



(12) 发明专利申请

(10) 申请公布号 CN 116312561 A

(43) 申请公布日 2023. 06. 23

(21) 申请号 202310297886.X

G10L 21/0208 (2013.01)

(22) 申请日 2023.03.23

G06F 21/32 (2013.01)

G10L 21/0272 (2013.01)

(71) 申请人 广东电网有限责任公司广州供电局  
地址 510000 广东省广州市天河区天河南二路2号

G10L 17/02 (2013.01)

G06Q 50/06 (2012.01)

(72) 发明人 崔兆阳 衷宇清 张雄威 凌健文  
徐武华 蒋盛智 彭丽文 周上  
罗慕尧 骆雅菲 刘晨辉 孔嘉麟  
陈文文 张思敏 周菲 吴若迪  
冯雅雯

(74) 专利代理机构 佛山市君创知识产权代理事务  
所(普通合伙) 44675  
专利代理师 陈誉宏

(51) Int. Cl.

G10L 17/04 (2013.01)

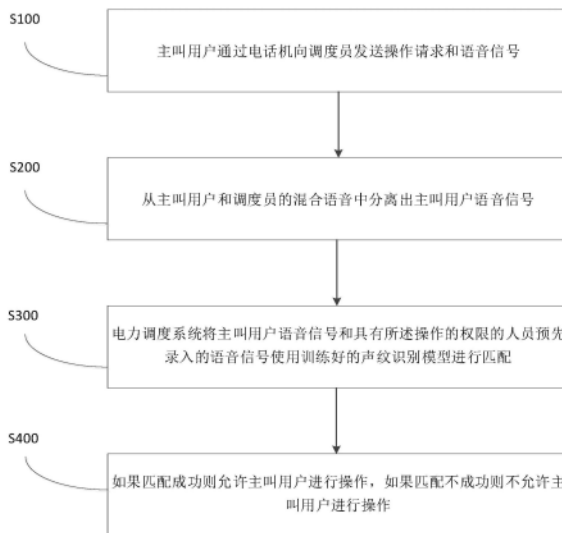
权利要求书4页 说明书11页 附图4页

(54) 发明名称

一种电力调度系统人员声纹识别鉴权降噪和语音增强方法、系统及装置

(57) 摘要

本发明提供一种电力调度系统人员声纹识别鉴权降噪和语音增强方法、系统及装置,该方法包括:主叫用户通过电话机向调度员发送操作请求和语音信号;从主叫用户和调度员的混合语音中分离出主叫用户语音信号;对主叫用户语音信号进行降噪;对主叫用户语音信号进行语音增强;电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;如果匹配成功则允许主叫用户进行操作,如果匹配不成功则不允许主叫用户进行操作。本发明能够在受到电流和噪声干扰的情况下准确识别用户语音。



1. 一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,包括:  
 主叫用户通过电话机向调度员发送操作请求和语音信号;  
 从主叫用户和调度员的混合语音中分离出主叫用户语音信号;  
 对主叫用户语音信号进行降噪;  
 对主叫用户语音信号进行语音增强;

电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;

如果匹配成功则允许主叫用户进行操作,如果匹配不成功则不允许主叫用户进行操作。

2. 根据权利要求1所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,所述从主叫用户和调度员的混合语音中分离出主叫用户语音信号包括:

从电话机端获取的第一语音信号;

在电力调度系统传输线路内部增加的消侧音电路,从电话机听筒端获取的第二语音信号;

使用短时过零率、端点检测和语音能量谱对第一语音信号和第二语音信号进行语音信号强度分析与信号比对,分离出主叫用户语音信号;

分离后获得四个受不同噪声影响的语音信号为:

$$y_{\text{主叫用户}}^{\text{电话机端}}(t) = s_{\text{主叫用户}}^{\text{电话机端}}(t) + n_{\text{主叫用户环境噪声+电话传输信道噪声}}(t)$$

$$y_{\text{调度员}}^{\text{电话机端}}(t) = s_{\text{调度员}}^{\text{电话机端}}(t) + n_{\text{电话机本身噪声}}(t)$$

$$y_{\text{主叫用户}}^{\text{电话机听筒端}}(t) = s_{\text{主叫用户}}^{\text{电话机听筒端}}(t) + n_{\text{主叫用户环境噪声+电话传输信道噪声}}(t) + n_{\text{电话机本身噪声}}$$

$$y_{\text{调度员}}^{\text{电话机听筒端}}(t) = s_{\text{主叫用户调度员}}^{\text{电话机听筒端}}(t) + n_{\text{低噪声影响}}(t)$$

3. 根据权利要求1所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,所述对主叫用户语音信号进行降噪包括:

采用相关特征法对主叫用户语音信号进行降噪:

假定主叫用户语音信号与主叫用户环境噪声、电话传输信道噪声互不相干,对带噪声信号做自相关处理,可得到与不带噪语音信号的自相关帧序列:

$$R_y(\tau) = \frac{1}{T} \int_{-h}^t y(t)y(t-\tau)w(t)dt \approx \frac{1}{T} \int_{-h}^t s(t)s(t-\tau)w(t)dt = R_s(\tau)$$

其中 $s(t)$ 为纯净语音信号, $n(t)$ 为噪声信号, $w(t)$ 是为达成短时平稳所加的窗函数, $R_y(\tau)$ 和 $R_s(\tau)$ 分别是主叫人语音信号带噪和不带噪的自相关函数;

采用维纳滤波法对语音信号进行降噪:

带噪语音信号经过维纳滤波器后的输出 $s'(t)$ 满足 $E[|s'(t) - s(t)|^2]$ 的值最小,维纳滤波法基于短时平稳语音信号为前提,对于维纳滤波器,得到下式:

$$H|\omega| = \frac{P_s(\omega)}{P_s(\omega) + P_n(\omega)}$$

上式中 $H|\omega|$ 为维纳滤波器频域的冲激响应, $P_s(\omega), P_n(\omega)$ 分为信号功率谱和噪声功率谱;

$$S_0(\omega) = H(\omega) \cdot Y(\omega)$$

上式中 $S_0(\omega)$ 为维纳滤波器的输出信号频谱, $Y(\omega)$ 为主叫人带噪电话语音信号频谱。

4. 根据权利要求1所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,所述对主叫用户语音信号进行语音增强包括:

倒谱均值规整降噪CMN方法清除带非加性噪声主叫人电话语音信号倒谱中的噪声成分,利用CMN方法处理后得到的增强语音倒谱表示为:

$$\hat{C}_s(t) = C_{sn}(t) - \bar{C}_{sn}(t) = C_s(t) + C_n(t) - \frac{1}{N} \sum_{t=1}^N C_s(t) - C_n(t) = C_s(t) - \frac{1}{N} \sum_{t=1}^N C_s(t)$$

其中 $\hat{C}_s(t)$ 为增强语音的倒谱, $C_{sn}(t)$ 为带噪语音的倒谱, $C_s(t)$ 为纯净语音的倒谱, $\bar{C}_{sn}(t)$ 为主叫人被采集语音段的倒谱平均值。

5. 根据权利要求2所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,所述使用短时过零率、端点检测和语音能量谱对第一语音信号和第二语音信号进行语音信号强度分析与信号比对,分离出主叫用户语音信号包括:

