

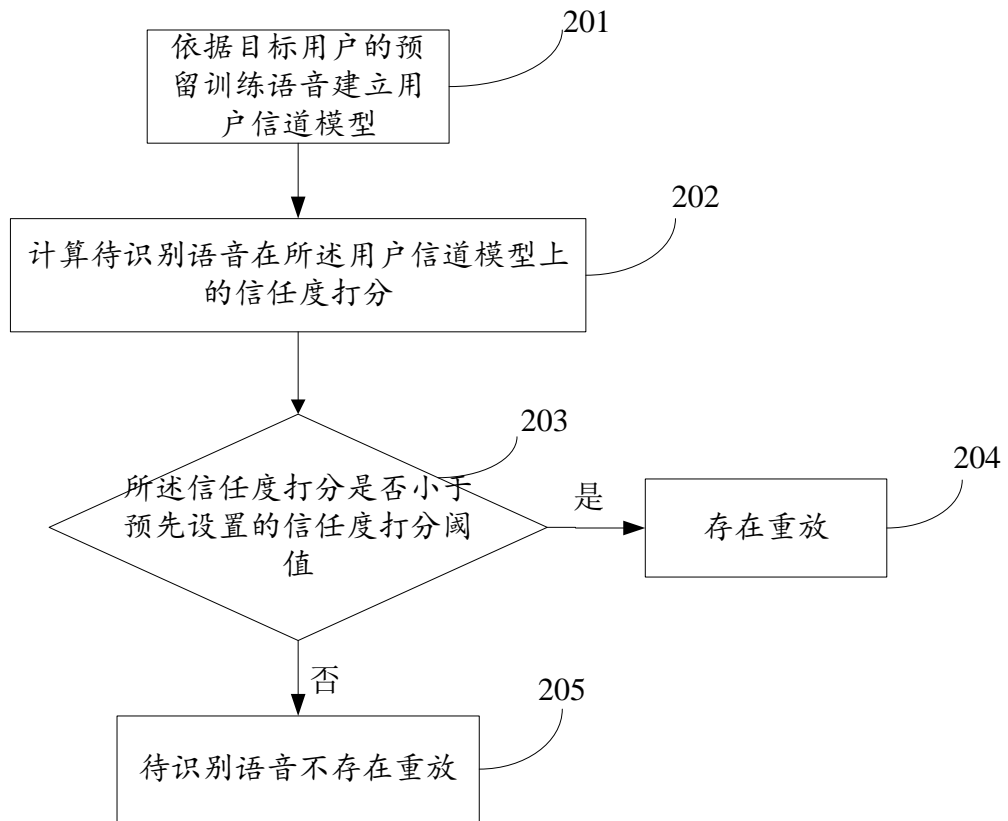
# 说明书摘要

---

本申请公开了一种语音重放检测方法和装置，包括：根据目标用户的预留训练语音建立用户信道模型；计算待识别语音在所述用户信道模型上的信任度打分；若所述信任度打分小于设定阈值，则认定待识别语音存在重放，返回认证失败；反之，通过重放检测。从而解决了现有说话人识别技术中语音重放攻击的问题。

