

权 利 要 求 书

1. 一种声纹识别方法，其特征在于，所述声纹识别方法包括：

获取数据库中每个坐席员的呼叫权限，并根据预设的权限分组方式，将每个所述坐席员分配到所述呼叫权限对应的分组中；

针对每个所述分组，对该分组中的每个所述坐席员进行声纹特征采集，得到该分组的基础声纹特征集；

对每个所述基础声纹特征集中的声纹特征进行训练，得到每个所述基础声纹特征集的综合声纹特征，并基于所述分组、所述呼叫权限和所述综合声纹特征之间的声纹对应关系，构建动态声纹识别模型；

若检测到待识别坐席员的呼叫操作，则获取被呼叫客户的客户级别和所述待识别坐席员的语音数据，并对所述语音数据进行声纹特征提取，得到目标声纹特征；

基于所述动态声纹识别模型，对所述目标声纹特征进行识别，确定所述待识别坐席员的呼叫权限；

若所述待识别坐席员的呼叫权限不满足所述客户级别的要求，则执行预设的预警处理措施。

2. 如权利要求 1 所述的声纹识别方法，其特征在于，所述针对每个所述分组，对该分组中的每个所述坐席员进行声纹特征采集，得到该分组的基础声纹特征集，包括：

针对每个所述分组中的每个所述坐席员，获取该坐席员对预设的多个不同语料进行朗读的录音样本，并对所述录音样本进行声纹解析，得到该坐席员的初始声纹样本；

对所述初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

采用分帧和加窗的方式，对所述加重处理声纹样本进行分帧处理，得到初始语音帧；

对所述初始语音帧信号进行静默音分离，得到目标语音帧；

基于所述目标语音帧，提取声纹特征；

在每个所述分组中，将该分组中每个所述坐席员的所述声纹特征组成该分组的所述基础声纹特征集。

3. 如权利要求 2 所述的声纹识别方法，其特征在于，所述基于所述目标语音帧，提取声纹特征，包括：

使用莱文森-德宾算法对所述目标语音帧的预测误差的计算公式 (F1) 中的线性预测系数 $a_1, a_2, \dots, a_p$ ：

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^p a_i s(n-i) \quad \text{F1}$$

其中， $e(n)$ 为所述预测误差， $\hat{s}(n)$ 为所述目标语音帧序列的预测值， n 为所述目标语音帧序列包含的总点数， $s(n-i)$ 为所述目标语音帧序列在前 $n-i$ 个点的取值总和， a_i 为所述目标语音帧序列中第 i 点的线性预测系数， p 为预设的线性预测的阶数， $i \in [1, p]$ ， n 为正整数；

根据所述目标线性预测系数，按照公式 (F2) 计算所述目标语音帧的线性预测倒谱系数，并将所述线性预测倒谱系数作为所述声纹特征：

$$\hat{h}(n) = \begin{cases} a_1' & n=1 \\ a_n' + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) a_k' \hat{h}(n-k) & 1 < n \leq p \\ \sum_{k=1}^p \left(1 - \frac{k}{n}\right) a_k' \hat{h}(n-k) & n > p \end{cases} \quad \text{F2}$$

其中, $\hat{h}(n)$ 为所述线性预测倒谱系数。

4. 如权利要求 1 所述的声纹识别方法, 其特征在于, 所述对每个所述基础声纹特征集中的声纹特征进行训练, 得到每个所述基础声纹特征集的综合声纹特征, 包括:

针对每个所述基础声纹特征集, 采用遗传算法对该基础声纹特征集中的声纹特征进行优化, 得到该基础声纹特征集对应的码本集;

针对第 x 个所述基础声纹特征集, 获取该基础声纹特征集对应的所述码本集内的码本总数 q , 在该码本集中任意选取 k 个码本作为训练样本, 将所述训练样本的码本矩阵按对应位取算术平均值得到类别码本 D_x , 将该类别码本 D_x 作为第 x 个分组的基础声纹特征集的综合声纹特征, 其中 $k \in [1, q]$, x 和 q 为正整数。

5. 如权利要求 4 所述的声纹识别方法, 其特征在于, 所述基于所述动态声纹识别模型, 对所述目标声纹特征进行识别, 确定所述待识别坐席员的呼叫权限, 包括:

