



(12) 发明专利

(10) 授权公告号 CN 112151056 B

(45) 授权公告日 2023. 08. 04

(21) 申请号 202011032720.8

CN 108711419 A, 2018.10.26

(22) 申请日 2020.09.27

CN 109559757 A, 2019.04.02

(65) 同一申请的已公布的文献号

CN 106653044 A, 2017.05.10

申请公布号 CN 112151056 A

CN 106658323 A, 2017.05.10

(43) 申请公布日 2020.12.29

WO 2008092183 A1, 2008.08.07

(73) 专利权人 浙江诺尔康神经电子科技股份有限公司

US 2012215519 A1, 2012.08.23

地址 311100 浙江省杭州市余杭区仓前街道龙潭路17号

CN 101763858 A, 2010.06.30

CN 103761974 A, 2014.04.30

CN 109961785 A, 2019.07.02

(72) 发明人 韩彦 黄穗 孙晓安

陈又圣; 宫琴. 基于双TP型麦克风的电子耳蜗前端指向性语音增强系统的研制. 仪器仪表学报. 2010, (09), 全文.

(51) Int. Cl.

审查员 彭鼎原

G10L 21/0208 (2013.01)

G10L 25/30 (2013.01)

(56) 对比文件

CN 106098077 A, 2016.11.09

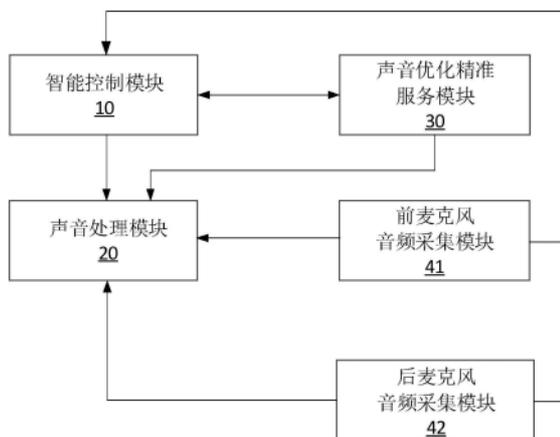
权利要求书5页 说明书12页 附图3页

(54) 发明名称

带自定义的智能耳蜗声音处理系统和方法

(57) 摘要

本发明公开了一种带自定义的智能耳蜗声音处理系统和方法, 智能控制模块与声音处理模块、声音精准优化模块、前麦克风音频采集模块和后麦克风音频采集模块分别连接, 声音处理模块还与声音精准优化模块连接, 前麦克风音频采集模块和后麦克风音频采集模块还与声音处理模块连接, 智能控制模块对环境声进行监听和识别, 根据识别的结果, 实时调整声音处理模块的参数配置; 声音处理模块对声音信号进行处理, 包括降噪和语音增强; 声音精准优化模块提供精准的个性化处理及优化, 可生成自定义的声音处理策略; 前麦克风音频采集模块和后麦克风音频采集模块分别设置在智能耳蜗体外机的前侧和后侧, 对外界声音进行采集。



1. 一种带自定义的智能耳蜗声音处理系统,其特征在于,包括智能控制模块、声音处理模块、声音精准优化模块、前麦克风音频采集模块和后麦克风音频采集模块,其中,智能控制模块与声音处理模块、声音精准优化模块、前麦克风音频采集模块和后麦克风音频采集模块分别连接,声音处理模块还与声音精准优化模块连接,前麦克风音频采集模块和后麦克风音频采集模块还与声音处理模块连接,智能控制模块对环境声进行监听和识别,根据识别的结果,实时调整声音处理模块的参数配置;声音处理模块对声音信号进行处理,包括降噪和语音增强;声音精准优化模块提供精准的个性化处理及优化,可生成自定义的声音处理策略;前麦克风音频采集模块和后麦克风音频采集模块分别设置在智能耳蜗体外机的前侧和后侧,对外界声音进行采集;

所述智能控制模块包括音频预处理模块、音频成分检测模块和音频成分综合分析决策模块,其中,音频预处理模块包括能量检测模块和音频信号特征分析模块,所述能量检测模块通过监测能量的方式,从声音信号流里检测静音期,如处于静音期,则音频成分检测模块将其判定为安静环境,同时避开音频信号特征分析模块的处理和音频成分检测模块的判定,直接传送给音频成分综合分析模块进行统计分析;如非静音期,则将音频信号送给音频信号特征分析模块,将信号的物理特征提取出来并进行统计分析,形成特征集合,将该特征集合提供给音频成分检测模块,计算上述特征的值时,以4ms为一个帧,统计计算0.5s即125帧特征的值的均值和方差;

音频成分检测模块对特征集合进行分析,识别出该特征集合表征的音频信号包含的成分或者属于的环境,将分类结果传递给音频成分综合分析决策模块;

音频成分综合分析决策模块对环境向量进行统计分析,调控声音处理模块的处理策略,包括状态反馈模块、统计分析模块和决策输出模块;状态反馈模块将环境向量与一个状态转移矩阵相乘,得到一个通过先验概率加权评估后的环境向量,传递给统计分析模块,状态转移矩阵里的每一个元素代表了一类声音转换到另一类声音的概率,其值越大,代表可能性越高;统计分析模块对一段时间内的环境向量进行统计,将统计结果超过一定阈值的最多两种声音类别作音频成分的最终决定,发送给决策输出模块,上述一段时间为6-15s;决策输出模块根据统计分析模块的对音频成分的最终判定,控制声音处理模块的处理策略,包括开启声音处理算法以及选择算法的参数配置;

所述声音处理模块在智能控制模块的调控下对声音进行降噪和语音增强的处理,包括双麦克风降噪模块、单麦克风降噪模块、语音增强模块和参数配置模块,声音信号被前麦克风音频采集模块和后麦克风音频采集模块采集之后,依次经过双麦克风降噪、单麦克风降噪和语音增强的处理;

所述单麦克风降噪模块通过噪声估计和噪声消除,提高语音信号的信噪比;

所述双麦克风降噪模块处理不同方位的声源发出的声音,到达一前一后两个麦克风的时间延时不同,反向判断目标声源以及噪声源的方向,进行延时相加,以及差分麦克风阵列中分别控制目标声源及噪声源的增益;

所述语音增强模块通过改变时域输出的幅度,拟合汉语4声调的基频,提高用户对汉语声调的感知能力;

所述参数配置模块包括双麦克风降噪参数配置表、单麦克风降噪参数配置表和语音增强参数配置表,每个配置表均包括一组默认参数配置作为预设值、一组可选参数配置作为

备用选项和4组自定义参数配置,其中可选参数配置和自定义参数配置根据用户需求可个性化设置,可选参数配置在默认参数的基础上,根据实际听声感受进行微调;自定义参数配置为匹配自定义环境检测而设置的模块,当检测到自定义的声音环境时,使用与之对应的自定义参数配置;语音增强参数配置表中包括一组13阶低通滤波器的系数、平滑因子、增益以及增益变化的步长;单麦克风降噪参数配置表中包括两个常数项和两个平滑因子;双麦克风降噪参数配置表中包括一个侧向的增益因子和后向三个不同角度的增益因子;

所述声音精准优化模块包括手机应用模块、云端服务器和用户定制服务单元,其中,手机应用模块包括用户自定义引导单元、服务请求单元和服务反馈单元,配置手机应用模块首先通过手机应用模块控制耳蜗采集环境声并通过音频信号特征分析模块提取其特征值,并将特征值和音频发送给云端服务器等待处理,然后通过手机应用模块递交服务请求,云端服务器收到请求后,启动用户定制服务单元,对上传的数据进行分析和处理,训练自定义环境检测单元的参数,并调整与之匹配的声音处理模块的自定义参数配置,云端服务器将训练和调试好的参数,下载到手机,再通过手机应用模块的服务反馈单元将参数部署到耳蜗中;所述用户自定义引导单元通过音频信号特征分析模块将所需自定义的声音环境的音频信号特征进行提取,并将特征值发送到云端服务器进行存储;

所述服务请求单元将系统信息及用户信息提供给云端服务器并发送给用户定制服务单元;

所述云端服务器存储和管理用户数据和信息;

所述用户定制服务单元根据用户递交的请求对数据进行处理,然后通过数据训练自定义环境检测单元的参数,并调整自定义参数配置,待完成参数的训练和调试后,通过云端服务器下载到用户的手机端;

所述服务反馈单元在接收到云端服务器传来的配置文件后,将配置部署到声音处理模块中。

2. 根据权利要求1所述的带自定义的智能耳蜗声音处理系统,其特征在于,所述特征集合包括谱质心、谱熵、不同子频段划分的子带能量率、单帧谱通量、多帧谱通量和频谱能量互相关。