10

# 摘要附图



## 权 利 要 求 书

1. 一种语音重放检测方法，其特征在于，包括：

依据目标用户的预留训练语音建立用户信道模型；

计算待识别语音在所述用户信道模型上的信任度打分，获得待识别语音的信任度打分；

5 若所述信任度打分小于设定阈值，则认定待识别语音存在重放，返回认证失败；反之，通过重放检测。

2. 根据权利要求 1 所述的方法，依据目标用户的预留训练语音建立用户信道模型的步骤包括：

提取目标用户的预留训练语音的低能量语音段；

10 若低能量语音段的时间长度小于设定阈值，则根据信道通用背景模型自适应获取用户信道模型；反之，则采用最大期望算法直接获取用户信道模型。

3. 根据权利要求 2 所述的方法，其特征在于，所述提取目标用户的预留训练语音的低能量语音段的步骤之前还包括：

15 计算当前预留训练语音段的采样值的平方和得到当前预留训练语音段的能量，若所述能量低于设定阈值，则认定为低能量语音段。

4. 根据权利要求 2 所述的方法，其特征在于，依据信道通用背景模型自适应获取用户信道模型的步骤包括：

提取目标用户的预留训练语音的低能量语音段的多复合声学特征；

20 依据最大后验概率估计算法，利用所述低能量训练语音数据的多复合特征对所述信道通用背景模型进行模型自适应更新，得到用户信道模型。

5. 根据权利要求 4 所述的方法，其特征在于，所述提取所述低能量语音段的多复合声学特征的步骤包括：

提取低能量语音段的多类语音声学特征，并拼接成高维度声学特征；

25 使用主成分分析 PCA 对高维度声学特征进行处理，获得正交化的声学特征；

使用线性判别分析 LDA 对正交化后的声学特征进行处理，获得低维度声学特征，将所述低维度声学特征作为多复合声学特征。

6. 根据权利要求 2 所述的方法，其特征在于，训练所述信道通用背景

模型的步骤包括：

获取系统开发集语音数据的多复合声学特征；

计算多复合声学特征中每一帧声学特征在信道通用背景模型上的后验概率；

5 通过最大期望算法计算模型参数的极值，使其在多复合声学特征上的总体概率最大；

通过最大期望算法反复迭代使模型参数不断更新，直到得到收敛的模型参数值，将所述收敛的模型参数值对应的模型作为信道通用背景模型。

7. 根据权利要求 1 所述的方法，其特征在于，计算待识别语音在目标  
10 用户信道模型上的信任度打分的步骤包括：

提取待识别语音的低能量段的多复合声学特征；

计算每一帧多复合声学特征在目标用户信道模型上的概率似然分；

计算全部概率似然分的平均值作为待识别语音信任度打分。

8. 一种语音重放检测装置，其特征在于，包括：

15 用户信道模块，用于依据目标用户的预留训练语音建立用户信道模型；  
计算模块，用于计算待识别语音在目标用户信道模型上的信任度打分，  
获得待识别语音的信任度打分；

第一判断模块，用于若所述信任度打分小于设定阈值，则认定待识别语音存在重放，返回认证失败；反之，通过重放检测。

20 9. 根据权利要求 8 所述的装置，用户信道模块包括：

第一提取模块，用于提取目标用户的预留训练语音的低能量语音段；

多复合声学特征模块，用于提取所述低能量语音段的多复合声学特征；

25 第二判断模块，用于若低能量语音段的时间长度小于设定阈值，则根据信道通用背景模型自适应获取用户信道模型；反之，则采用最大期望算法直接获取用户信道模型。

10、根据权利要求 9 所述的方法，其特征在于，第一提取模块之前还包括：

第三判断模块，用于计算当前预留训练语音段的采样值的平方和得到当

前预留训练语音段的能量，若所述能量低于设定阈值，则认定为低能量语音段。

## 语音重放检测方法和装置

### 技术领域

本申请涉及计算机信息服务技术领域，特别是涉及一种语音重放检测方法和装置。

### 5 背景技术

说话人识别技术，又称为声纹识别技术，主要是基于语音中说话人包含的个性特征的信息，利用计算机以及各种信息识别技术，自动地实现说话人身份的确认。

10 近几年来，随着互联网的飞速发展，语音作为一种非接触性信息载体，人们可以依靠各种移动终端设备，例如：手机、麦克风和 IP 电话等，随时随地的完成语音采集，并通过网络传输和后台服务器来实现人机交互和说话人身份识别。

随着移动互联网时代的到来，在给予人类方便快捷的同时，也带来了许多安全隐患。如图 1 所示，若说话人 A 的语音被企图闯入者预先窃取并录  
15 制，继而将录制的语音重新播放至说话人识别系统，说话人识别系统通常将难以区分，从而导致企图闯入者成功地以说话人 A 的录音闯入说话人 A 的账号，对说话人 A 的账号带来极大的安全隐患。