将所述目标声纹特征进行预处理, 得到所述目标声纹特征的目标优化码本;

按照公式 (F3) 计算所述目标优化码本到每个所述类别码本的欧式距离:

$$\rho_j(A, D_j) = \sqrt{\sum_{t=1}^z (a_t - b_t)^2} \quad \text{F3}$$

其中, $\rho_j(A, D_j)$ 为目标优化码本 A 到第 j 个所述类别码本 D_j 的欧式距离, a_t 为目标优化码本 A 中第 t 个编码序列特征, b_t 为所述类别码本 D_j 中第 t 个编码序列特征, $j \in [1, J]$, J 为所述基础声纹特征集的个数;

从 J 个所述欧式距离中选取数值最小的欧氏距离, 作为目标欧氏距离;

获取所述目标欧式距离对应的类别码本所在的分组, 将该分组作为当前外呼组别;

根据所述对应关系, 确定所述当前外呼组别对应的当前呼叫权限, 作为所述待识别坐席员的呼叫权限。

6. 一种声纹识别装置, 其特征在于, 所述声纹识别装置包括:

坐席权限分组模块, 用于获取数据库中每个坐席员的呼叫权限, 并根据预设的权限分组方式, 将每个所述坐席员分配到所述呼叫权限对应的分组中;

声纹特征采集模块, 用于针对每个所述分组, 对该分组中的每个所述坐席员进行声纹特征采集, 得到该分组的基础声纹特征集;

识别模型构建模块, 用于对每个所述基础声纹特征集中的声纹特征进行训练, 得到每个所述基础声纹特征集的综合声纹特征, 并基于所述分组、所述呼叫权限和所述综合声纹特征之间的声纹对应关系, 构建动态声纹识别模型;

目标特征获取模块, 用于若检测到待识别坐席员的呼叫操作, 则获取被呼叫客户的客户级别和所述待识别坐席员的语音数据, 并对所述语音数据进行声纹特征提取, 得到目标声纹特征;

呼叫权限确定模块, 用于基于所述动态声纹识别模型, 对所述目标声纹特征进行识别,

确定所述待识别坐席员的呼叫权限；

实时语音预警模块，用于若所述待识别坐席员的呼叫权限不满足所述客户级别的要求，则执行预设的预警处理措施。

7. 如权利要求6所述的声纹识别装置，其特征在于，声纹特征采集模块，包括：

初始声纹获取单元，用于针对每个所述分组中的每个所述坐席员，获取该坐席员对预设的多个不同语料进行朗读的录音样本，并对所述录音样本进行声纹解析，得到该坐席员的初始声纹样本；

声纹加重处理单元，用于对所述初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

初始语音帧生成单元，用于采用分帧和加窗的方式，对所述加重处理声纹样本进行分帧处理，得到初始语音帧；

目标语音帧获取单元，用于对所述初始语音帧信号进行静默音分离，得到目标语音帧；

声纹特征提取单元，用于基于所述目标语音帧，提取声纹特征；

声纹特征集整合单元，用于在每个所述分组中，将该分组中每个所述坐席员的所述声纹特征组成该分组的所述基础声纹特征集。

8. 如权利要求7所述的声纹识别装置，其特征在于，所述声纹特征提取单元，包括：

目标线性预测系数递推子单元，用于使用莱文森-德宾算法对所述目标语音帧的预测误差的计算公式(F1)中的线性预测系数进行预测，并根据预测结果确定目标线性预测系数 $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ ；

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^p a_i s(n-i) \quad \text{F1}$$

其中， $e(n)$ 为所述预测误差， $\hat{s}(n)$ 为所述目标语音帧序列的预测值， n 为所述目标语音帧序列包含的总点数， $s(n-i)$ 为所述目标语音帧序列在前 $n-i$ 个点的取值总和， a_i 为所述目标语音帧序列中第 i 点的线性预测系数， p 为预设的线性预测的阶数， $i \in [1, p]$ ， n 为正整数；

线性预测倒谱系数计算子单元，用于根据所述目标线性预测系数，按照公式(F2)计算所述目标语音帧的线性预测倒谱系数，并将所述线性预测倒谱系数作为所述声纹特征：

$$\hat{h}(n) = \begin{cases} \hat{a}_1 & n=1 \\ \hat{a}_n + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \hat{a}_k \hat{h}(n-k) & 1 < n \leq p \\ \sum_{k=1}^p \left(1 - \frac{k}{n}\right) \hat{a}_k \hat{h}(n-k) & n > p \end{cases} \quad \text{F2}$$