用短时能量和过零率检测结合的短时过零率检测算法来检测清音,用短时能量来检测浊音;

根据语音信号的浊音和清音选择对应的清音模型和浊音模型进行语音信号端点检测从而获取主叫用户语音信号;

所述根据语音信号的浊音和清音选择对应的清音模型和浊音模型进行语音信号端点检测从而获取主叫用户语音信号包括:

发清音时,对应的清音激励模型模拟成随机白噪声,使用均值为零,方差为1,并且在时间和幅度值上为白色分布的序列;

发浊音时,产生间歇的脉冲波,数学表达为:

$$g(n) = \begin{cases} (1/2)[1 - \cos(\pi n / N_1)], & 0 \leq n \leq N_1 \\ \cos[\pi(n - N_1) / 2N_2], & N_1 \leq n \leq N_1 + N_2 \\ 0, & \text{其它} \end{cases}$$

上式中, $N_1$ 为斜三角波上升部分的时间, $N_2$ 为其下降部分的时间;

语音信号进行分帧之后,第 $n$ 帧语音信号 $x_n(m)$ 的能量可表示为:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m)$$

短时过零率为一帧语音中语音信号波形穿过横轴即零电平的次数,可表示为:

$$Z_n = \frac{1}{2} \sum_{m=0}^{N-1} |\text{sgn}[x_n(m)] - \text{sgn}[x_n(m-1)]|$$

其中 $\text{sgn}()$ 为符号函数,通过考察当前采样信号与上一采样信号之间是否出现了波形上符号变化来评估过零次数;

对主叫用户语音信号进行能量谱估计:

语音信号进行分帧之后,第n帧语音信号 $x_n(m)$ 的能量表示为:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m)$$

采用自相关法提取主叫用户语音信号:

语音信号 $x_n(m)$ 的短时自相关函数 $R_n(k)$ 可表示为:

$$R_n(k) = \sum_{m=0}^{N-1-k} x_n(m)x_n(m+k) \quad (0 \leq k \leq K)$$

其中,K为最大的延迟点数,假设语音序列具有周期性,则自相关函数也是同周期的周期函数,对于浊音信号用自相关函数求出语音波形序列的基音周期,噪声信号和含噪语音的自相关函数在峰值幅度有较大差异,根据噪声的大小,设置阈值,确定其端点。

6. 根据权利要求1所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,所述声纹识别模型由卷积神经网络CNN和长短期记忆网络LSTM网络串联构建而成。

7. 根据权利要求1所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,所述电力调度系统将收到的用户语音信号和预先录入的声纹信息使用训练好的声纹识别模型进行匹配之前,还包括对声纹识别模型进行训练,具体为:

将预处理后的多个语音信号分为训练集和测试集;

将训练集输入声纹识别模型;

声纹识别模型输出语音信号的判断结果;

迭代训练声纹识别模型至错误率小于预设值。

8. 根据权利要求1所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,其特征在于,所述电力调度系统将收到的用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配包括:

对降噪后的主叫用户语音信号进行快速傅里叶变换,得到每个音源信号对应的频谱特征;

将频谱特征经梅尔滤波器滤波后取对数,得到主叫用户电话语音信号对应的梅尔频率对数能量谱;

将梅尔频率对数能量谱经离散余弦变换,得到主叫用户语音信号对应的梅尔系数谱;

基于对应的梅尔系数谱进行声纹识别处理,判别主叫用户身份并鉴权。

9. 一种电力调度系统人员声纹识别鉴权降噪和语音增强系统,其特征在于,包括:

接收模块,用于接收主叫用户通过电话机向调度员发送操作请求和语音信号;

第一数据处理模块,用于从主叫用户和调度员的混合语音中分离出主叫用户语音信号;

第二数据处理模块,用于电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;

结果输出模块,如果匹配成功则允许用户进行操作,如果匹配不成功则不允许用户进行操作。

10. 一种电力调度系统人员声纹识别鉴权降噪和语音增强装置,其特征在于,通过数据传送路径与电力调度系统人员声纹识别鉴权系统相连接,使所述电力调度系统人员声纹识

别鉴权装置执行权利要求1-8所述的一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,包括:

    数据获取单元,用于接收主叫用户通过电话机向调度员发送操作请求和语音信号;

    数据处理单元,用于从主叫用户和调度员的混合语音中分离出主叫用户语音信号;

    判断单元,用于电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;

    输出单元,如果匹配成功则允许用户进行操作,如果匹配不成功则不允许用户进行操作。

## 一种电力调度系统人员声纹识别鉴权降噪和语音增强方法、系统及装置

### 技术领域

[0001] 本发明涉及人工智能技术领域,特别涉及声纹识别技术领域,尤其涉及一种电力调度系统人员声纹识别鉴权降噪和语音增强方法、系统及装置。

### 背景技术

[0002] 在电力调度系统中,电话调度是常见基础形式。当通过电话接收到主叫方的调度指令时,对主叫方进行身份认证并鉴权是提高调度系统安全性和可靠性的一个核心问题。

[0003] 通过对做为主叫方的调度申请者的语音信号进行声纹识别是对其进行鉴权的一种可行方法。

[0004] 进行相应操作时,会涉及到电话语音的提取,语音信号的预处理,基于语音样本信号的深度神经网络训练和基于主叫方实际调度语音信号的判决与鉴权等一系列流程。

[0005] 其中,对电力调度系统人员声纹识别鉴权成功率和可靠性影响很大的一个因素是通过调度电话所提取到的语音信号的质量和干扰问题。

[0006] 当电力系统的调度员通过调度电话打入电话时,会不可避免地遇到工作环境内的噪声干扰与电话信道噪声干扰的问题。因此,如何有效地对这两种噪声进行抑制,并有针对性地进行语音增强,是提升系统性能的关键性问题。

### 发明内容

[0007] 本发明的目的是提供一种电力调度系统人员声纹识别鉴权降噪和语音增强方法、系统及装置,该方法能够在受到电流和噪声干扰的情况下准确识别用户语音。

[0008] 一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,包括:

[0009] 主叫用户通过电话机向调度员发送操作请求和语音信号;

[0010] 从主叫用户和调度员的混合语音中分离出主叫用户语音信号;

[0011] 对主叫用户语音信号进行降噪;

[0012] 对主叫用户语音信号进行语音增强;

[0013] 电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;

[0014] 如果匹配成功则允许主叫用户进行操作,如果匹配不成功则不允许主叫用户进行操作。

[0015] 从主叫用户和调度员的混合语音中分离出主叫用户语音信号包括:

[0016] 从电话机端获取的第一语音信号;

[0017] 在电力调度系统传输线路内部增加的消侧音电路,从电话机听筒端获取的第二语音信号;

[0018] 使用短时过零率、端点检测和语音能量谱对第一语音信号和第二语音信号进行语音信号强度分析与信号比对,分离出主叫用户语音信号;