3. 根据权利要求1所述的带自定义的智能耳蜗声音处理系统,其特征在于,所述音频成分检测模块包括至少10个预设检测模块和4个自定义环境检测单元,10个预设检测模块包含言语成分检测单元、耳语成分检测单元、安静环境检测单元、车内噪声环境检测单元、空调环境检测单元、言语噪声环境检测单元、街道环境检测单元、雨噪环境检测单元、风噪环境检测单元和强噪声环境检测单元,每个检测单元均包括基于二分类神经网络模型的分器,模型参数通过包含相应声音的音频数据库训练得出,该音频数据库包括言语、耳语、车内噪声、空调噪声、言语噪声、街道噪声、雨噪、风噪、强噪,每个二分类器均采用一对多的分类模式,即在训练言语成分检测模块的分器时,将数据库中的全部音频划分为言语声和非言语声,如是训练耳语成分检测单元的分器时,将数据库中的全部音频划分为耳语声和非耳语声,以此类推,在均衡了两类数据样本之后,对各自的分类器进行训练;自定义检测单元用于检测用户自定义的声音或环境类别,所有检测单元的输出结果组成一个一维的向量,即环境向量,提供给音频成分综合分析决策模块,环境向量中的每一个元素是归一后从0到1之间的小数,值越大代表音频中包含该元素对应的声音类别的可能性越大。

4. 一种采用权利要求1-3之一所述的系统的单麦克风降噪方法,其特征在於,包括以下步骤:

所述单麦克风降噪模块进行噪声估计和噪声消除,首先使用改进的最小值跟踪对噪声能量进行估计,然后根据估计的SNR计算相应的增益值,最后再对噪声进行消除。

5. 根据权利要求4所述的系统的单麦克风降噪方法,其特征在於,所述噪声估计为,设输入信号 $y(n)$ 由语音信号 $x(n)$ 和噪音信号 $d(n)$ 组成:

$$y(n) = x(n) + d(n) \quad (1)$$

对上述输入信号进行短时傅里叶变换,变换后第 k 通道,第 λ 帧的功率谱记为 $Y(\lambda, k)$,平滑后可得:

$$P(\lambda, k) = \eta P(\lambda - 1, k) + (1 + \eta) |Y(\lambda, k)|^2 \quad (2)$$

其中, $P(\lambda, k)$ 表示平滑后的功率谱, η 表示平滑因子;

对带噪信号的局部最小能量值 $P_{\min}(\lambda, k)$ 进行估计:

$$P_{\min}(\lambda, k) = \begin{cases} P(\lambda, k), P_{\min}(\lambda - 1, k) \geq P(\lambda, k) \\ \gamma P_{\min}(\lambda - 1, k) + \frac{1 - \gamma}{1 - \beta} (P(\lambda, k) - \beta P(\lambda - 1, k)), P_{\min}(\lambda - 1, k) < P(\lambda, k) \end{cases} \quad (3)$$

其中, β 和 γ 为两个常数,根据式(3)估计的局部最小功率谱来计算语音信号的出现概率,首先,计算当前功率谱与局部最小功率谱的比值 $S_r(\lambda, k)$:

$$S_r(\lambda, k) = \frac{P(\lambda, k)}{P_{\min}(\lambda, k)} \quad (4)$$

之后,引入一个阈值 $T(\lambda, k)$ 来确定当前帧是否包含语音信号:

$$I(\lambda, k) = \begin{cases} 1, S_r(\lambda, k) \geq T(\lambda, k) \\ 0, S_r(\lambda, k) < T(\lambda, k) \end{cases} \quad (5)$$

其中, $I(\lambda, k)$ 表示当前帧中是否包含语音信号的判断标识,阈值 $T(\lambda, k)$ 为动态阈值,根据上一帧所估计的SNR得出,SNR越高,阈值越低,反之亦然;

对 $I(\lambda, k)$ 进行平滑后可得:

$$K(\lambda, k) = \alpha K(\lambda, k) + (1 - \alpha) I(\lambda, k) \quad (6)$$

其中, $K(\lambda, k)$ 表示语音信号在当前帧出现的概率,由此结果来更新噪声估计的平滑因子:

$$\alpha_s(\lambda, k) = \alpha_d + (1 - \alpha_d) K(\lambda, k) \quad (7)$$

其中, α_d 为一个常数,最终得到当前帧每个通道的噪声功率谱估计:

$$D(\lambda, k) = \alpha_s(\lambda, k) D(\lambda - 1, k) + (1 - \alpha_s(\lambda, k)) |Y(\lambda, k)|^2 \quad (8)$$

所述噪声消除为,根据上述估计的噪声功率谱计算信噪比SNR:

$$\text{SNR}(\lambda, k) = \delta \text{SNR}(\lambda - 1, k) + (1 - \delta) |Y(\lambda, k)|^2 \quad (9)$$

由此可得到一个增益系数:

$$G(\lambda, k) = \frac{SNR(\lambda, k)}{SNR(\lambda, k) + 1} \quad (10)$$

最终根据增益系数 $G(\lambda, k)$ 对信号进行降噪处理,得到降噪后的语音信号功率谱:

$$S(\lambda, k) = G_0(\lambda, k)P(\lambda, k) \quad (11)$$

其中, $G_0(\lambda, k)$ 表示最终的降噪增益系数。

6.一种采用权利要求1-3之一所述的系统的双麦克风降噪方法,其特征在于,所述双麦克风降噪模块利用不同方位的声源发出的声音,到达前后两个麦克风的时间延时不同的特点,反向判断目标声源以及噪声源的方向,再利用延时相加,以及差分麦克风阵列中分别控制目标声源及噪声源的增益方法,提升信噪比,降低噪声干扰。

7.根据权利要求6所述的系统的双麦克风降噪方法,其特征在于,在语音信号进行频域信号处理时,对后麦克风音频采集模块的语音信号进行频域延时处理,频域信号延时处理采用离散傅里叶变换的时间位移定理,即频域的乘积相当于时域的延时 $s(t+\tau)$:

$$s(t+\tau) = S(f) \cdot e^{\left(-j\omega \frac{d}{c}\right)} \quad (12)$$

其中, $s(t+\tau)$ 表示延时 $-\tau$ 后的信号, $S(f)$ 表示原信号的频谱, d 为前后两个麦克风间的直线距离, c 为声音在空气中的传播速度, $\omega = 2\pi \frac{k}{N/2} * f_s$, N 为FFT点数, $k=0 \sim N/2$, f_s

为系统信号采样率;

频域信号的延时相加利用离散傅里叶变换的线性性质实现,即频域两个信号的相加等价于时域两个信号的叠加,在频域中前麦克风音频采集模块采集的语音信号与进行延时处理后的后向信号进行叠加处理:

$$s_{\text{sum}}(t) = s_F(t) + s_B(t+d/c) \quad (13)$$

其中, $s_{\text{sum}}(t)$ 表示叠加处理后的信号, $s_F(t)$ 表示前置麦克风接收信号, $s_B(t+d/c)$ 表示经过时延矫正的后置麦克风接收信号,在进行完信号的延时相加处理后,对其进行相位补偿,其参考相位以前麦克风音频采集模块采集的语言信号的相位信息为标准;

使用差分麦克风阵列是为了获取两个背靠背的单极性心形麦克风信号输出,并通过分别控制前向增益系数,如前向增益为1,保留正前方信号;和后向增益系数 β_0 ,即衰减后向信号强度的大小,以获取所需要的心形输出,差分麦克风阵列由下式实现:

$$\begin{aligned} C_F(f) &= \Phi(S_{\text{sum}}(f) - e^{-i\omega d/c} \cdot S_{B_delay}(f)) \\ C_B(f) &= \Phi(S_{\text{sum}}(f) - e^{i\omega d/c} \cdot S_{B_delay}(f)) \\ \Phi &= 1/(1 - e^{2i\omega d/c}) \end{aligned} \quad (14)$$

其中, C_F 表示前向差分后的信号频谱, C_B 表示后向差分后的信号频谱, Φ 表示差分后的补偿滤波器, $S_{\text{sum}}(f)$ 表示 $s_{\text{sum}}(t)$ 的频谱, $S_{B_delay}(f)$ 表示 $s_B(t+d/c)$ 的频谱;在获取前向和后向背靠背的心形输出信号后,通过调整后向输出的后向增益系数 β_0 ,再以前向心形信号减去后向心形信号,以获得最终差分麦克风阵列输出;其实现方式如下式所示:

$$s_{\text{out}}(t) = c_F(t) - \beta_0 c_B(t), \beta_0 \approx (1 + \cos\theta_n)/(1 - \cos\theta_n) \quad (15)$$