其中， $\hat{h}(n)$ 为所述线性预测倒谱系数。

9. 如权利要求6所述的声纹识别装置，其特征在于，所述识别模型构建模块包括：

特征码本优化单元，用于针对每个所述基础声纹特征集，采用遗传算法对该基础声纹特征集中的声纹特征进行优化，得到该基础声纹特征集对应的码本集；

类别码本训练单元，用于针对第 x 个所述基础声纹特征集，获取该基础声纹特征集对应的所述码本集内的码本总数 q ，在该码本集中任意选取 k 个码本作为训练样本，将所述训练样本的码本矩阵按对应位取算术平均值得到类别码本 D_x ，将该类别码本 D_x 作为第 x

个分组的基础声纹特征集的综合声纹特征，其中 $k \in [1, q]$ ， x 和 q 为正整数。

10. 如权利要求 9 所述的声纹识别装置，其特征在于，所述呼叫权限确定模块包括：

目标优化码本获取单元，用于将所述目标声纹特征进行预处理，得到所述目标声纹特征的目标优化码本；

欧氏距离计算单元，用于按照公式 (F3) 计算所述目标优化码本到每个所述类别码本的欧式距离：

$$\rho_j(A, D_j) = \sqrt{\sum_{t=1}^z (a_t - b_t)^2} \quad \text{F3}$$

其中， $\rho_j(A, D_j)$ 为目标优化码本 A 到第 j 个所述类别码本 D_j 的欧式距离， a_t 为目标优化码本 A 中第 t 个编码序列特征， b_t 为所述类别码本 D_j 中第 t 个编码序列特征， $j \in [1, J]$ ， J 为所述基础声纹特征集的个数；

目标欧式距离选取单元，用于从 J 个所述欧式距离中选取数值最小的欧氏距离，作为目标欧氏距离；

当前外呼组别确定单元，用于获取所述目标欧式距离对应的类别码本所在的分组，将该分组作为当前外呼组别；

当前呼叫权限获取单元，用于根据所述对应关系，确定所述当前外呼组别对应的当前呼叫权限，作为所述待识别坐席员的呼叫权限。

11. 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：

获取数据库中每个坐席员的呼叫权限，并根据预设的权限分组方式，将每个所述坐席员分配到所述呼叫权限对应的分组中；

针对每个所述分组，对该分组中的每个所述坐席员进行声纹特征采集，得到该分组的基础声纹特征集；

对每个所述基础声纹特征集中的声纹特征进行训练，得到每个所述基础声纹特征集的综合声纹特征，并基于所述分组、所述呼叫权限和所述综合声纹特征之间的声纹对应关系，构建动态声纹识别模型；

若检测到待识别坐席员的呼叫操作，则获取被呼叫客户的客户级别和所述待识别坐席员的语音数据，并对所述语音数据进行声纹特征提取，得到目标声纹特征；

基于所述动态声纹识别模型，对所述目标声纹特征进行识别，确定所述待识别坐席员的呼叫权限；

若所述待识别坐席员的呼叫权限不满足所述客户级别的要求，则执行预设的预警处理措施。

12. 如权利要求 11 所述的终端设备，其特征在于，所述针对每个所述分组，对该分组中的每个所述坐席员进行声纹特征采集，得到该分组的基础声纹特征集，包括：

针对每个所述分组中的每个所述坐席员，获取该坐席员对预设的多个不同语料进行朗读的录音样本，并对所述录音样本进行声纹解析，得到该坐席员的初始声纹样本；

对所述初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

采用分帧和加窗的方式，对所述加重处理声纹样本进行分帧处理，得到初始语音帧；

对所述初始语音帧信号进行静默音分离，得到目标语音帧；

基于所述目标语音帧，提取声纹特征；

在每个所述分组中，将该分组中每个所述坐席员的所述声纹特征组成该分组的所述基础声纹特征集。

13. 如权利要求 12 所述的终端设备，其特征在于，的计算公式 (F1) 中的线性预测系数进行预测，并根据预测结果确定目标线性预测系数 $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ ：