[0019] 分离后获得四个受不同噪声影响的语音信号为:

$$[0020] \quad y_{\text{主叫用户}}^{\text{电话机端}}(t) = s_{\text{主叫用户}}^{\text{电话机端}}(t) + n_{\text{主叫用户环境噪声+电话传输信道噪声}}(t)$$

$$[0021] \quad y_{\text{调度员}}^{\text{电话机端}}(t) = s_{\text{调度员}}^{\text{电话机端}}(t) + n_{\text{电话机本身噪声}}(t)$$

$$[0022] \quad y_{\text{主叫用户}}^{\text{电话机听筒端}}(t) = s_{\text{主叫用户}}^{\text{电话机听筒端}}(t) + n_{\text{主叫用户环境噪声+电话传输信道噪声}}(t) + n_{\text{电话机本身噪声}}$$

$$[0023] \quad y_{\text{调度员}}^{\text{电话机听筒端}}(t) = s_{\text{主叫用户调度员}}^{\text{电话机听筒端}}(t) + n_{\text{低噪声影响}}(t)$$

[0024] 对主叫用户语音信号进行降噪包括:

[0025] 采用相关特征法对主叫用户语音信号进行降噪:

[0026] 假定主叫用户语音信号与主叫用户环境噪声、电话传输信道噪声互不相干,对带噪信号做自相关处理,可得到与不带噪语音信号的自相关帧序列:

$$[0027] \quad R_y(\tau) = \frac{1}{T} \int_{-h}^t y(t)y(t-\tau)w(t)dt \approx \frac{1}{T} \int_{-h}^t s(t)s(t-\tau)w(t)dt = R_s(\tau)$$

[0028] 其中 $s(t)$ 为纯净语音信号, $n(t)$ 为噪声信号, $w(t)$ 是为达成短时平稳所加的窗函数, $R_y(\tau)$ 和 $R_s(\tau)$ 分别是主叫人语音信号带噪和不带噪的自相关函数;

[0029] 采用维纳滤波法对语音信号进行降噪:

[0030] 带噪语音信号经过维纳滤波器后的输出 $s'(t)$ 满足 $E[|s'(t) - s(t)|^2]$ 的值最小,维纳滤波法基于短时平稳语音信号为前提,对于维纳滤波器,得到下式:

$$[0031] \quad H|\omega| = \frac{P_s(\omega)}{P_s(\omega) + P_n(\omega)}$$

[0032] 上式中 $H|\omega|$ 为维纳滤波器频域的冲激响应, $P_s(\omega)$ , $P_n(\omega)$ 分为信号功率谱和噪声功率谱;

$$[0033] \quad S_0(\omega) = H(\omega) \cdot Y(\omega)$$

[0034] 上式中 $S_0(\omega)$ 为维纳滤波器的输出信号频谱, $Y(\omega)$ 为主叫人带噪电话语音信号频谱。

[0035] 对主叫用户语音信号进行语音增强包括:

[0036] 倒谱均值规整降噪CMN方法清除带非加性噪声主叫人电话语音信号倒谱中的噪声成分,利用CMN方法处理后得到的增强语音倒谱表示为:

$$[0037] \quad \hat{C}_s(t) = C_{sn}(t) - \bar{C}_{sn}(t) = C_s(t) + C_n(t) - \frac{1}{N} \sum_{i=1}^N C_s(t) - C_n(t) = C_s(t) - \frac{1}{N} \sum_{i=1}^N C_s(t)$$

[0038] 其中 $\hat{C}_s(t)$ 为增强语音的倒谱, $C_{sn}(t)$ 为带噪语音的倒谱, $C_s(t)$ 为纯净语音的倒谱, $\bar{C}_{sn}(t)$ 为主叫人被采集语音段的倒谱平均值。

[0039] 使用短时过零率、端点检测和语音能量谱对第一语音信号和第二语音信号进行语音信号强度分析与信号比对,分离出主叫用户语音信号包括:

[0040] 用短时能量和过零率检测结合的短时过零率检测算法来检测清音,用短时能量来检测浊音;

[0041] 根据语音信号的浊音和清音选择对应的清音模型和浊音模型进行语音信号端点检测从而获取主叫用户语音信号;

[0042] 所述根据语音信号的浊音和清音选择对应的清音模型和浊音模型进行语音信号端点检测从而获取主叫用户语音信号包括:

[0043] 发清音时,对应的清音激励模型模拟成随机白噪声,使用均值为零,方差为1,并且在时间和幅度值上为白色分布的序列;

[0044] 发浊音时,产生间歇的脉冲波,数学表达为:

$$[0045] \quad g(n) = \begin{cases} (1/2)[1 - \cos(\pi n / N_1)], & 0 \leq n \leq N_1 \\ \cos[\pi(n - N_1) / 2N_2], & N_1 \leq n \leq N_1 + N_2 \\ 0, & \text{其它} \end{cases}$$

[0046] 上式中,N1为斜三角波上升部分的时间,N2为其下降部分的时间;

[0047] 语音信号进行分帧之后,第n帧语音信号 $x_n(m)$ 的能量可表示为:

$$[0048] \quad E_n = \sum_{m=0}^{N-1} x_n^2(m)$$

[0049] 短时过零率为一帧语音中语音信号波形穿过横轴即零电平的次数,可表示为:

$$[0050] \quad Z_n = \frac{1}{2} \sum_{m=0}^{N-1} |\text{sgn}[x_n(m)] - \text{sgn}[x_n(m-1)]|$$

[0051] 其中 $\text{sgn}()$ 为符号函数,通过考察当前采样信号与上一采样信号之间是否出现了波形上符号变化来评估过零次数;

[0052] 对主叫用户语音信号进行能量谱估计:

[0053] 语音信号进行分帧之后,第n帧语音信号 $x_n(m)$ 的能量表示为:

$$[0054] \quad E_n = \sum_{m=0}^{N-1} x_n^2(m)$$

[0055] 采用自相关法提取主叫用户语音信号:

[0056] 语音信号 $x_n(m)$ 的短时自相关函数 $R_n(k)$ 可表示为:

$$[0057] \quad R_n(k) = \sum_{m=0}^{N-1-k} x_n(m)x_n(m+k) \quad (0 \leq k \leq K)$$

[0058] 其中,K为最大的延迟点数,假设语音序列具有周期性,则自相关函数也是同周期的周期函数,对于浊音信号用自相关函数求出语音波形序列的基音周期,噪声信号和含噪语音的自相关函数在峰值幅度有较大差异,根据噪声的大小,设置阈值,确定其端点。

[0059] 声纹识别模型由卷积神经网络CNN和长短期记忆网络LSTM网络串联构建而成。

[0060] 电力调度系统将收到的用户语音信号和预先录入的声纹信息使用训练好的声纹识别模型进行匹配之前,还包括对声纹识别模型进行训练,具体为:

[0061] 将预处理后的多个语音信号分为训练集和测试集;

[0062] 将训练集输入声纹识别模型;

[0063] 声纹识别模型输出语音信号的判断结果;

[0064] 迭代训练声纹识别模型至错误率小于预设值。

[0065] 电力调度系统将收到的用户语音信号和具有所述操作的权限的人员预先录入的

语音信号使用训练好的声纹识别模型进行匹配包括：

[0066] 对降噪后的主叫用户语音信号进行快速傅里叶变换,得到每个音源信号对应的频谱特征;