### 发明内容

20 本申请提供一种录音重放检测方法和装置，以解决说话人识别技术中语音重放的问题。

为了解决上述问题，本申请公开了一种语音重放检测方法，包括：

依据目标用户的预留训练语音建立用户信道模型；

计算待识别语音在所述用户信道模型上的信任度打分，获得待识别语音的信任度打分；

若所述信任度打分小于设定阈值，则认定待识别语音存在重放，返回认证失败；反之，通过重放检测。

优选地，依据目标用户的预留训练语音建立用户信道模型的步骤包括：

提取目标用户的预留训练语音的低能量语音段；

- 5 若低能量语音段的时间长度小于设定阈值，则依据信道通用背景模型自适应获取用户信道模型；反之，则采用最大期望算法直接获取用户信道模型。

优选地，所述提取目标用户的预留训练语音的低能量语音段的步骤之前还包括：

- 10 计算当前预留训练语音段的采样值的平方和得到当前预留训练语音段的能量，若所述能量低于设定阈值，则认定为低能量语音段。

优选地，依据信道通用背景模型自适应获取用户信道模型的步骤包括：

提取目标用户的预留训练语音的低能量段的多复合声学特征；

依据最大后验概率估计算法，利用所述低能量训练语音数据的多复合特征对所述信道通用背景模型进行模型自适应更新，得到用户信道模型。

- 15 优选地，所述提取所述低能量语音段的多复合声学特征的步骤包括：

提取低能量语音段的多类语音声学特征，并拼接成高维度声学特征；

使用主成分分析 PCA 对高维度声学特征进行处理，获得正交化的声学特征；

- 20 使用线性判别分析 LDA 对正交化后的声学特征进行处理，获得低维度声学特征，将所述低维度声学特征作为多复合声学特征。

优选地，训练所述信道通用背景模型的步骤包括：

获取系统开发集语音数据的多复合声学特征；

计算多复合声学特征中每一帧声学特征在信道通用背景模型上的后验概率；

- 25 通过最大期望算法计算模型参数的极值，使其在多复合声学特征上的总体概率最大；

通过最大期望算法反复迭代使模型参数不断更新，直到得到收敛的模型参数值，将所述收敛的模型参数值对应的模型作为信道通用背景模型。

优选地，计算待识别语音在目标用户信道模型上的信任度打分的步骤包括：

提取待识别语音的低能量段的多复合声学特征；

计算每一帧多复合声学特征在目标用户信道模型上的概率似然分；

5 计算全部概率似然分的平均值作为待识别语音的信任度打分。

为了解决上述问题，本申请还公开了一种语音重放检测装置，包括：

用户信道模块，用于依据目标用户的预留训练语音建立用户信道模型；

计算模块，用于计算待识别语音在目标用户信道模型上的信任度打分，

获得待识别语音的信任度打分；

10 第一判断模块，用于若所述信任度打分小于设定阈值，则认定待识别语音存在重放，返回认证失败；反之，通过重放检测。

优选地，用户信道模块包括：

第一提取模块，用于提取目标用户的预留训练语音的低能量语音段；

多复合声学特征模块，用于提取所述低能量语音段的多复合声学特征；

15 第二判断模块，用于若低能量语音段的时间长度小于设定阈值，则依据信道通用背景模型自适应获取用户信道模型；反之，则采用最大期望算法直接获取用户信道模型。

20 优选地，第一提取模块之前还包括：第三判断模块，用于计算当前预留训练语音段的采样值的平方和得到当前预留训练语音段的能量，若所述能量低于设定阈值，则认定为低能量语音段。

与现有技术相比，本申请包括以下优点：

25 本申请通过目标用户的预留训练语音建立用户信道模型，依据用户信道模型计算待识别语音的信任度打分，将信任度打分与设定的阈值比较，若信任度打分小于设定阈值，则认定待识别语音存在重放，返回待识别语音认证失败，反之，则通过重放检测，即待识别语音认证成功，通过计算待识别语音在所述用户信道模型上的信任度打分，从而避免了闯入者重放攻击的问题。



## 附图说明

图 1 是本申请录音重放应用场景的示意图；

图 2 是本申请实施例一所述一种语音重放检测方法的流程图；

图 3 是本申请语音数据在录音前后的时域图；

5 图 4 是本申请语音数据在录音前后的频域图；

图 5 是本申请实施例二所述一种语音重放检测方法的流程图；

图 6 是本申请一种语音重放检测方法应用的实例；

图 7 是本申请实施例三所述一种语音重放检测装置的结构框图。

## 具体实施方式

10 为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

### 实施例一

参照图 2，示出了本申请实施例一所述一种语音重放检测方法的流程图，具体包括：

15 步骤 201：依据目标用户的预留训练语音建立用户信道模型。

预先获取目标用户的预留训练语音，根据获取的目标用户的预留训练语音建立用户信道模型。

可以通过从后台服务器或者目标用户的客户端获取预留训练语音，也可以采用其他方式获取预留训练语音，对此本申请不做具体限制。

20 步骤 202：计算待识别语音在所述用户信道模型上的信任度打分。

本申请使用用户信道模型对用户端输入的待识别语音进行信任度打分，获取待识别语音的信任度打分，根据信任度打分判断输入的待识别语音是否存在重放。

25 待识别语音包括录音数据、模仿语音数据和合成语音数据等等，录音数据、模仿语音数据和合成语音数据均对应有说话人的真实语音数据。

从信号处理的角度来看，即使企图闯入者使用高保真的录音设备录音说话人的语音数据，由于说话人的真实语音数据在进入说话人识别系统之前必

定经过一个相同的或不同的录音和放音系统，这两个额外的系统就会对真实语音数据带来额外的频谱损伤，所以得到的录音数据必定与真实语音数据存在差异，本申请从信号处理角度出发，分析了录音重放前后语音数据的变化过程。