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^p \hat{a}_i s(n-i) \quad \text{F1}$$

其中， $e(n)$ 为所述预测误差， $\hat{s}(n)$ 为所述目标语音帧序列的预测值， n 为所述目标语音帧序列包含的总点数， $s(n-i)$ 为所述目标语音帧序列在前 $n-i$ 个点的取值总和， $\hat{a}_i$ 为所述目标语音帧序列中第 i 点的线性预测系数， p 为预设的线性预测的阶数， $i \in [1, p]$ ， n 为正整数；

根据所述目标线性预测系数，按照公式 (F2) 计算所述目标语音帧的线性预测倒谱系数，并将所述线性预测倒谱系数作为所述声纹特征：

$$\hat{h}(n) = \begin{cases} \hat{a}_1 & n=1 \\ \hat{a}_n + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \hat{a}_k \hat{h}(n-k) & 1 < n \leq p \\ \sum_{k=1}^p \left(1 - \frac{k}{n}\right) \hat{a}_k \hat{h}(n-k) & n > p \end{cases} \quad \text{F2}$$

其中， $\hat{h}(n)$ 为所述线性预测倒谱系数。

14. 如权利要求 11 所述的终端设备，其特征在于，所述对每个所述基础声纹特征集中的声纹特征进行训练，得到每个所述基础声纹特征集的综合声纹特征，包括：

针对每个所述基础声纹特征集，采用遗传算法对该基础声纹特征集中的声纹特征进行优化，得到该基础声纹特征集对应的码本集；

针对第 x 个所述基础声纹特征集，获取该基础声纹特征集对应的所述码本集内的码本总数 q ，在该码本集中任意选取 k 个码本作为训练样本，将所述训练样本的码本矩阵按对应位取算术平均值得到类别码本 D_x ，将该类别码本 D_x 作为第 x 个分组的基础声纹特征集的综合声纹特征，其中 $k \in [1, q]$ ， x 和 q 为正整数。

15. 如权利要求 14 所述的终端设备，其特征在于，所述基于所述动态声纹识别模型，对所述目标声纹特征进行识别，确定所述待识别坐席员的呼叫权限，包括：

将所述目标声纹特征进行预处理，得到所述目标声纹特征的目标优化码本；

按照公式 (F3) 计算所述目标优化码本到每个所述类别码本的欧式距离：

$$\rho_j(A, D_j) = \sqrt{\sum_{t=1}^J (a_t - b_t)^2} \quad \text{F3}$$

其中， $\rho_j(A, D_j)$ 为目标优化码本 A 到第 j 个所述类别码本 D_j 的欧式距离， a_t 为目标优化码本 A 中第 t 个编码序列特征， b_t 为所述类别码本 D_j 中第 t 个编码序列特征， $j \in [1, J]$ ， J 为所述基础声纹特征集的个数；

从 J 个所述欧式距离中选取数值最小的欧氏距离，作为目标欧氏距离；

获取所述目标欧式距离对应的类别码本所在的分组，将该分组作为当前外呼组别；

根据所述对应关系，确定所述当前外呼组别对应的当前呼叫权限，作为所述待识别坐席员的呼叫权限。

16. 一个或多个存储有计算机可读指令的非易失性可读存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如下步骤：

获取数据库中每个坐席员的呼叫权限，并根据预设的权限分组方式，将每个所述坐席员分配到所述呼叫权限对应的分组中；

针对每个所述分组，对该分组中的每个所述坐席员进行声纹特征采集，得到该分组的基础声纹特征集；

对每个所述基础声纹特征集中的声纹特征进行训练，得到每个所述基础声纹特征集的综合声纹特征，并基于所述分组、所述呼叫权限和所述综合声纹特征之间的声纹对应关系，构建动态声纹识别模型；

若检测到待识别坐席员的呼叫操作，则获取被呼叫客户的客户级别和所述待识别坐席员的语音数据，并对所述语音数据进行声纹特征提取，得到目标声纹特征；

基于所述动态声纹识别模型，对所述目标声纹特征进行识别，确定所述待识别坐席员的呼叫权限；

若所述待识别坐席员的呼叫权限不满足所述客户级别的要求，则执行预设的预警处理措施。

17. 如权利要求 16 所述的非易失性可读存储介质，其特征在于，所述针对每个所述分组，对该分组中的每个所述坐席员进行声纹特征采集，得到该分组的基础声纹特征集，包括：