[0067] 将频谱特征经梅尔滤波器滤波后取对数,得到主叫用户电话语音信号对应的梅尔频率对数能量谱;

[0068] 将梅尔频率对数能量谱经离散余弦变换,得到主叫用户语音信号对应的梅尔系数谱;

[0069] 基于对应的梅尔系数谱进行声纹识别处理,判别主叫用户身份并鉴权。

[0070] 一种电力调度系统人员声纹识别鉴权降噪和语音增强系统,包括:

[0071] 接收模块,用于接收主叫用户通过电话机向调度员发送操作请求和语音信号;

[0072] 第一数据处理模块,用于从主叫用户和调度员的混合语音中分离出主叫用户语音信号;

[0073] 第二数据处理模块,用于电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;

[0074] 结果输出模块,如果匹配成功则允许用户进行操作,如果匹配不成功则不允许用户进行操作。

[0075] 一种电力调度系统人员声纹识别鉴权降噪和语音增强装置,通过数据传送路径与电力调度系统人员声纹识别鉴权系统相连接,使所述电力调度系统人员声纹识别鉴权装置执行一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,包括:

[0076] 数据获取单元,用于接收主叫用户通过电话机向调度员发送操作请求和语音信号;

[0077] 数据处理单元,用于从主叫用户和调度员的混合语音中分离出主叫用户语音信号;

[0078] 判断单元,用于电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;

[0079] 输出单元,如果匹配成功则允许用户进行操作,如果匹配不成功则不允许用户进行操作。

[0080] 本发明通过主叫用户通过电话机向调度员发送操作请求和语音信号;从主叫用户和调度员的混合语音中分离出主叫用户语音信号;对主叫用户语音信号进行降噪;对主叫用户语音信号进行语音增强;电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;如果匹配成功则允许主叫用户进行操作,如果匹配不成功则不允许主叫用户进行操作。能够同时从调度电话的输入端和话筒端进行电话语音信号提取,通过电话输入端和话筒端的语音对比,剔除不属于主叫人的语音,提高了对用户语音信号提纯精度,处理后的用户语音信号能够让声纹识别模型对用户声音信息的判断更加准确,减少了调度员的工作,提高了调度效率。

## 附图说明

[0081] 此处的附图被并入说明书中并构成本说明书的一部分,标示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。

[0082] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,对于本领域普通技术人员而言,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0083] 图1为本发明的流程图;

[0084] 图2为本发明的获取纯净用户语音信号流程图;

[0085] 图3为本发明的声纹识别模型训练流程图;

[0086] 图4为本发明的声纹识别模型运行流程图。

## 具体实施方式

[0087] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明的一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0088] 需要说明,本发明实施例中所有方向性指示(诸如上、下、左、右、前、后……)仅用于解释在某一特定姿态(如附图所示)下各部件之间的相对位置关系、运动情况等,如果该特定姿态发生改变时,则该方向性指示也相应地随之改变。

[0089] 另外,在本发明中涉及“第一”、“第二”等的描述仅用于描述目的,而不能理解为指示或暗示其相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一种该特征。另外,各个实施例之间的技术方案可以相互结合,但是必须是以本领域普通技术人员能够实现为基础,当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在,也不在本发明要求的保护范围之内。

[0090] 调度语音作为调度员下达口令最直接的方式,也是调度信息传递最常用的载体,在人工智能水平越来越高的今天,更加需要智能化的调度语音处理平台,识别、分析和诊断各类调度语音信息,辅助调度员做出最及时的反应、最精确的判断和最高效的分析。时频分析法是声信号处理领域的一种常用手段。但运行中的调度员的声信号不可避免地会受到电流、噪声干扰等的影响,使得不同时间监测到的声波信号亦会随之改变且呈现宽带非平稳特征,其时频特性表现出一定复杂性,难以直接进行分析以区分调度员的不同工作状态。如何提高调度员工作状态识别的准确性是亟待解决的问题。

[0091] 目前基于神经网络的语音识别方法容易受到外界环境噪声和其它人声的干扰而导致识别结果不准确,本发明可以排除外界环境噪声和其它人声的干扰获取到纯净的目标人声语音信号,提高了声纹识别模型的识别准确率,并且单一的卷积神经网络模型提取的特征单一,识别结果不够准确,本发明采用卷积神经网络和长短期记忆网络相结合构成声纹识别模型,大大提高了声音识别正确率。

[0092] 实施例1

[0093] 一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,包括:

[0094] S100,主叫用户通过电话机向调度员发送操作请求和语音信号;

[0095] S200,从主叫用户和调度员的混合语音中分离出主叫用户语音信号;

[0096] S300,对主叫用户语音信号进行降噪;

[0097] S400,对主叫用户语音信号进行语音增强;

[0098] S500,电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配;

[0099] S600,如果匹配成功则允许主叫用户进行操作,如果匹配不成功则不允许主叫用户进行操作。

[0100] S200从主叫用户和调度员的混合语音中分离出主叫用户语音信号包括:

[0101] S201,从电话机端获取的第一语音信号;

[0102] S202,在电力调度系统传输线路内部增加的消侧音电路,从电话机听筒端获取的第二语音信号;

[0103] S203,使用短时过零率、端点检测和语音能量谱对第一语音信号和第二语音信号进行语音信号强度分析与信号比对,分离出主叫用户语音信号;

[0104] 分离后获得四个受不同噪声影响的语音信号为:

$$[0105] \quad y_{\text{主叫用户}}^{\text{电话机端}}(t) = s_{\text{主叫用户}}^{\text{电话机端}}(t) + n_{\text{主叫用户环境噪声+电话传输信道噪声}}(t)$$

$$[0106] \quad y_{\text{调度员}}^{\text{电话机端}}(t) = s_{\text{调度员}}^{\text{电话机端}}(t) + n_{\text{电话机本身噪声}}(t)$$

$$[0107] \quad y_{\text{主叫用户}}^{\text{电话机听筒端}}(t) = s_{\text{主叫用户}}^{\text{电话机听筒端}}(t) + n_{\text{主叫用户环境噪声+电话传输信道噪声}}(t) + n_{\text{电话机本身噪声}}$$

$$[0108] \quad y_{\text{调度员}}^{\text{电话机听筒端}}(t) = s_{\text{主叫用户调度员}}^{\text{电话机听筒端}}(t) + n_{\text{低噪声影响}}(t)$$

[0109] 考虑其噪声特性,可根据其能量谱的强度,对信号进行适当的放大,使得不同采集端的主叫人和被叫人电话语音强度信号接近。将电话机端主叫人与电话机听筒端主叫人进行对消,可提取出电话机本身噪声,将电话机端被叫人与电话机听筒端被叫人进行对消,再消除此前提取出的电话机本身噪声信号,可获得当前通话状态下的静音时 $n_{\text{主叫用户环境噪声+电话传输信道噪声}}$ 的噪声影响。对比此噪声影响和电话机端主叫人表达式,可分解出 $s_{\text{主叫用户}}^{\text{电话机端}}(t)$ 。该方法可大幅提升电话采集语音信号的噪声抑制特性,提升后续对主叫人进行声纹识别的准确率。