其中, $s_{out}(t)$ 表示经过差分波束形成的信号, c_F 表示前向差分后的信号, c_B 表示后向差分后的信号, θ_n 表示抑制噪声的方向, 如固定后向差分麦克风阵列输出信号的后向增益系数 β_0 , 则为定向消除某一方向的噪声信号, 而保留正前方语音信号; 如自动修改或选择相应的后向增益系数 β_0 , 则实现正前方语音信号保留, 同时自适应的抑制除正前方目标语音信号外的其他方向的噪声信号, 为了实现不同方向的噪声源选择相应的后向增益系数 β_0 的策略, 则首先需判定出噪声源的位置, 将某一方向的语音信号映射分布在半径相同的球面上, 各个方向的延时时间不同, 计算出语音信号分布在 0° 到 180° 方向的信号能量大小, 以判断噪声源的位置; 然后, 根据噪声源所分别的区域选择相应的后向增益系数 β_0 , 以实现自适应追踪噪声源位置, 抑制不同方向的噪声信号, 由下式:

$$Energy(\theta) = \sum_{k=0}^{k-\frac{N}{2}} \left(S_F(k) + e^{-i\omega^d \cos\theta} / c S_B(k) \right)^2 \quad (16)$$

其中, θ 为语音信号映射方位角, $Energy(\theta)$ 表示 θ 方向信号能量强度, S_F 表示前置麦克风接收到的信号频谱, S_B 表示后置麦克风接收到的信号频谱; 在获得当前帧语音信号在 0° 到 180° 范围内且角度分辨率为 p 的各个角度的能量分布后, 通过比较大小选择出能量最大的位置 m_{max} , 即可判断出当前信号的噪声源位置, 由于此时获得帧与帧间的 m_{max} 值变化很迅速, 需对其进行平滑处理, 减缓其变化速度, 进一步准确判断声源方位, 并能稳定的维持一定的时间, 其平滑公式:

$$E_{max}(k) = \alpha E_{max}(k-1) + (1-\alpha) m_{max} \quad (17)$$

其中, α 为平滑因子, k 为信号分帧处理的帧序数, E_{max} 为当前帧经平滑处理后的最大能量角度;

前述的后向增益系数 β_0 , 当其参数固定时, 则为定向抑制某一方向的噪声信号, 保留正前方的目标语音; 当其参数可变时, 则为自适应追踪噪声源位置, 并抑制其噪声信号而保留正前方的目标语音。

8. 一种采用权利要求1-3之一所述的系统的语音增强方法, 其特征在于, 所述语音增强模块提高汉语声调感知, 包括以下步骤:

首先, 将声音信号分为 M 个通道进行包络提取; 然后提取声音信号的基频信息, 并根据基频信息对每个通道的包络进行调整, 模拟音调的变化; 最后, 耳蜗系统根据包络调整后的信号对各通道进行刺激。

带自定义的智能耳蜗声音处理系统和方法

技术领域

[0001] 本发明属于信号处理领域,特别涉及一种带自定义的智能耳蜗声音处理系统和方法。

背景技术

[0002] 人工耳蜗是目前市场上唯一一种可以有效地让重度或极重度耳聋患者恢复听力的医疗器械。一般的人工耳蜗的工作原理是将麦克风采集的声音信号,经过信号处理单元,将声音信号转为刺激编码发送给植入体,植入体按照刺激编码,通过微电极刺激听神经,从而使植入者恢复听力。环境往往存在很多的噪声,会严重影响辅听设备用户的听声感受。通常的做法是使用降噪或者语音增强手段对声音进行处理。然而,单一的声音处理算法或者某一处理算法的单一参数配置难以应付复杂多变的声环境。因此,这类设备大都配有多种声音处理方法以应对不同的情况。在早期,在何种情况下使用何种方法需要用户手动选择,随着人工智能技术的不断发展,现在出现了自动选择处理方法的声感知系统。实验表明,相比自主选择,系统自动选择的处理方法更能提高用户的言语辨识能力。但是,人工耳蜗这类设备,由于每个患者的情况不同,具体说是每个患者耳蜗内残留的听神经情况不同;同时,植入体电极插入耳蜗的位置和电极的分布又不可能完全一致,因此,实际情况是,通用的方法加上通用的配置,虽然能在大量用户样本上起作用,不同程度的提高了这类用户的言语辨识能力,但对单个用户并不是最优的,至少少量用户起到负作用。其次,系统选择的通用最佳方案不一定会被用户采纳,有些用户甚至拒绝接受系统推荐的方案,坚持自己选择,因此,大大降低该系统的收益。

发明内容

[0003] 有鉴于此,本发明提供了一种带自定义的智能耳蜗声音处理系统,包括智能控制模块、声音处理模块、声音精准优化模块、前麦克风音频采集模块和后麦克风音频采集模块,其中,智能控制模块与声音处理模块、声音精准优化模块、前麦克风音频采集模块和后麦克风音频采集模块分别连接,声音处理模块还与声音精准优化模块连接,前麦克风音频采集模块和后麦克风音频采集模块还与声音处理模块连接,智能控制模块对环境声进行监听和识别,根据识别的结果,实时调整声音处理模块的参数配置;声音处理模块对声音信号进行处理,包括降噪和语音增强;声音精准优化模块提供精准的个性化处理及优化,可生成自定义的声音处理策略;前麦克风音频采集模块和后麦克风音频采集模块分别设置在智能耳蜗体外机的前侧和后侧,对外界声音进行采集。

[0004] 优选地,所述智能控制模块包括音频预处理模块、音频成分检测模块和音频成分综合分析决策模块,其中,音频预处理模块包括能量检测模块和音频信号特征分析模块,所述能量检测模块通过监测能量的方式,从声音信号流里检测静音期,如处于静音期,则音频成分检测模块将其判定为安静环境,同时避开音频信号特征分析模块的处理和音频成分检测模块的判定,直接传送给音频成分综合分析模块进行统计分析;如非静音期,则将音频信

号送给音频信号特征分析模块,将信号的物理特征提取出来并进行统计分析,形成特征集合,将该特征集合提供给音频成分检测模块,计算上述特征值时,以4ms为一个帧,统计计算0.5s即125帧特征值的均值和方差;

[0005] 音频成分检测模块对特征集合进行分析,识别出该特征集合表征的音频信号包含的成分或者属于的环境,将分类结果传递给音频成分综合分析决策模块;

[0006] 音频成分综合分析决策模块对环境向量进行统计分析,调控声音处理模块的处理策略,包括状态反馈模块、统计分析模块和决策输出模块;状态反馈模块将环境向量与一个状态转移矩阵相乘,得到一个通过先验概率加权评估后的环境向量,传递给统计分析模块,状态转移矩阵里的每一个元素代表了一类声音转换到另一类声音的概率,其值越大,代表可能性越高;统计分析模块对一段时间内的环境向量进行统计,将统计结果超过一定阈值的最多两种声音类别作音频成分的最终决定,发送给决策输出模块,上述一段时间为6-15s;决策输出模块根据统计分析模块的对音频成分的最终判定,控制声音处理模块的处理策略,包括开启声音处理算法以及选择算法的参数配置;

[0007] 所述声音处理模块在智能控制模块的调控下对声音进行降噪和语音增强的处理,包括双麦克风降噪模块、单麦克风降噪模块、语音增强模块和参数配置模块,声音信号被前麦克风音频采集模块和后麦克风音频采集模块采集之后,依次经过双麦克风降噪、单麦克风降噪和语音增强的处理;

[0008] 所述单麦克风降噪模块通过噪声估计和噪声消除,提高语音信号的信噪比;

[0009] 所述双麦克风降噪模块处理不同方位的声源发出的声音,到达一前一后两个麦克风的时间延时不同,反向判断目标声源以及噪声源的方向,进行延时相加,以及差分麦克风阵列中分别控制目标声源及噪声源的增益;

[0010] 所述语音增强模块通过改变时域输出的幅度,拟合汉语4声调的基频,提高用户对汉语声调的感知能力;

[0011] 所述参数配置模块包括双麦克风降噪参数配置表、单麦克风降噪参数配置表和语音增强参数配置表,每个配置表均包括一组默认参数配置作为预设值、一组可选参数配置作为备用选项和4组自定义参数配置,其中可选参数配置和自定义参数配置根据用户需求可个性化设置,可选参数配置在默认参数的基础上,根据实际听声感受进行微调;自定义参数配置为匹配自定义环境检测而设置的模块,当检测到自定义的声音环境时,使用与之对应的自定义参数配置;语音增强参数配置表中包括一组13阶低通滤波器的系数、平滑因子、增益以及增益变化的步长;单麦克风降噪参数配置表中包括两个常数项和两个平滑因子;双麦克风降噪参数配置表中包括一个侧向的增益因子和后向三个不同角度的增益因子;