针对每个所述分组中的每个所述坐席员，获取该坐席员对预设的多个不同语料进行朗读的录音样本，并对所述录音样本进行声纹解析，得到该坐席员的初始声纹样本；

对所述初始声纹样本进行预加重处理，生成具有平坦频谱的加重处理声纹样本；

采用分帧和加窗的方式，对所述加重处理声纹样本进行分帧处理，得到初始语音帧；

对所述初始语音帧信号进行静默音分离，得到目标语音帧；

基于所述目标语音帧，提取声纹特征；

在每个所述分组中，将该分组中每个所述坐席员的所述声纹特征组成该分组的所述基础声纹特征集。

18. 如权利要求 17 所述的非易失性可读存储介质，其特征在于，所述基于所述目标语音帧，提取声纹特征，包括：

使用莱文森-德宾算法对所述目标语音帧的预测误差的计算公式 (F1) 中的线性预测系数 $a_1, a_2, \dots, a_p$ 进行预测，并根据预测结果确定目标线性预测系数 $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ ：

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^p a_i s(n-i) \quad \text{F1}$$

其中， $e(n)$ 为所述预测误差， $\hat{s}(n)$ 为所述目标语音帧序列的预测值， n 为所述目标语音帧序列包含的总点数， $s(n-i)$ 为所述目标语音帧序列在前 $n-i$ 个点的取值总和， a_i 为所述目标语音帧序列中第 i 点的线性预测系数， p 为预设的线性预测的阶数， $i \in [1, p]$ ， n 为正整数；

根据所述目标线性预测系数，按照公式 (F2) 计算所述目标语音帧的线性预测倒谱系数，并将所述线性预测倒谱系数作为所述声纹特征：

$$\hat{h}(n) = \begin{cases} a'_1 & n=1 \\ a'_n + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) a'_k \hat{h}(n-k) & 1 < n \leq p \\ \sum_{k=1}^p \left(1 - \frac{k}{n}\right) a'_k \hat{h}(n-k) & n > p \end{cases} \quad \text{F2}$$

其中, $\hat{h}(n)$ 为所述线性预测倒谱系数。

19. 如权利要求 16 所述的非易失性可读存储介质, 其特征在于, 所述对每个所述基础声纹特征集中的声纹特征进行训练, 得到每个所述基础声纹特征集的综合声纹特征, 包括:

针对每个所述基础声纹特征集, 采用遗传算法对该基础声纹特征集中的声纹特征进行优化, 得到该基础声纹特征集对应的码本集;

针对第 x 个所述基础声纹特征集, 获取该基础声纹特征集对应的所述码本集内的码本总数 q , 在该码本集中任意选取 k 个码本作为训练样本, 将所述训练样本的码本矩阵按对应位取算术平均值得到类别码本 D_x , 将该类别码本 D_x 作为第 x 个分组的基础声纹特征集的综合声纹特征, 其中 $k \in [1, q]$, x 和 q 为正整数。

20. 如权利要求 19 所述的非易失性可读存储介质, 其特征在于, 所述基于所述动态声纹识别模型, 对所述目标声纹特征进行识别, 确定所述待识别坐席员的呼叫权限, 包括:

将所述目标声纹特征进行预处理, 得到所述目标声纹特征的目标优化码本;

按照公式 (F3) 计算所述目标优化码本到每个所述类别码本的欧式距离:

$$\rho_j(A, D_j) = \sqrt{\sum_{t=1}^z (a_t - b_t)^2} \quad \text{F3}$$

其中, $\rho_j(A, D_j)$ 为目标优化码本 A 到第 j 个所述类别码本 D_j 的欧式距离, a_t 为目标优化码本 A 中第 t 个编码序列特征, b_t 为所述类别码本 D_j 中第 t 个编码序列特征, $j \in [1, J]$, J 为所述基础声纹特征集的个数;

从 J 个所述欧式距离中选取数值最小的欧氏距离, 作为目标欧氏距离;

获取所述目标欧式距离对应的类别码本所在的分组, 将该分组作为当前外呼组别;

根据所述对应关系, 确定所述当前外呼组别对应的当前呼叫权限, 作为所述待识别坐席员的呼叫权限。