[0110] 端点检测可基于双门限法、自相关法、谱熵法、比例法和对数频谱距离法等多种不同方法进行展开。

[0111] 双门限法:短时能量检测可较好区分出浊音和静音。对于清音,由于其能量较小,在短时能量检测中会因为能量低于能量门限而被误判为静音;短时过零检测则可以从语音中区分出静音和清音。以上两种方面相互结合,可以检测出语音段和静音段。

[0112] 自相关法:语音信号 $x_n(m)$ 的短时自相关函数 $R_n(k)$ 可表示为:

$$[0113] \quad R_n(k) = \sum_{m=0}^{N-1-k} x_n(m)x_n(m+k) \quad (0 \leq k \leq K)$$

[0114] 其中, $K$ 为最大的延迟点数。

[0115] 假设语音序列具有周期性,则其自相关函数也是同周期的周期函数。对于浊音信号可以用自相关函数求出语音波形序列的基音周期。噪声信号和含噪语音的自相关函数在峰值幅度有较大差异,根据噪声的大小,设置适当的阈值,判定对应的是否存在对应的语音

信号,并确定其端点。

[0116] 对数频谱距离法:设含噪语音信号为 $x(n)$ ,加窗分帧处理后得到的第 $i$ 帧语音信号 $x_i(m)$ ,帧长为 $N$ 。针对 $x_i(m)$ 做FFT(快速傅里叶变换),可得:

$$[0117] \quad X_i(k) = \sum_{m=0}^{N-1} x_i(m) \exp(j \frac{2\pi mk}{N}) \quad k = 0, 1, \dots, N-1$$

[0118] 对频谱 $X_i(k)$ 取模值后再取对数,可得:

$$[0119] \quad \hat{X}_i(k) = 20 \lg |X_i(k)|$$

[0120] 由于噪声信号和含噪语音信号的能量谱有显著区别(噪声信号能量谱远低于含噪语音信号的能量谱,故而可通过两帧信号之间的对数频谱差值来确定语音信号的端点。

[0121] 结合短时过零率、端点检测语音和能量谱判断,对两种不同方法所提取的语音信号进行信号比对,可有效地提取出电力调度系统中主叫人的语音信号,用于后续进行语音信号的声纹识别深度学习神经网络框架训练和主叫人声纹识别身份判决与鉴权。

[0122] 在处理过程中,电话手柄端的听筒接口所得到的语音信号由于电话消侧音电路的存在,会导致主叫人与被叫人之间的语音信号有明显的强度区别,结合短时过零率和端点检测,可有效对电话主叫人和被叫人信号进行分段和截取。

[0123] S300对主叫用户语音信号进行降噪,具体为:

[0124] 采用相关特征法对主叫用户语音信号进行降噪:

[0125] 假定主叫用户语音信号与主叫用户环境噪声、电话传输信道噪声互不相干,对带噪信号做自相关处理,可得到与不带噪语音信号类似的自相关帧序列:

$$[0126] \quad R_y(\tau) = \frac{1}{T} \int_{-n}^t y(t)y(t-\tau)w(t)dt \approx \frac{1}{T} \int_{-n}^t s(t)s(t-\tau)w(t)dt = R_s(\tau)$$

[0127] 其中 $s(t)$ 为纯净语音信号, $n(t)$ 为噪声信号, $w(t)$ 是为达成短时平稳所加的窗函数, $R_y(\tau)$ 和 $R_s(\tau)$ 分别是主叫人语音信号带噪和不带噪的自相关函数;

[0128] 采用维纳滤波法对语音信号进行降噪:

[0129] 带噪语音信号经过维纳滤波器后的输出 $s'(t)$ 满足 $E[|s'(t) - s(t)|^2]$ 的值最小,维纳滤波法基于短时平稳语音信号为前提,对于维纳滤波器,得到下式:

$$[0130] \quad H|\omega| = \frac{P_s(\omega)}{P_s(\omega) + P_n(\omega)}$$

[0131] 上式中 $H|\omega|$ 为维纳滤波器频域的冲激响应, $P_s(\omega)$ , $P_n(\omega)$ 分为信号功率谱和噪声功率谱;

$$[0132] \quad S_0(\omega) = H(\omega) \cdot Y(\omega)$$

[0133] 上式中 $S_0(\omega)$ 为维纳滤波器的输出信号频谱, $Y(\omega)$ 为主叫人带噪电话语音信号频谱。

[0134] 对于电力调度电话声纹识别系统而言,由于调度员工作环境、调度电话传输信道、电话机本身所引入的噪声,导致在进行主叫人身份识别和声纹识别处理时,会因为主叫人电话语音质量的降低和干扰问题,导致所采集的主叫人语音信号与大规模语音样本训练时所采用的语音信号具有相差较大的背景和干扰噪声偏差,从而导致声纹识别率的大幅降低。

[0135] 为有效地提高系统的声纹识别成功率,需要尽可能降低主叫人语音信号中背景干扰、电话传输信道干扰和电话机本身引入的干扰。

[0136] 可用的降噪和语音增强方法如下:

[0137] 主动降噪:是基于声波的叠加原理,即通过声波的相互抵消实现噪声去除。通过找到一种与所要消除的噪声频谱完全一样,仅仅相位相反的声音,通过叠加从而将噪声进行抵消。这种方法的难点是噪声的频率与语音信号的频谱融为一体,难以找到一个相位严格相反的声音,进行后续的噪声抵消。

[0138] 对说话人识别的特征提取方法进行了分类整理,将无噪声补偿技术的特征提取方法分为以下几类进行阐述:基于高/低层次的特征提取,变换类型,言语产生/听觉系统,特征提取技术的类型,时变性,语音处理技术。另外将噪声补偿特征提取方法分为噪声屏蔽特征、特征归一化方法、特征补偿方法。

[0139] 基于稀疏约束的非负矩阵分解(NMF)算法。结合普通幅度频谱或梅尔频谱特性和非负矩阵分解的原理,采用了基于梅尔频谱为数据进行矩阵分解的方法,使用基于梅尔频谱的稀疏约束的NMF算法。现有的稀疏约束的NMF算法使用固定的噪声字典和语音字典,当带噪语音的噪声和噪声字典不匹配时,去噪性能会有所下降。

[0140] 将谱减法与理想二值掩蔽(Ideal Binary Mask,IBM)算法结合,对待增强语音先掩蔽增强再谱减降噪。

[0141] 降噪处理需将环境音频信号、电话信道信号、电话机干扰信号与说话人语音信号进行信号分离,得到较纯净的主叫人语音信息。

[0142] 所采集的主叫人语音信息经过降噪处理后,将与预录的说话人音频信号进行声纹识别匹配。

[0143] S400对主叫用户语音信号进行语音增强包括:

[0144] 倒谱均值规整降噪CMN方法清除带非加性噪声主叫人电话语音信号倒谱中的噪声成分,利用CMN方法处理后得到的增强语音倒谱表示为:

$$[0145] \quad \hat{C}_s(t) = C_{sn}(t) - \bar{C}_{sn}(t) = C_s(t) + C_n(t) - \frac{1}{N} \sum_{i=1}^N C_s(t) - C_n(t) = C_s(t) - \frac{1}{N} \sum_{i=1}^N C_s(t)$$