[0012] 所述声音精准优化模块包括手机应用模块、云端服务器和用户定制服务单元,其中,手机应用模块包括用户自定义引导单元、服务请求单元和服务反馈单元,配置手机应用模块首先通过手机应用模块控制耳蜗采集环境声并通过音频信号特征分析模块提取其特征值,并将特征值和音频发送给云端服务器等待处理,然后通过手机应用模块递交服务请求,云端服务器收到请求后,启动用户定制服务单元,对上传的数据进行分析和处理,训练自定义环境检测单元的参数,并调整与之匹配的声音处理模块的自定义参数配置,云端服务器将训练和调试好的参数,下载到手机,再通过手机应用模块的服务反馈单元将参数部署到耳蜗中;所述用户自定义引导单元通过音频信号特征分析模块将所需自定义的声音环

境的音频信号特征进行提取,并将特征值发送到云端服务器进行存储;

[0013] 所述服务请求单元将系统信息及用户信息提供给云端服务器并发送给用户定制服务单元;

[0014] 所述云端服务器存储和管理用户数据和信息;

[0015] 所述用户定制服务单元根据用户递交的请求对数据进行处理,然后通过数据训练自定义环境检测单元的参数,并调整自定义参数配置,待完成参数的训练和调试后,通过云端服务器下载到用户的手机端;

[0016] 所述服务反馈单元在接收到云端服务器传来的配置文件后,将配置部署到声音处理模块中。

[0017] 优选地,所述特征集合包括谱质心、谱熵、不同子频段划分的子带能量率、单帧谱通量、多帧谱通量和频谱能量互相关。

[0018] 优选地,所述音频成分检测模块包括至少10个预设检测模块和4个自定义环境检测单元,10个预设检测模块包含言语成分检测单元、耳语成分检测单元、安静环境检测单元、车内噪声环境检测单元、空调环境检测单元、言语噪声环境检测单元、街道环境检测单元、雨噪环境检测单元、风噪环境检测单元和强噪声环境检测单元,每个检测单元均包括基于二分类神经网络模型的分分类器,模型参数通过包含相应声音的音频数据库训练得出,该音频数据库包括言语、耳语、车内噪声、空调噪声、言语噪声、街道噪声、雨噪、风噪、强噪,每个二分类器均采用一对多的分类模式,即在训练言语成分检测模块的分分类器时,将数据库中的全部音频划分为言语声和非言语声,如是训练耳语成分检测单元的分分类器时,将数据库中的全部音频划分为耳语声和非耳语声,以此类推,在均衡了两类数据样本之后,对各自的分类器进行训练;自定义检测单元用于检测用户自定义的声音或环境类别,所有检测单元的输出结果组成一个一维的向量,即环境向量,提供给音频成分综合分析决策模块,环境向量中的每一个元素是归一后从0到1之间的小数,值越大代表音频中包含该元素对应的声音类别的可能性越大。

[0019] 基于上述目的,本发明还提供了一种采用上述系统的带自定义的智能耳蜗声音处理系统的单麦克风降噪方法,包括以下步骤:

[0020] 所述单麦克风降噪模块进行噪声估计和噪声消除,首先使用改进的最小值跟踪对噪声能量进行估计,然后根据估计的SNR计算相应的增益值,最后再对噪声进行消除。

[0021] 优选地,所述噪声估计为,设输入信号 $y(n)$ 由语音信号 $x(n)$ 和噪音信号 $d(n)$ 组成:

$$[0022] \quad y(n) = x(n) + d(n) \quad (1)$$

[0023] 对上述输入信号进行短时傅里叶变换,变换后第 k 通道,第 λ 帧的功率谱记为 $Y(\lambda, k)$,平滑后可得:

$$[0024] \quad P(\lambda, k) = \eta P(\lambda - 1, k) + (1 + \eta) |Y(\lambda, k)|^2 \quad (2)$$

[0025] 其中, $P(\lambda, k)$ 表示平滑后的功率谱, η 表示平滑因子;

[0026] 对带噪信号的局部最小能量值 $P_{\min}(\lambda, k)$ 进行估计:

$$[0027] \quad P_{\min}(\lambda, k) = \begin{cases} P(\lambda, k), & P_{\min}(\lambda - 1, k) \geq P(\lambda, k) \\ \gamma P_{\min}(\lambda - 1, k) + \frac{1 - \gamma}{1 - \beta} (P(\lambda, k) - \beta P(\lambda - 1, k)), & P_{\min}(\lambda - 1, k) < P(\lambda, k) \end{cases} \quad (3)$$

[0028] 其中, β 和 γ 为两个常数,根据式(3)估计的局部最小功率谱来计算语音信号的出

现概率,首先,计算当前功率谱与局部最小功率谱的比值 $S_r(\lambda, k)$:

$$[0029] \quad S_r(\lambda, k) = \frac{P(\lambda, k)}{P_{min}(\lambda, k)} \quad (4)$$

[0030] 之后,引入一个阈值 $T(\lambda, k)$ 来确定当前帧是否包含语音信号:

$$[0031] \quad I(\lambda, k) = \begin{cases} 1, S_r(\lambda, k) \geq T(\lambda, k) \\ 0, S_r(\lambda, k) < T(\lambda, k) \end{cases} \quad (5)$$

[0032] 其中, $I(\lambda, k)$ 表示当前帧中是否包含语音信号的判断标识,阈值 $T(\lambda, k)$ 为动态阈值,根据上一帧所估计的SNR得出,SNR越高,阈值越低,反之亦然;

[0033] 对 $I(\lambda, k)$ 进行平滑后可得:

$$[0034] \quad K(\lambda, k) = \alpha K(\lambda, k) + (1-\alpha) I(\lambda, k) \quad (6)$$

[0035] 其中, $K(\lambda, k)$ 表示语音信号在当前帧出现的概率,由此结果来更新噪声估计的平滑因子:

$$[0036] \quad \alpha_s(\lambda, k) = \alpha_d + (1-\alpha_d) K(\lambda, k) \quad (7)$$

[0037] 其中, α_d 为一个常数,最终得到当前帧每个通道的噪声功率谱估计:

$$[0038] \quad D(\lambda, k) = \alpha_s(\lambda, k) D(\lambda-1, k) + (1-\alpha_s(\lambda, k)) |Y(\lambda, k)|^2 \quad (8)$$

[0039] 所述噪声消除为,根据上述估计的噪声功率谱计算信噪比SNR:

$$[0040] \quad SNR(\lambda, k) = \delta SNR(\lambda-1, k) + (1-\delta) |Y(\lambda, k)|^2 \quad (9)$$

[0041] 由此可得到一个增益系数:

$$[0042] \quad G(\lambda, k) = \frac{SNR(\lambda, k)}{SNR(\lambda, k) + 1} \quad (10)$$

[0043] 最终根据增益系数 $G(\lambda, k)$ 对信号进行降噪处理,得到降噪后的语音信号功率谱:

$$[0044] \quad S(\lambda, k) = G_0(\lambda, k) P(\lambda, k) \quad (11)$$

[0045] 其中, $G_0(\lambda, k)$ 表示最终的降噪增益系数。

[0046] 基于上述目的,本发明还提供了一种采用上述系统的带自定义的智能耳蜗声音处理系统的双麦克风降噪方法,所述双麦克风降噪模块利用不同方位的声源发出的声音,到达前后两个麦克风的时间延时不同的特点,反向判断目标声源以及噪声源的方向,再利用延时相加,以及差分麦克风阵列中分别控制目标声源及噪声源的增益方法,提升信噪比,降低噪声干扰。

[0047] 优选地,在语音信号进行频域信号处理时,对后麦克风音频采集模块的语音信号进行频域延时处理,频域信号延时处理采用离散傅里叶变换的时间位移定理,即频域的乘积相当于时域的延时 $s(t+\tau)$:

$$[0048] \quad s(t+\tau) = S(f) \cdot e^{-j\omega \frac{d}{c}} \quad (12)$$

[0049] 其中, $s(t+\tau)$ 表示延时 τ 后的信号, $S(f)$ 表示原信号的频谱, d 为前后两个麦克风间

的直线距离, c 为声音在空气中的传播速度, $\omega = 2\pi \frac{k}{N/2} * f_s$, N 为FFT点数, $k = 0 \sim N/2$,

f_s 为系统信号采样率。

[0050] 频域信号的延时相加利用离散傅里叶变换的线性性质实现,即频域两个信号的相加等价于时域两个信号的叠加,在频域中前麦克风音频采集模块采集的语音信号与进行延时处理后的后向信号进行叠加处理:

$$[0051] \quad s_{\text{sum}}(t) = s_F(t) + s_B(t+d/c) \quad (13)$$

[0052] 其中, $s_{\text{sum}}(t)$ 表示叠加处理后的信号, $s_F(t)$ 表示前置麦克风接收信号, $s_B(t+d/c)$ 表示经过时延矫正的后置麦克风接收信号,在进行完信号的延时相加处理后,对其进行相位补偿,其参考相位以前麦克风音频采集模块采集的语言信号的相位信息为标准;