5 参见图 3 其示出了针对同一条语音数据在录音前后的时域图，包括：录音前的语音数据的时域图 301 和录音后的语音数据的时域图 302。其中，录音前的语音数据是指说话人发出的真实语音数据，可以称为真实语音数据，真实语音数据是指未经过录音、模仿和合成的语音数据。

从图 3 中可以看出，原语音数据经过录音重放后，整体信号强度减弱，  
10 能量幅值也降低，并且信号强度减弱程度并不稳定，有的地方减弱程度较大，有的地方减弱程度相对较小。

参见图 4 其示出了针对同一条语言数据在录音前后的频域图，包括：录音前的语音数据的频域图 401 和录音后的语音数据的频域图 402，从图 4 可以看出语音数据经过录音重放后，频域分布变得平滑，响度降低。并且从整个频域包络线可以看出，录音重放前后信号频率分布也发生了变化，因此可以  
15 从信道特性和倒谱系数分析语音数据重放。

步骤 203：信任度打分与预先设置的信任度打分阈值进行比较，若小于预设的信任度打分阈值，则执行步骤 204，否则执行步骤 205。

其中，信任度打分阈值的设定可以由本领域技术人员采用任意适当方式  
20 进行设定，如可以采用人工经验设定阈值，或者针对历史数据的差异值设定阈值，本申请对此不作限制。

步骤 204：待识别语音存在重放，返回认证失败。

步骤 205：待识别语音不存在重放，通过重放检测。

通过目标用户的预留训练语音建立用户信道模型，依据用户信道模型计  
25 算待识别语音的信任度打分，将信任度打分与设定的阈值比较，若信任度打分小于设定阈值，则认定待识别语音存在重放，返回待识别语音认证失败，反之，则通过重放检测，即待识别语音认证成功，通过计算待识别语音在所述用户信道模型上的信任度打分，，从而避免了闯入者重放攻击的问题。

## 实施例二

参照图 5，示出了本申请实施例二所述一种语音重放检测方法的流程图，具体包括：

步骤 501：依据目标用户的预留训练语音建立用户信道模型。

5 步骤 501 包括以下子步骤：

子步骤 5011：计算当前预留训练语音段的采样值的平方和得到当前预留训练语音段的能量，若所述能量低于设定阈值，则认定训练语音段为低能量语音段。

子步骤 5012：提取目标用户的预留训练语音段的低能量语音段。

10 对目标用户的预留训练语音进行提取获取预留训练语音的低能量语音段，采用基于短时能量算法检测得到满足条件的低能量语音段。

由于低能量语音段中含有噪音，因此需要在提取低能量语音段后进行去噪处理，目前较为流行的去噪方法有谱减法、RASTA 滤波法、主分量分析法等。

15 语音段的一个短段的能量称为短时能量，第  $n$  段的短时能量用  $E_n$  表示，可以利用公式 (1) 得到  $E_n$ 。

$$E_n = \sum_{m=-\infty}^{\infty} [X(m)w(n-m)]^2 \quad (1)$$

其中，本窗短时能量序列  $E_n$  反映了语音振幅或能量随时间变化的规律， $w(n)$  代表了固定数据窗， $X(m)$  代表了语音窗长，根据公式 (1) 可以计算出语音段上一组  $E_n$  值。

20

根据本窗短时能量序列可以使用公式 (2) 得出整个语音段的平均短时能量。

$$E = \frac{1}{N} \sum_{i=1}^N E_i \quad (2)$$

将平均短时能量与当前窗的短时能量进行比较，若连续  $K$  窗都小于平均短时能量，则认为连续  $K$  窗语音段是低能量语音段， $K$  一般取 10-20，根据实际使用环境可以进行适当的选择。

25

子步骤 5013: 若低能量语音段的时间长度小于低能量语音段的设定阈值, 则执行步骤 5015, 若低能量语音段的时间长度大于低能量语音段的设定阈值, 则执行步骤 5014。