[0146] 其中  $\hat{C}_s(t)$  为增强语音的倒谱,  $C_{sn}(t)$  为带噪语音的倒谱,  $C_s(t)$  为纯净语音的倒谱,  $\bar{C}_{sn}(t)$  为主叫人被采集语音段的倒谱平均值;

[0147] 同态滤波法:对于加性噪声,可采用线性处理的方法,对于非加性噪声,可采用同态滤波法进行处理。由于在语音信号处理中,倒谱信号被广泛采用,因此,可基于倒谱处理的过程,一并达成降噪目标。卷积信号经过同态滤波器后,卷积运算成为了复倒谱的求和运算,因此可分离出乘性噪声。此后,可由复倒谱提取音调参数,并通过频谱分析得到相应的共振峰,从而可进一步得到降噪后的语音信号。利用倒谱均值规整降噪(Cepstral Mean Normalization,CMN)方法可清除带非加性噪声主叫人电话语音信号倒谱中的噪声成分,从而提升语音的质量。

[0148] S203使用短时过零率、端点检测和语音能量谱对第一语音信号和第二语音信号进行语音信号强度分析与信号比对,分离出主叫用户语音信号包括:

[0149] S2031,用短时能量和过零率检测结合的短时过零率检测算法来检测清音,用短时

能量来检测浊音；

[0150] S2032,根据语音信号的浊音和清音选择对应的清音模型和浊音模型进行语音信号端点检测从而获取主叫用户语音信号。

[0151] S2032根据语音信号的浊音和清音选择对应的清音模型和浊音模型进行语音信号端点检测从而获取主叫用户语音信号包括：

[0152] 发清音时,对应的清音激励模型模拟成随机白噪声,使用均值为零,方差为1,并且在时间和幅度值上为白色分布的序列；

[0153] 发浊音时,产生间歇的脉冲波,数学表达为：

$$[0154] \quad g(n) = \begin{cases} (1/2)[1 - \cos(\pi n / N_1)], & 0 \leq n \leq N_1 \\ \cos[\pi(n - N_1) / 2N_2], & N_1 \leq n \leq N_1 + N_2 \\ 0, & \text{其它} \end{cases}$$

[0155] 上式中,N1为斜三角波上升部分的时间,N2为其下降部分的时间；

[0156] 语音信号进行分帧之后,第n帧语音信号 $x_n(m)$ 的能量可表示为：

$$[0157] \quad E_n = \sum_{m=0}^{N-1} x_n^2(m)$$

[0158] 短时过零率为一帧语音中语音信号波形穿过横轴即零电平的次数,可表示为：

$$[0159] \quad Z_n = \frac{1}{2} \sum_{m=0}^{N-1} |\text{sgn}[x_n(m)] - \text{sgn}[x_n(m-1)]|$$

[0160] 其中 $\text{sgn}()$ 为符号函数,通过考察当前采样信号与上一采样信号之间是否出现了波形上符号变化来评估过零次数；

[0161] 对主叫用户语音信号进行能量谱估计：

[0162] 语音信号进行分帧之后,第n帧语音信号 $x_n(m)$ 的能量表示为：

$$[0163] \quad E_n = \sum_{m=0}^{N-1} x_n^2(m)$$

[0164] 采用自相关法提取主叫用户语音信号：

[0165] 语音信号 $x_n(m)$ 的短时自相关函数 $R_n(k)$ 可表示为：

$$[0166] \quad R_n(k) = \sum_{m=0}^{N-1-k} x_n(m)x_n(m+k) \quad (0 \leq k \leq K)$$

[0167] 其中,K为最大的延迟点数,假设语音序列具有周期性,则自相关函数也是同周期的周期函数,对于浊音信号用自相关函数求出语音波形序列的基音周期,噪声信号和含噪语音的自相关函数在峰值幅度有较大差异,根据噪声的大小,设置阈值,确定其端点。

[0168] 声纹识别模型由卷积神经网络CNN和长短期记忆网络LSTM网络串联构建而成。

[0169] S500电力调度系统将收到的用户语音信号和预先录入的声纹信息使用训练好的声纹识别模型进行匹配之前,还包括S410对声纹识别模型进行训练,具体为：

[0170] S411,将预处理后的多个语音信号分为训练集和测试集；

[0171] S412,将训练集输入声纹识别模型；

[0172] S413,声纹识别模型输出语音信号的判断结果；

[0173] S414,迭代训练声纹识别模型至错误率小于预设值。

[0174] S500电力调度系统将收到的用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配包括：

[0175] S501,对降噪后的主叫用户语音信号进行快速傅里叶变换,得到每个音源信号对应的频谱特征；

[0176] 因为语音波是一个非平稳过程,因此,适用于周期、瞬变或平稳随机信号的标准傅里叶变换不能直接表示语音信号,而应该使用短时傅里叶变换对语音信号的频谱进行处理。相应的频谱称为短时谱。

[0177] S502,将频谱特征经梅尔滤波器滤波后取对数,得到主叫用户电话语音信号对应的梅尔频率对数能量谱；

[0178] S503,将梅尔频率对数能量谱经离散余弦变换,得到主叫用户语音信号对应的梅尔系数谱；

[0179] S504,基于对应的梅尔系数谱进行声纹识别处理,判别主叫用户身份并鉴权。

[0180] 实施例2

[0181] 一种电力调度系统人员声纹识别鉴权降噪和语音增强系统,包括：

[0182] 接收模块,用于接收主叫用户通过电话机向调度员发送操作请求和语音信号；

[0183] 第一数据处理模块,用于从主叫用户和调度员的混合语音中分离出主叫用户语音信号；

[0184] 第二数据处理模块,用于电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配；

[0185] 结果输出模块,如果匹配成功则允许用户进行操作,如果匹配不成功则不允许用户进行操作。

[0186] 实施例3

[0187] 一种电力调度系统人员声纹识别鉴权降噪和语音增强装置,通过数据传送路径与电力调度系统人员声纹识别鉴权系统相连接,使电力调度系统人员声纹识别鉴权装置执行一种电力调度系统人员声纹识别鉴权降噪和语音增强方法,包括：

[0188] 数据获取单元,用于接收主叫用户通过电话机向调度员发送操作请求和语音信号；

[0189] 数据处理单元,用于从主叫用户和调度员的混合语音中分离出主叫用户语音信号；

[0190] 判断单元,用于电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配；

[0191] 输出单元,如果匹配成功则允许用户进行操作,如果匹配不成功则不允许用户进行操作。

[0192] 本发明通过主叫用户通过电话机向调度员发送操作请求和语音信号；从主叫用户和调度员的混合语音中分离出主叫用户语音信号；对主叫用户语音信号进行降噪；对主叫用户语音信号进行语音增强；电力调度系统将主叫用户语音信号和具有所述操作的权限的人员预先录入的语音信号使用训练好的声纹识别模型进行匹配；如果匹配成功则允许主叫用户进行操作,如果匹配不成功则不允许主叫用户进行操作。能够同时从调度电话的输入端和话筒端进行电话语音信号提取,通过电话输入端和话筒端的语音对比,剔除不属于主