[0053] 使用差分麦克风阵列是为了获取两个背靠背的单极性心形麦克风信号输出,并通过分别控制前向增益系数,如前向增益为1,保留正前方信号;和后向增益系数 β_0 ,即衰减后向信号强度的大小,以获取所需要的心形输出,差分麦克风阵列由下式实现:

$$[0054] \quad C_F(f) = \Phi \left(S_{\text{sum}}(f) - e^{-i\omega d/c} \cdot S_{B_delay}(f) \right)$$

$$[0055] \quad C_B(f) = \Phi \left(S_{\text{sum}}(f) - e^{i\omega d/c} \cdot S_{B_delay}(f) \right)$$

$$[0056] \quad \Phi = 1 / \left(1 - e^{2i\omega d/c} \right) \quad (14)$$

[0057] 其中, C_F 表示前向差分后的信号频谱, C_B 表示后向差分后的信号频谱, Φ 表示差分后的补偿滤波器, $S_{\text{sum}}(f)$ 表示 $s_{\text{sum}}(t)$ 的频谱, $S_{B_delay}(f)$ 表示 $s_B(t+d/c)$ 的频谱。在获取前向和后向背靠背的心形输出信号后,通过调整后向输出的增益系数,再以前向心形信号减去后向心形信号,以获得最终差分麦克风阵列输出。其实现方式如下式所示:

$$[0058] \quad s_{\text{out}}(t) = c_F(t) - \beta_0 c_B(t), \beta_0 \approx (1 + \cos \theta_n) / (1 - \cos \theta_n) \quad (15)$$

[0059] 其中, $s_{\text{out}}(t)$ 表示经过差分波束形成的信号, c_F 表示前向差分后的信号, c_B 表示后向差分后的信号, θ_n 表示抑制噪声的方向,如固定后向差分麦克风阵列输出信号的增益大小 β_0 ,则为定向消除某一方向的噪声信号,而保留正前方语音信号;如自动修改或选择相应的增益大小 β_0 ,则实现正前方语音信号保留,同时自适应的抑制除正前方目标语音信号外的其他方向的噪声信号,为了实现不同方向的噪声源选择相应的增益系数 β_0 的策略,则首先需判定出噪声源的位置,将某一方向的语音信号映射分布在半径相同的球面上,各个方向的延时时间不同,计算出语音信号分布在 0° 到 180° 方向的信号能量大小,以判断噪声源的位置;然后,根据噪声源所分别的区域选择相应的后向增益系数 β_0 ,以实现自适应追踪噪声源位置,抑制不同方向的噪声信号,由下式:

$$[0060] \quad \text{Energy}(\theta) = \sum_{k=0}^{k=N/2} \left(S_F(k) + e^{-i\omega d \cos \theta / c} S_B(k) \right)^2 \quad (16)$$

[0061] 其中, θ 为语音信号映射方位角, $\text{Energy}(\theta)$ 表示 θ 方向信号能量强度, S_F 表示前置

麦克风接收到的信号频谱, S_b 表示后置麦克风接收到的信号频谱。在获得当前帧语音信号在 0° 到 180° 范围内且角度分辨率为 p 的各个角度的能量分布后, 通过比较大小选择出能量最大的位置 m_{\max} , 即可判断出当前信号的噪声源位置, 由于此时获得帧与帧间的 m_{\max} 值变化很迅速, 需对其进行平滑处理, 减缓其变化速度, 进一步准确判断声源方位, 并能稳定的维持一定的时间, 其平滑公式:

$$[0062] \quad E_{\max}(k) = \alpha E_{\max}(k-1) + (1-\alpha) m_{\max} \quad (17)$$

[0063] 其中, α 为平滑因子, k 为信号分帧处理的帧序数, E_{\max} 为当前帧经平滑处理后的最大能量角度;

[0064] 前述的后向增益因子 β_0 , 当其参数固定时, 则为定向抑制某一方向的噪声信号, 保留正前方的目标语音; 当其参数可变时, 则为自适应追踪噪声源位置, 并抑制其噪声信号而保留正前方的目标语音。

[0065] 基于上述目的, 本发明还提供了一种采用上述系统的带自定义的智能耳蜗声音处理系统的语音增强方法, 所述语音增强模块提高汉语声调感知, 包括以下步骤:

[0066] 首先, 将声音信号分为 M 个通道进行包络提取; 然后提取声音信号的基频信息, 并根据基频信息对每个通道的包络进行调整, 模拟音调的变化; 最后, 耳蜗系统根据包络调整后的信号对各通道进行刺激。

[0067] 本发明的有益效果在于: 不仅包含了智能化的多种声环境声检测和与之对应的处理模块, 还可以可通过用户定制服务, 将用户在日常生活中所处的常用环境添加到该系统中。系统在运行期间, 当用户处在该环境时, 便可自动识别并加以处理, 提高用户的声音感知能力, 达到精准服务的目的。用户在添加自定义环境时, 通过手机APP开启系统的引导程序, 引导程序会引导用户录制几段当前的环境声, 通过蓝牙发送给手机; 手机应用模块会将音频上传到服务器; 服务请求在云端服务器登记后, 工程师利用这些录制好的音频及其扩展对模型进行训练, 然后训练好的模型参数将通过用户的手机应用模块下载到声音优化系统中, 从而完成了环境声及其处理的自定义。本系统将用户自定义模块与系统默认模块进行了整合, 采用配置与模型隔离的方式, 对各个模块进行统一管理。

附图说明

[0068] 为了使本发明的目的、技术方案和有益效果更加清楚, 本发明提供如下附图进行说明:

[0069] 图1为本发明实施例的一种带自定义的智能耳蜗声音处理系统结构示意图;

[0070] 图2为本发明一具体实施例的一种带自定义的智能耳蜗声音处理系统结构示意图;

[0071] 图3为本发明实施例的一种带自定义的智能耳蜗声音处理系统的双麦克风降噪方法的调整增益系数来抑制不同角度声源示意图。

具体实施方式

[0072] 下面将结合附图, 对本发明的优选实施例进行详细的描述。

[0073] 参见图1、图2, 所示为本发明实施例的一种带自定义的智能耳蜗声音处理系统, 包括智能控制模块10、声音处理模块20、声音精准优化模块30、前麦克风音频采集模块41和后

麦克风音频采集模块42,其中,智能控制模块10与声音处理模块20、声音精准优化模块30、前麦克风音频采集模块41和后麦克风音频采集模块42分别连接,声音处理模块20还与声音精准优化模块30连接,前麦克风音频采集模块41和后麦克风音频采集模块42还与声音处理模块20连接,智能控制模块10对环境声进行监听和识别,根据识别的结果,实时调整声音处理模块20的参数配置;声音处理模块20对声音信号进行处理,包括降噪和语音增强;声音精准优化模块30提供精准的个性化处理及优化,可生成自定义的声音处理策略;前麦克风音频采集模块41和后麦克风音频采集模块42分别设置在智能耳蜗体外机的前侧和后侧,对外界声音进行采集。

[0074] 智能控制模块10包括音频预处理模块11、音频成分检测模块12和音频成分综合分析决策模块13,其中,音频预处理模块11包括能量检测模块111和音频信号特征分析模块112,能量检测模块111通过监测能量的方式,从声音信号流里检测静音期,如处于静音期,则音频成分检测模块12将其判定为安静环境,同时避开音频信号特征分析模块112的处理和音频成分检测模块12的判定,直接传送给音频成分综合分析模块进行统计分析;如非静音期,则将音频信号送给音频信号特征分析模块112,将信号的物理特征提取出来并进行统计分析,形成特征集合,将该特征集合提供给音频成分检测模块12,计算上述特征值时,以4ms为一个帧,统计计算0.5s即125帧特征值的均值和方差;

[0075] 音频成分检测模块12对特征集合进行分析,识别出该特征集合表征的音频信号包含的成分或者属于的环境,将分类结果传递给音频成分综合分析决策模块13;