步骤 5014: 则采用最大期望算法直接获取用户信道模型。

5 步骤 5015: 依据信道通用背景模型自适应获取用户信道模型。

信道通用背景模型 (channel universal background model, 简称 CUBM) 的训练一般需要利用大量的不同信道下的语音来建立一个稳定的信道通用背景模型, CUBM 模型中的信道信息覆盖的情况越均匀广泛, 其适用性越好。

10 其中, 低能量语音段的设定阈值可以由本领域技术人员采用任意适当方式进行设定, 如可以采用人工经验设定阈值, 或者针对历史数据的差异值设定阈值, 本申请对此不作限制。

在得到 CUBM 模型后, 利用用户提供的预留训练语音在 CUBM 上自适应得到对应用户信道模型(channel speaker model, 简称 CSM), 该 CSM 模型代表了用户提供预留训练语音所处的信道性质。

具体地, 依据信道通用背景模型自适应获取用户信道模型的步骤包括: 提取目标用户的预留训练语音的低能量段的多复合声学特征。

依据最大后验概率估计算法, 利用所述低能量训练语音数据的多复合特征对所述信道通用背景模型进行模型自适应更新, 得到用户信道模型。

20 采用最大后验概率估计算法(Maximum a posteriori, 简称 MPA)对 CUBM 信道无关模型, 进行信道通用背景模型自适应更新, 得到用户信道相关对应的信道模型 CSM。CSM 模型的估计是由用户信道训练数据的平均特性和 CUBM 模型先验信息的一个折中, 即得到的用户信道 CSM 模型是当前用户信道数据的期望 $E(X)$ 和 CUBM 先验均值的加权平均得到的结果, 可表示为:

$$\lambda^{CSM} = \alpha E(X) + (1 - \alpha) \lambda^{CUBM}$$

其中,  $\alpha$  称为信道灵敏度, 取值范围在 0~1 之间; 显然, 当  $\alpha$  趋近于 1 时, 用户信道 CSM 模型更趋向于当前用户信道数据; 反之 CSM 模型更趋向

于 CUBM 的先验知识。在实际应用中， $\alpha$  通常由用户信道语音数据量的大小决定。当用户信道数据充分时， $\alpha$  可取值偏大；反之  $\alpha$  取值偏小，在自适应得到用户的 CSM 信道模型后，将所述用户信道模型存入录音信道模型库。

- 5 子步骤 5012: 所述提取所述低能量语音段的多复合声学特征的步骤包括以下子步骤: 50121: 提取低能量语音段的多类语音声学特征, 并拼接成高维度声学特征。

50122: 使用主成分分析 PCA 对高维度声学特征进行处理, 获得正交化的声学特征。

- 10 PCA 算法去除声学特征的相关性按照如下的方式进行: 假设输入的数据集合  $X$ ,  $X=(x_1, \dots, x_n)$  中维度的声学特征均是在  $D$  维空间中变动的点, 则需要从这个  $D$  维空间中寻找  $D$  个相互正交的基向量, 作为空间的基底来表示当前空间。

- 15 为了求解正交的基向量集合, 假设其为  $\{u_i, i=1, \dots, D\}$ ,  $u_i$  为第  $i$  个正交基向量, 由于这个基向量集合是完备的, 所以使用公式 (8) 对  $D$  维空间中的变动点  $x_n$  表示为基向量的线性组合。

$$x_n = \sum_{i=1}^D \alpha_{ni} u_i \quad (8)$$

运用拉格朗日乘子法对公式 (8) 进行运算, 得到的对应求解为:

$$S_T u_i = \lambda_i u_i$$

- 20 其中, 矩阵  $S_T$  为对训练数据的协方差矩阵, 使用公式 (9) 计算公式  $S_T$ :

$$S_T = \sum_{n=1}^N (x_n - \mu)(x_n - \mu)^T \quad (9)$$

其中,  $\mu$  数据集  $X$  的平均值。

通过公式 (9) 计算出的对应的特征向量即为所需要的正交基向量组。

- 25 通过 PCA 算法可以进行数据降维, 因此可以用 PCA 算法对多种声学特征进行正交化, 定义正交化的投影矩阵为  $W_{PCA}$ , 那么经过 PCA 变化以后,

高维度的声学特征特征向量  $X$  经过变换后输出为  $Y = W_{PCA}^T X$ ，且投影矩阵为：

$$W_{PCA} = \arg \max_W |W^T S_T W| \quad \text{其中，} W^T \text{ 为正交化投影矩阵 } W \text{ 的转置。}$$

