它的形式。

[0089] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机、服务器或者网络设备)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0090] 以上所述仅为本申请的优选实施例而已,并不用于限制本申请,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

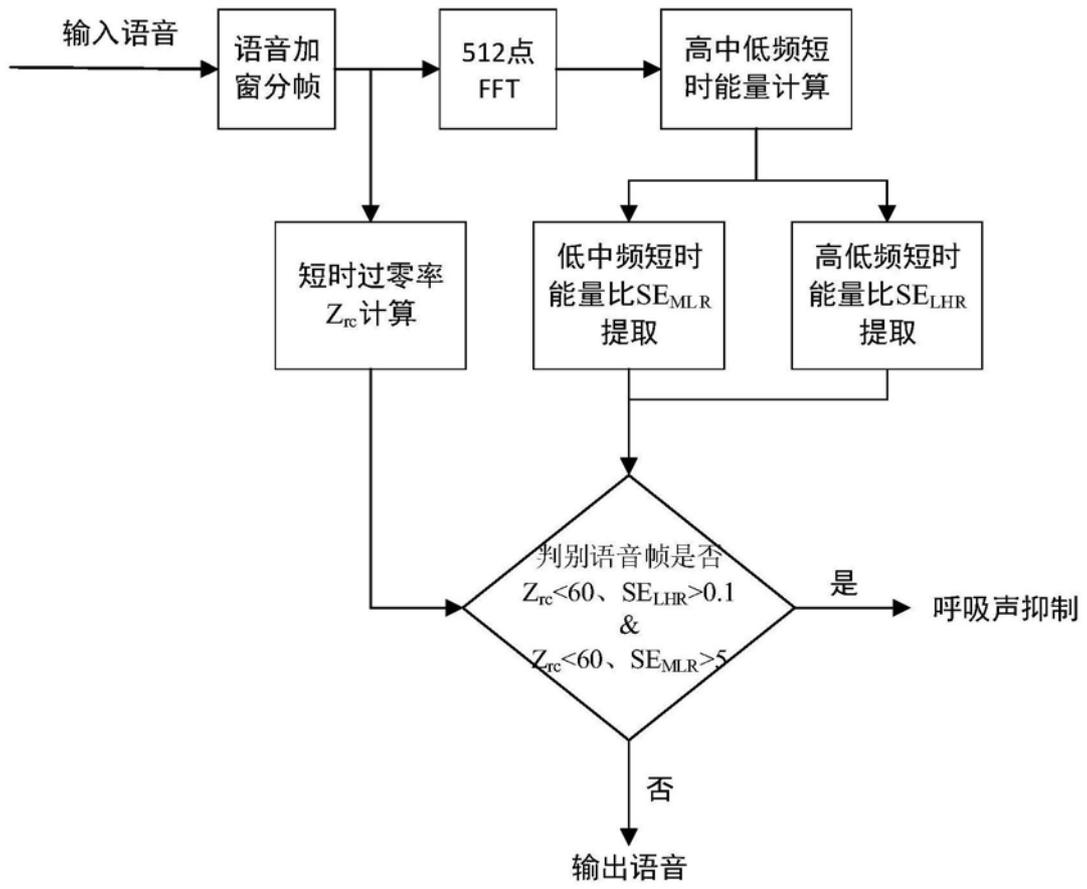


图1