[0076] 音频成分综合分析决策模块13对环境向量进行统计分析,调控声音处理模块20的处理策略,包括状态反馈模块131、统计分析模块132和决策输出模块133;状态反馈模块131将环境向量与一个状态转移矩阵相乘,得到一个通过先验概率加权评估后的环境向量,传递给统计分析模块132,状态转移矩阵里的每一个元素代表了一类声音转换到另一类声音的概率,其值越大,代表可能性越高;统计分析模块132对一段时间内的环境向量进行统计,将统计结果超过一定阈值的最多两种声音类别作音频成分的最终决定,发送给决策输出模块133,上述一段时间为6-15s;决策输出模块133根据统计分析模块132的对音频成分的最终判定,控制声音处理模块20的处理策略,包括开启声音处理算法以及选择算法的参数配置;

[0077] 声音处理模块20在智能控制模块10的调控下对声音进行降噪和语音增强的处理,包括双麦克风降噪模块21、单麦克风降噪模块22、语音增强模块23和参数配置模块24,声音信号被前麦克风音频采集模块41和后麦克风音频采集模块42采集之后,依次经过双麦克风降噪、单麦克风降噪和语音增强的处理;

[0078] 单麦克风降噪模块22通过噪声估计和噪声消除,提高语音信号的信噪比;

[0079] 双麦克风降噪模块21处理不同方位的声源发出的声音,到达一前一后两个麦克风的时间延时不同,反向判断目标声源以及噪声源的方向,进行延时相加,以及差分麦克风阵列中分别控制目标声源及噪声源的增益;

[0080] 语音增强模块23通过改变时域输出的幅度,拟合汉语4声调的基频,提高用户对汉语声调的感知能力;

[0081] 参数配置模块24包括双麦克风降噪参数配置表241、单麦克风降噪参数配置表242和语音增强参数配置表243,每个配置表均包括一组默认参数配置作为预设值、一组可选参

数配置作为备用选项和4组自定义参数配置,其中可选参数配置和自定义参数配置根据用户需求可个性化设置,可选参数配置在默认参数的基础上,根据实际听声感受进行微调;自定义参数配置为匹配自定义环境检测而设置的模块,当检测到自定义的声音环境时,使用与之对应的自定义参数配置;语音增强参数配置表243中包括一组13阶低通滤波器的系数、平滑因子、增益以及增益变化的步长;单麦克风降噪参数配置表242中包括两个常数项和两个平滑因子;双麦克风降噪参数配置表241中包括一个侧向的增益因子和后向三个不同角度的增益因子;

[0082] 声音精准优化模块30包括手机应用模块31、云端服务器32和用户定制服务单元33,其中,手机应用模块31包括用户自定义引导单元311、服务请求单元312和服务反馈单元313,配置手机应用模块31首先通过手机应用模块31控制耳蜗采集环境声并通过音频信号特征分析模块112提取其特征值,并将特征值和音频发送给云端服务器32等待处理,然后通过手机应用模块31递交服务请求,云端服务器32收到请求后,启动用户定制服务单元33,对上传的数据进行分析和处理,训练自定义环境检测单元1291的参数,并调整与之匹配的声音处理模块20的自定义参数配置,云端服务器32将训练和调试好的参数,下载到手机,再通过手机应用模块31的服务反馈单元313将参数部署到耳蜗中;用户自定义引导单元311通过音频信号特征分析模块112将所需自定义的声音环境的音频信号特征进行提取,并将特征值发送到云端服务器32进行存储;

[0083] 服务请求单元312将系统信息及用户信息提供给云端服务器32并发送给用户定制服务单元33;

[0084] 云端服务器32存储和管理用户数据和信息;

[0085] 用户定制服务单元33根据用户递交的请求对数据进行处理,然后通过数据训练自定义环境检测单元1291的参数,并调整自定义参数配置,待完成参数的训练和调试后,通过云端服务器32下传到用户的手机端;

[0086] 服务反馈单元313在接收到云端服务器32传来的配置文件后,将配置部署到声音处理模块20中。

[0087] 特征集合包括谱质心、谱熵、不同子频段划分的子带能量率、单帧谱通量、多帧谱通量和频谱能量互相关。

[0088] 音频成分检测模块12包括至少10个预设检测模块和4个自定义环境检测单元1291,10个预设检测模块包含言语成分检测单元121、耳语成分检测单元122、安静环境检测单元123、车内噪声环境检测单元124、空调环境检测单元125单元、言语噪声环境检测单元126、街道环境检测单元127、雨噪环境检测单元128、风噪环境检测单元129和强噪声环境检测单元1290,每个检测单元均包括基于二分类神经网络模型的分器,模型参数通过包含相应声音的音频数据库训练得出,该音频数据库包括言语、耳语、车内噪声、空调噪声、言语噪声、街道噪声、雨噪、风噪、强噪,每个二分类器均采用一对多的分类模式,即在训练言语成分检测模块的分器时,将数据库中的全部音频划分为言语声和非言语声,如是训练耳语成分检测单元122的分器时,将数据库中的全部音频划分为耳语声和非耳语声,以此类推,在均衡了两类数据样本之后,对各自的分类器进行训练;自定义检测单元用于检测用户自定义的声音或环境类别,所有检测单元的输出结果组成一个一维的向量,即环境向量,提供给音频成分综合分析决策模块13,环境向量中的每一个元素是归一后从0到1之间的小

数,值越大代表音频中包含该元素对应的声音类别的可能性越大。

[0089] 采用上述系统的带自定义的智能耳蜗声音处理系统的单麦克风降噪方法,包括以下步骤:

[0090] 所述单麦克风降噪模块进行噪声估计和噪声消除,首先使用改进的最小值跟踪对噪声能量进行估计,然后根据估计的SNR计算相应的增益值,最后再对噪声进行消除。

[0091] 噪声估计为,设输入信号 $y(n)$ 由语音信号 $x(n)$ 和噪音信号 $d(n)$ 组成:

$$[0092] \quad y(n) = x(n) + d(n) \quad (1)$$

[0093] 对上述输入信号进行短时傅里叶变换,变换后第 k 通道,第 λ 帧的功率谱记为 $Y(\lambda, k)$,平滑后可得:

$$[0094] \quad P(\lambda, k) = \eta P(\lambda - 1, k) + (1 - \eta) |Y(\lambda, k)|^2 \quad (2)$$

[0095] 其中, $P(\lambda, k)$ 表示平滑后的功率谱, η 表示平滑因子;

[0096] 对带噪信号的局部最小能量值 $P_{\min}(\lambda, k)$ 进行估计:

$$[0097] \quad P_{\min}(\lambda, k) = \begin{cases} P(\lambda, k), P_{\min}(\lambda - 1, k) \geq P(\lambda, k) \\ \gamma P_{\min}(\lambda - 1, k) + \frac{1 - \gamma}{1 - \beta} (P(\lambda, k) - \beta P(\lambda - 1, k)), P_{\min}(\lambda - 1, k) < P(\lambda, k) \end{cases} \quad (3)$$

[0098] 其中, β 和 γ 为两个常数,根据式(3)估计的局部最小功率谱来计算语音信号的出现概率,首先,计算当前功率谱与局部最小功率谱的比值 $S_r(\lambda, k)$:

$$[0099] \quad S_r(\lambda, k) = \frac{P(\lambda, k)}{P_{\min}(\lambda, k)} \quad (4)$$

[0100] 之后,引入一个阈值 $T(\lambda, k)$ 来确定当前帧是否包含语音信号:

$$[0101] \quad I(\lambda, k) = \begin{cases} 1, S_r(\lambda, k) \geq T(\lambda, k) \\ 0, S_r(\lambda, k) < T(\lambda, k) \end{cases} \quad (5)$$

[0102] 其中, $I(\lambda, k)$ 表示当前帧中是否包含语音信号的判断标识,阈值 $T(\lambda, k)$ 为动态阈值,根据上一帧所估计的SNR得出,SNR越高,阈值越低,反之亦然;

[0103] 对 $I(\lambda, k)$ 进行平滑后可得:

$$[0104] \quad K(\lambda, k) = \alpha K(\lambda, k) + (1 - \alpha) I(\lambda, k) \quad (6)$$

[0105] 其中, $K(\lambda, k)$ 表示语音信号在当前帧出现的概率,由此结果来更新噪声估计的平滑因子:

$$[0106] \quad \alpha_s(\lambda, k) = \alpha_d + (1 - \alpha_d) K(\lambda, k) \quad (7)$$

[0107] 其中, α_d 为一个常数,最终得到当前帧每个通道的噪声功率谱估计:

$$[0108] \quad D(\lambda, k) = \alpha_s(\lambda, k) D(\lambda - 1, k) + (1 - \alpha_s(\lambda, k)) |Y(\lambda, k)|^2 \quad (8)$$