叫人的语音,提高了对用户语音信号提纯精度,处理后的用户语音信号能够让声纹识别模型对用户声音信息的判断更加准确,减少了调度员的工作,提高了调度效率。

[0193] 以上所述仅是本发明的具体实施方式,使本领域技术人员能够理解或实现本发明。对这些实施例的多种修改对本领域的技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所申请的原理和新颖特点相一致的最宽的范围。

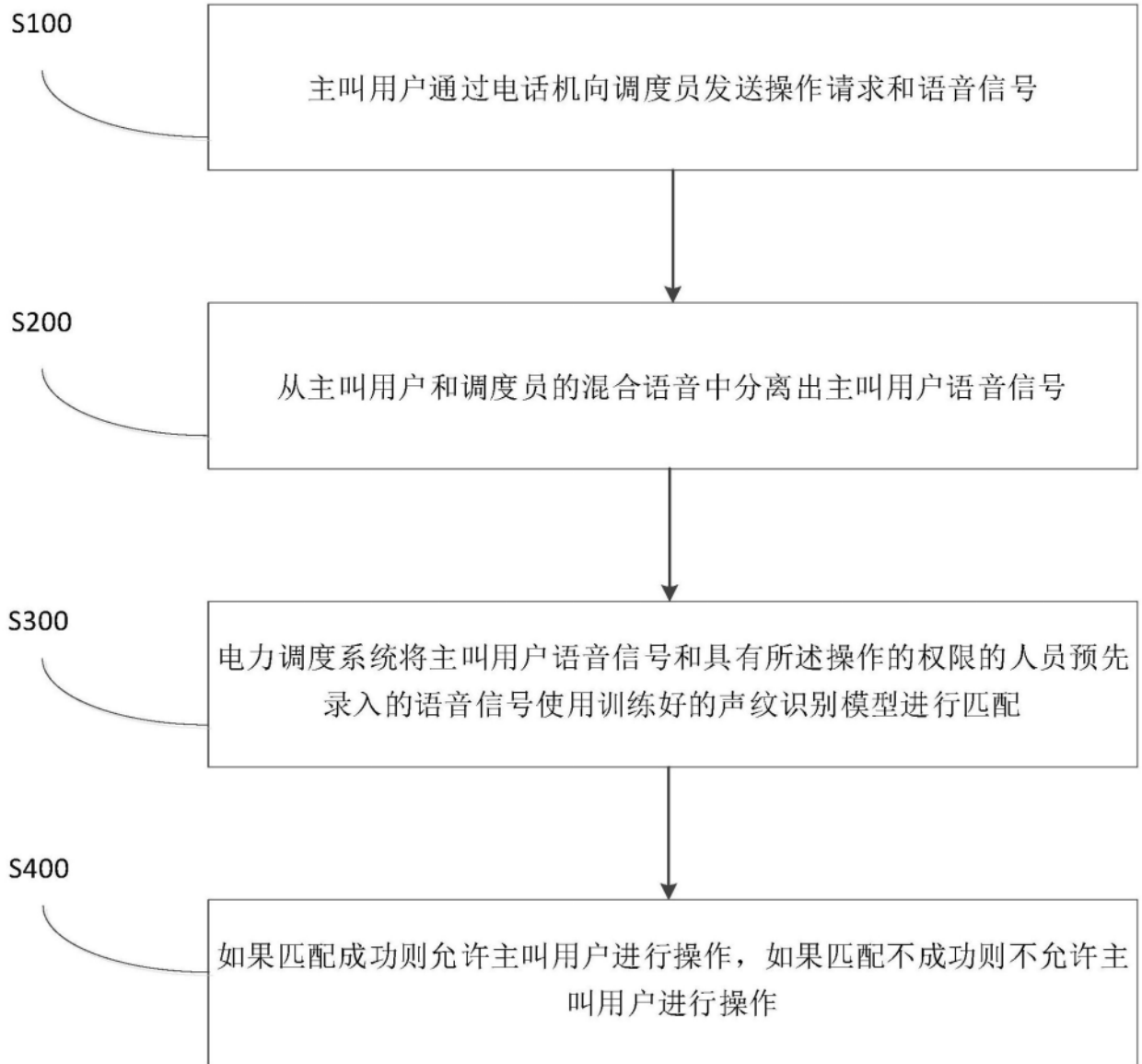


图1

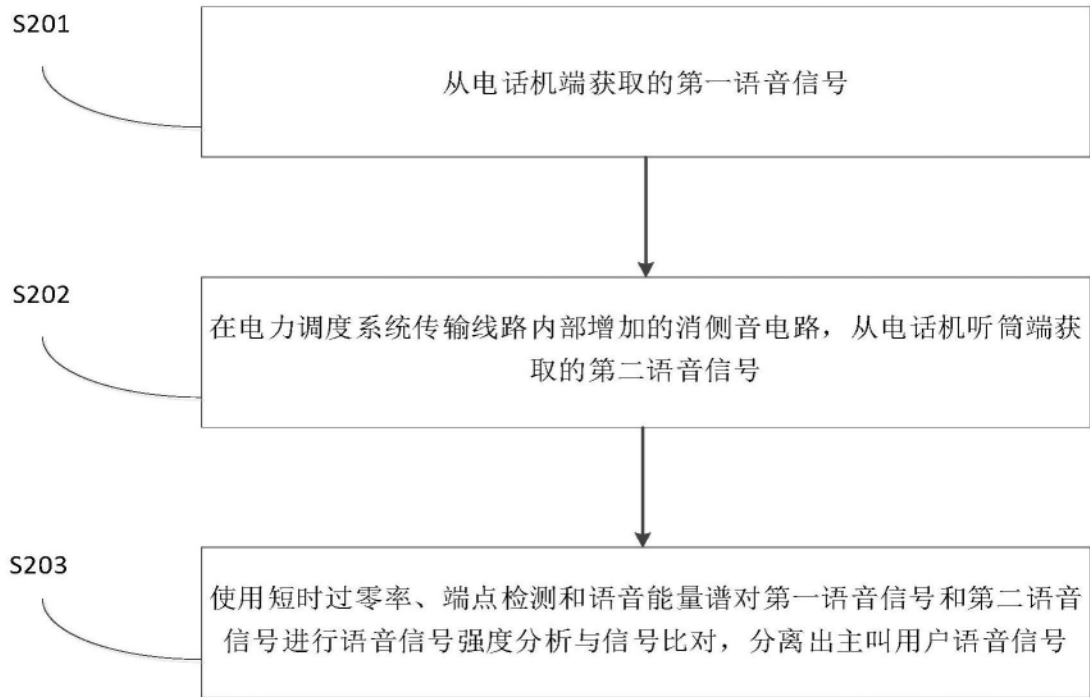


图2

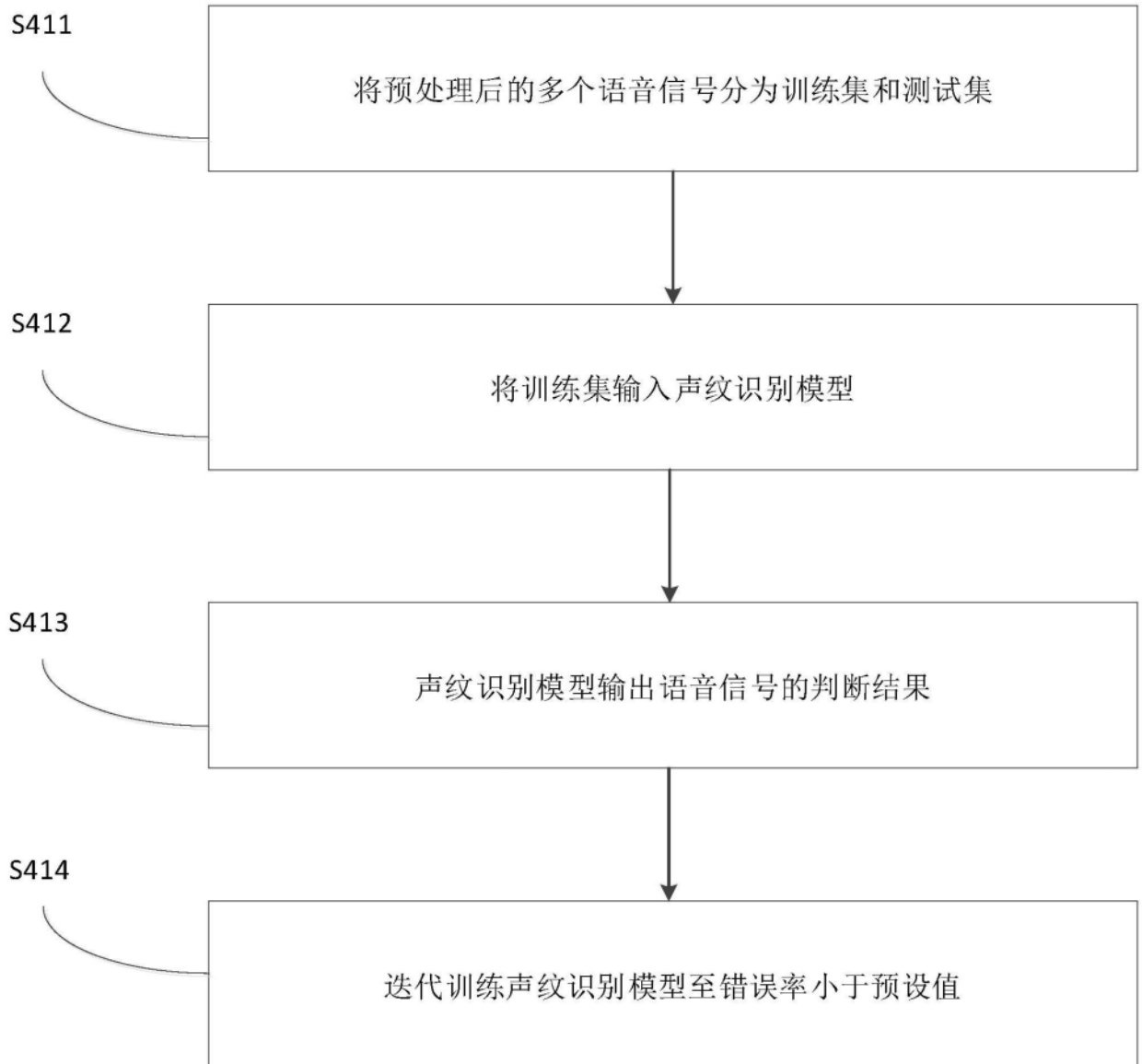


图3

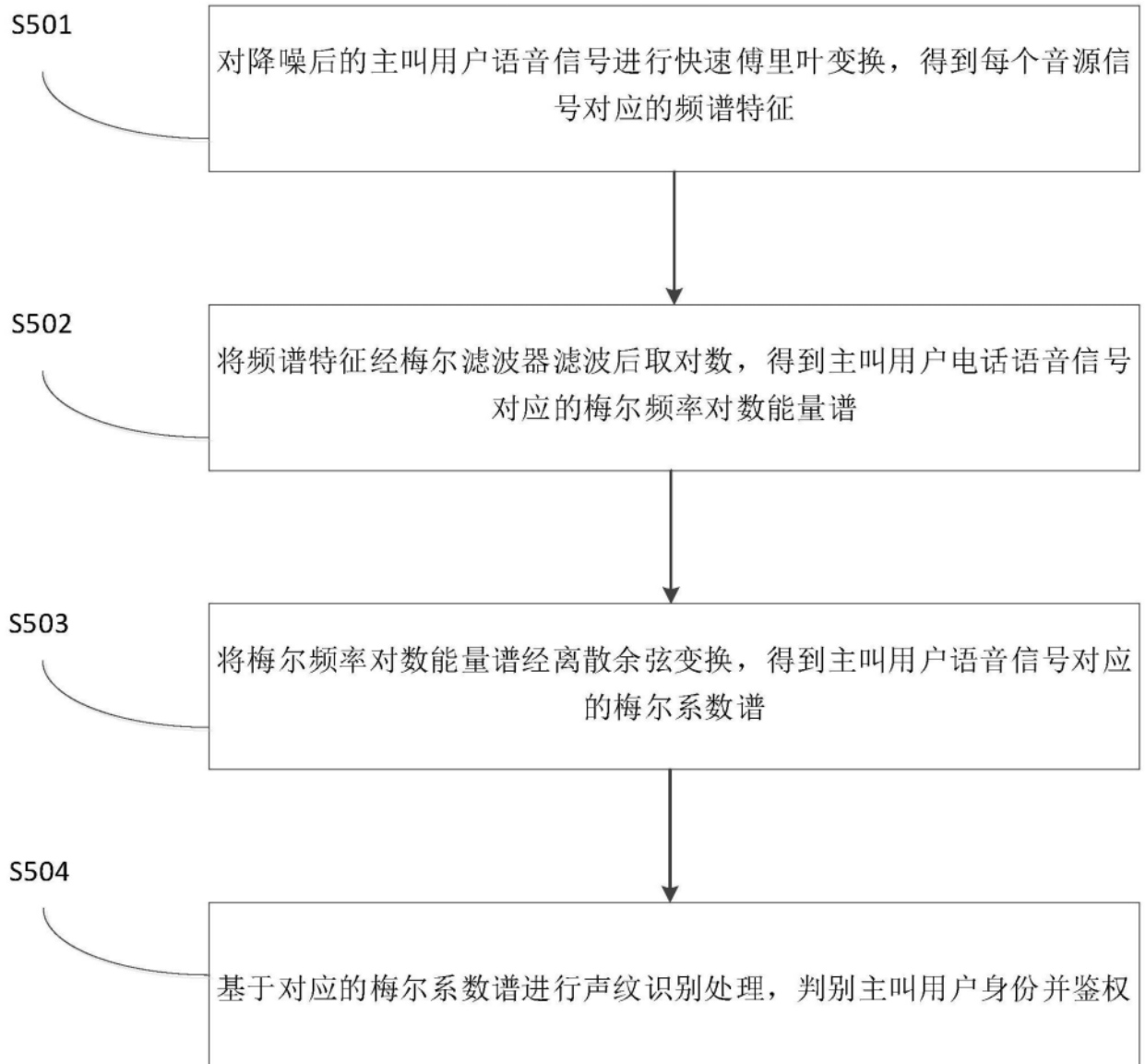


图4