[0109] 所述噪声消除为,根据上述估计的噪声功率谱计算信噪比SNR:

$$[0110] \quad \text{SNR}(\lambda, k) = \delta \text{SNR}(\lambda - 1, k) + (1 - \delta) |Y(\lambda, k)|^2 \quad (9)$$

[0111] 由此可得到一个增益系数:

$$[0112] \quad G(\lambda, k) = \frac{SNR(\lambda, k)}{SNR(\lambda, k) + 1} \quad (10)$$

[0113] 最终根据增益系数 $G(\lambda, k)$ 对信号进行降噪处理,得到降噪后的语音信号功率谱:

$$[0114] \quad S(\lambda, k) = G_0(\lambda, k) P(\lambda, k) \quad (11)$$

[0115] 其中, $G_0(\lambda, k)$ 表示最终的降噪增益系数。

[0116] 一种采用上述系统的带自定义的智能耳蜗声音处理系统的双麦克风降噪方法,双麦克风降噪模块利用不同方位的声源发出的声音,到达前后两个麦克风的时间延时不同的特点,反向判断目标声源以及噪声源的方向,再利用延时相加,以及差分麦克风阵列中分别控制目标声源及噪声源的增益方法,提升信噪比,降低噪声干扰。

[0117] 在语音信号进行频域信号处理时,对后麦克风音频采集模块的语音信号进行频域延时处理,频域信号延时处理采用离散傅里叶变换的时间位移定理,即频域的乘积相当于时域的延时 $s(t+\tau)$:

$$[0118] \quad s(t+\tau) = S(f) \cdot e^{-j\omega \frac{d}{c}} \quad (12)$$

[0119] 其中, $s(t+\tau)$ 表示延时 τ 后的信号, $S(f)$ 表示原信号的频谱, d 为前后两个麦克风间的直线距离, c 为声音在空气中的传播速度, $\omega = 2\pi \frac{k}{N/2} * f_s$, N 为FFT点数, $k=0 \sim N/2$,

f_s 为系统信号采样率。

[0120] 频域信号的延时相加利用离散傅里叶变换的线性性质实现,即频域两个信号的相加等价于时域两个信号的叠加,在频域中前麦克风音频采集模块采集的语音信号与进行延时处理后的后向信号进行叠加处理:

$$[0121] \quad s_{sum}(t) = s_F(t) + s_B(t+d/c) \quad (13)$$

[0122] 其中, $s_{sum}(t)$ 表示叠加处理后的信号, $s_F(t)$ 表示前置麦克风接收信号, $s_B(t+d/c)$ 表示经过时延矫正的后置麦克风接收信号,在进行完信号的延时相加处理后,对其进行相位补偿,其参考相位以前麦克风音频采集模块采集的语言信号的相位信息为标准;

[0123] 使用差分麦克风阵列是为了获取两个背靠背的单极性心形麦克风信号输出,并通过分别控制前向增益系数,如前向增益为1,保留正前方信号;和后向增益系数 β_0 ,即衰减后向信号强度的大小,以获取所需要的心形输出,差分麦克风阵列由下式实现:

$$[0124] \quad C_F(f) = \Phi \left(S_{sum}(f) - e^{-i\omega d/c} \cdot S_{B_delay}(f) \right)$$

$$[0125] \quad C_B(f) = \Phi \left(S_{sum}(f) - e^{i\omega d/c} \cdot S_{B_delay}(f) \right)$$

$$[0126] \quad \Phi = 1 / \left(1 - e^{2i\omega d/c} \right) \quad (14)$$

[0127] 其中, C_F 表示前向差分后的信号频谱, C_B 表示后向差分后的信号频谱, Φ 表示差分

后的补偿滤波器, $S_{\text{sum}}(f)$ 表示 $s_{\text{sum}}(t)$ 的频谱, $S_{B,\text{delay}}(f)$ 表示 $s_B(t+d/c)$ 的频谱。在获取前向和后向背靠背的心形输出信号后, 通过调整后向输出的增益系数, 再以前向心形信号减去后向心形信号, 以获得最终差分麦克风阵列输出。其实现方式如下式所示:

$$[0128] \quad s_{\text{out}}(t) = c_F(t) - \beta_0 c_B(t), \beta_0 \approx (1 + \cos\theta_n) / (1 - \cos\theta_n) \quad (15)$$

[0129] 其中, $s_{\text{out}}(t)$ 表示经过差分波束形成的信号, c_F 表示前向差分后的信号, c_B 表示后向差分后的信号, θ_n 表示抑制噪声的方向, 如固定后向差分麦克风阵列输出信号的增益大小 β_0 , 则为定向消除某一方向的噪声信号, 而保留正前方语音信号; 如自动修改或选择相应的增益大小 β_0 , 则实现正前方语音信号保留, 同时自适应的抑制除正前方目标语音信号外的其他方向的噪声信号, 为了实现不同方向的噪声源选择相应的增益系数 β_0 的策略, 则首先需判定出噪声源的位置, 将某一方向的语音信号映射分布在半径相同的球面上, 各个方向的延时时间不同, 计算出语音信号分布在 0° 到 180° 方向的信号能量大小, 以判断噪声源的位置; 然后, 根据噪声源所分别的区域选择相应的后向增益系数 β_0 , 以实现自适应追踪噪声源位置, 抑制不同方向的噪声信号, 由下式:

$$[0130] \quad \text{Energy}(\theta) = \sum_{k=0}^{k-\frac{N}{2}} \left(S_F(k) + e^{-i\omega^d \cos\theta / c} S_B(k) \right)^2 \quad (16)$$

[0131] 其中, θ 为语音信号映射方位角, $\text{Energy}(\theta)$ 表示 θ 方向信号能量强度, S_F 表示前置麦克风接收到的信号频谱, S_B 表示后置麦克风接收到的信号频谱。在获得当前帧语音信号在 0° 到 180° 范围内且角度分辨率为 p 的各个角度的能量分布后, 通过比较大小选择出能量最大的位置 m_{max} , 即可判断出当前信号的噪声源位置, 由于此时获得帧与帧间的 m_{max} 值变化很迅速, 需对其进行平滑处理, 减缓其变化速度, 进一步准确判断声源方位, 并能稳定的维持一定的时间, 其平滑公式:

$$[0132] \quad E_{\text{max}}(k) = \alpha E_{\text{max}}(k-1) + (1-\alpha) m_{\text{max}} \quad (17)$$

[0133] 其中, α 为平滑因子, k 为信号分帧处理的帧序数, E_{max} 为当前帧经平滑处理后的最大能量角度;

[0134] 前述的后向增益因子 β_0 , 当其参数固定时, 则为定向抑制某一方向的噪声信号, 保留正前方的目标语音; 当其参数可变时, 则为自适应追踪噪声源位置, 并抑制其噪声信号而保留正前方的目标语音。为了实现稳定的声源位置判定, 以至于不会出现频繁的增益因子 β 的修改调节, 我们需对目标声和噪声方位进行分区域设置。

[0135] 由于人与人交流通常为面对面沟通交流, 故设置 0° 到 75° 范围内的语音信号为用户需要听清楚并增强保留的目标语音; 设置 75° 到 105° 范围内的语音为侧向噪音, 需选择其相应的增益因子 β_s , 以抑制侧向噪音; 分别设置如下三个后向噪音区域, 并分别设置其对应的增益因子: 105° 到 120° 范围内为第一后向噪音区域, 设置其增益因子为 β_{b1} 、 120° 到 140° 范围内为第二后向噪音区域, 设置其增益因子为 β_{b2} 、 140° 到 180° 范围内为第三后向噪音区域, 设置其增益因子为 β_{b3} 。对不同角度抑制后的能量分布参见图3, 面对 90° , 120° , 180° 噪声源时, 极坐标增益输出。

[0136] 采用上述系统的带自定义的智能耳蜗声音处理系统的语音增强方法, 语音增强模块提高汉语声调感知, 包括以下步骤:

[0137] 首先, 将声音信号分为 M 个通道进行包络提取; 然后提取声音信号的基频信息, 并

根据基频信息对每个通道的包络进行调整,模拟音调的变化;最后,耳蜗系统根据包络调整后的信号对各通道进行刺激。

[0138] 最后说明的是,以上优选实施例仅用以说明本发明的技术方案而非限制,尽管通过上述优选实施例已经对本发明进行了详细的描述,但本领域技术人员应当理解,可以在形式上和细节上对其作出各种各样的改变,而不偏离本发明权利要求书所限定的范围。

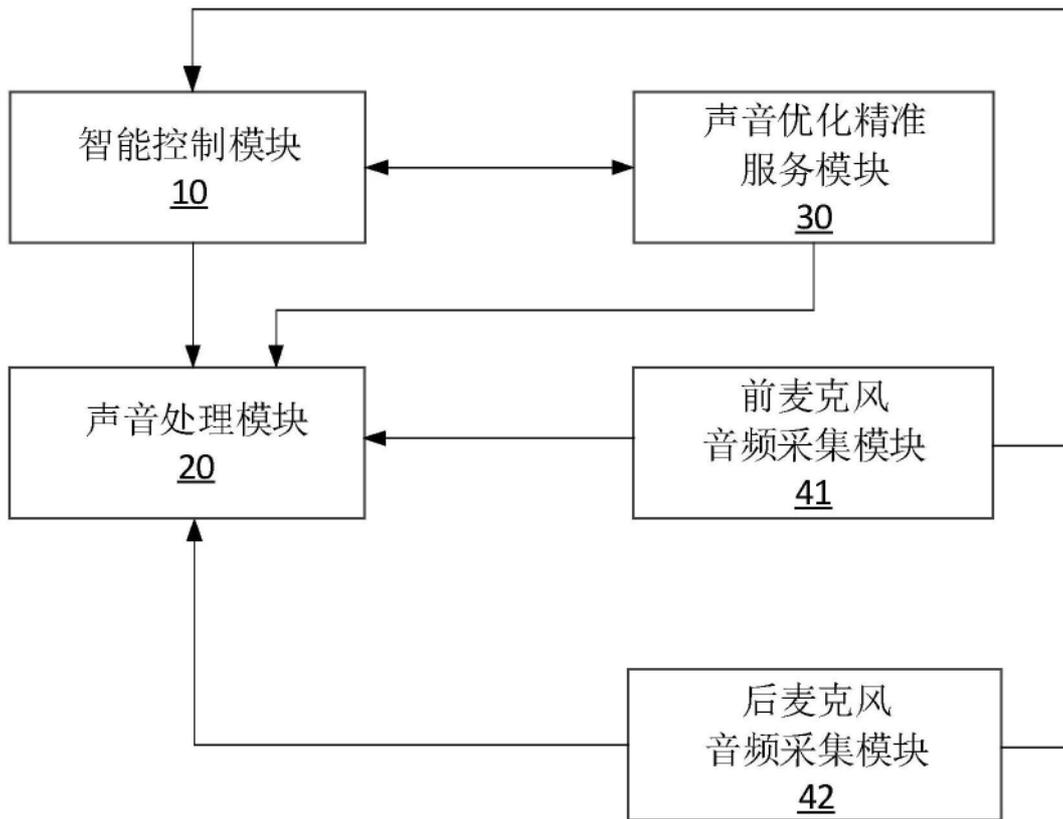


图1

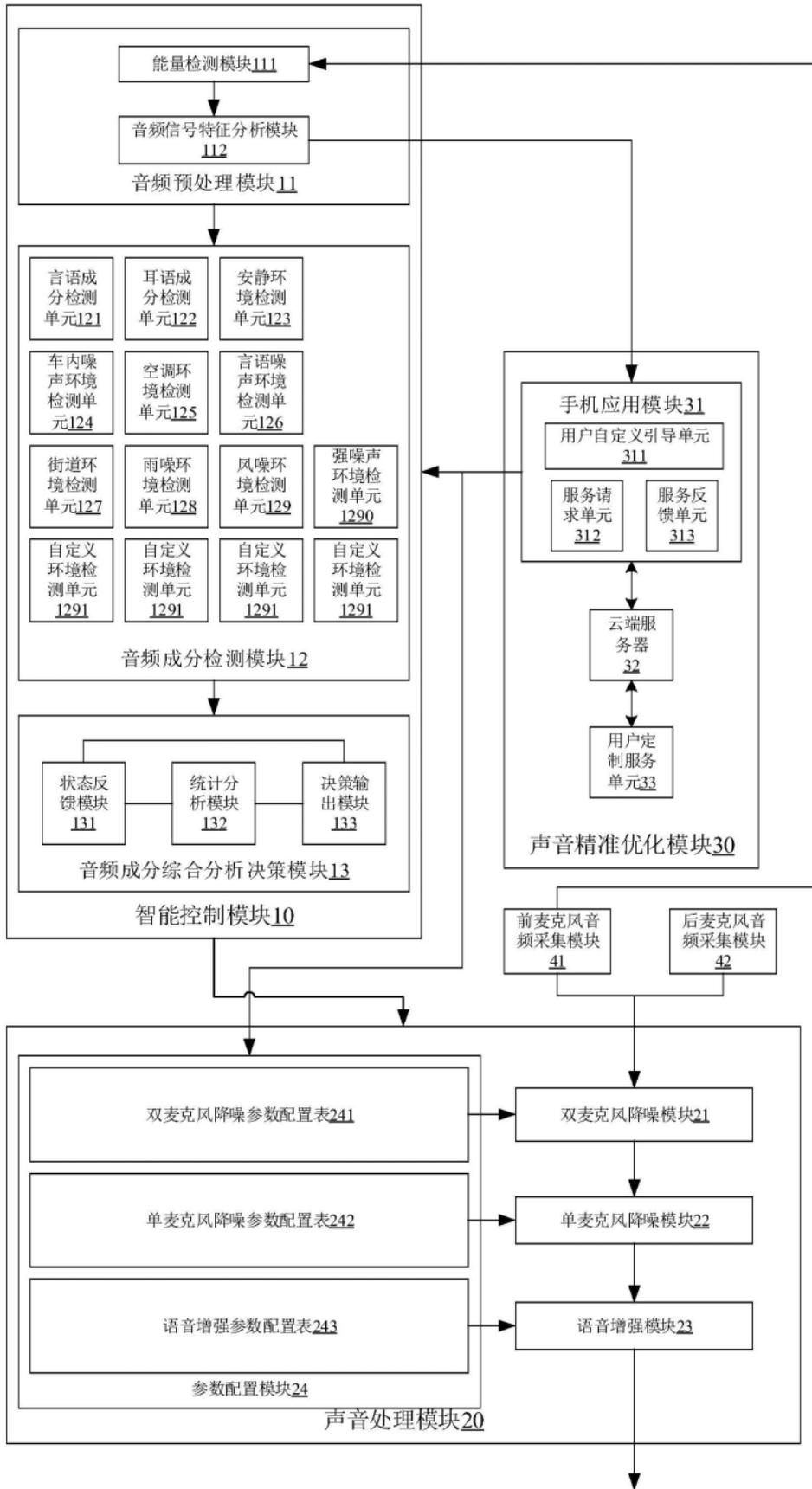


图2

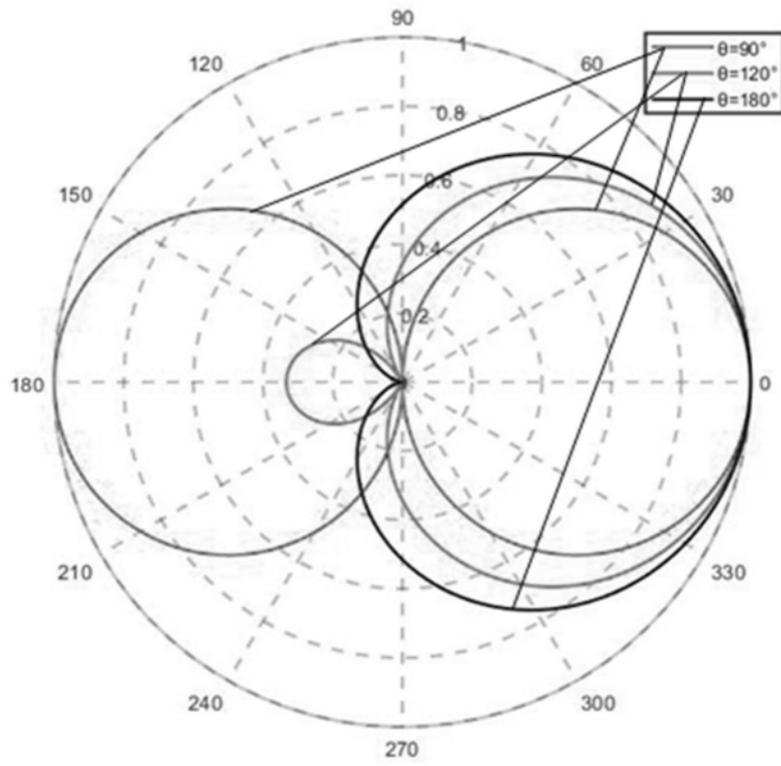


图3