50123：使用线性判别分析 LDA 对正交化后的声学特征进行处理，获得  
5 低维度声学特征，将所述低维度声学特征作为多复合声学特征。

线性判别分析 LDA 利用了数据上的分类标签，能够在低维空间上寻找不同类别数据间的最大区分性，先对数据进行 PCA 变化可以有效避免 LDA 计算过程中的矩阵奇异问题。

10  $n$  维空间  $R^n$  上处理后的高维度的声学特征训练集为  $X$ ，其中，总的数据量为  $N$  个声学特征训，总的数据类为  $c$  类，对于其中第  $i$  类 ( $i \in [1, c]$ ) 的数据样本集  $X_i$  包含  $N_i$  个数据点，所以  $N_1 + N_2 + \dots + N_c = N$ ，通过公式 (10) 可以求得第  $i$  类训练集的样本均值，通过公式 (10) 可以求得第  $i$  类训练集的总体样本的均值分：

$$\mu_i = \frac{1}{N} \sum_{x \in X_i} x \quad 1 \leq n \leq N \quad (10)$$

$$15 \quad \mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (11)$$

LDA 的目标是使声学特征经过投影以后，相同类的数据内聚，类间的数据尽量远离，因此通过公式 (12) 计算各类的相同类内离散度  $S_W$  和通过公式 (13) 计算不同类之间的类间离散度  $S_B$ ：

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (12)$$

$$20 \quad S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (13)$$

定义从  $R^n$  空间上向低维空间投影的矩阵为  $W$ ，那么投影以后的类内离散度和类间离散度为：

$$\tilde{S}_W = W^T S_W W, \tilde{S}_B = W^T S_B W$$

所以 LDA 的优化准则就是类间离散度与类内离散度的比值的最大化，

即：

$$W_{LDA} = \arg \max \frac{|W^T S_B W|}{|W^T S_w W|}$$

优选地，训练所述信道通用背景模型的步骤包括：

获取系统开发集语音数据的多复合声学特征。

- 5 计算多复合声学特征中每一帧声学特征在信道通用背景模型上的后验概率。

通过最大期望算法计算模型参数的极值，使其在多复合声学特征上的总体概率最大。

- 10 采用最大期望算法(Expectation Maximization,简称 EM)，定义训练 CUBM 模型的语音数据的多复合信道特征集合为  $X$ ，共有  $T$  帧，即， $X=(x_1, x_2 \dots x_t \dots x_T)$ ， $x_t$  是第  $t$  帧信道特征矢量，则定义 CUBM 模型的初始参数为  $\lambda$ ，则特征集合  $X$  在该模型上的概率分布可表示为： $p(X|\lambda)=\prod_{t=1}^T p(x_t|\lambda)$ 。

EM 算法依据最大似然准则，通过不断迭代使得当前特征集  $X$  在该 CUBM 模型参数  $\lambda$  上的似然概率最大。即为  $\arg \max_{\lambda} P(X|\lambda)$ 。

- 15 通过最大期望算法反复迭代使模型参数不断更新，直到得到收敛的模型参数值，将所述收敛的模型参数值对应的模型作为信道通用背景模型。

步骤 502：计算待识别语音在所述用户信道模型上的信任度打分，获得待识别语音的信任度打分。

可以采用以下方式获得待识别语音的信任度打分：

- 20 使用录音信道模型库计算所述待识别语音的多复合声学特征的概率似然分，获得第一概率似然分；

使用信道通用背景模型计算所述待识别语音的多复合声学特征的概率似然分，获得第二概率似然分；

- 25 计算所述第一概率似然分与所述第二概率似然分的差值，若所述差值小于设定阈值，则待识别语音存在重放，反之，在待识别语音不存在重放。

或者

提取待识别语音的低能量段的多复合声学特征。

计算每一帧多复合声学特征在目标用户信道模型上的概率似然分。

计算全部概率似然分的平均值作为待识别语音信任度打分。

步骤 503: 若信任度打分与预先设置的信任度打分阈值进行比较, 若小于预设的信任度打分阈值, 则执行步骤 504, 否则执行步骤 505。

- 5 其中, 预先设置的信任度打分阈值的设定可以由本领域技术人员采用任意适当方式进行设定, 如可以采用人工经验设定阈值, 或者针对历史数据的差异值设定阈值, 本申请对此不作限制。

步骤 504: 待识别语音存在重放, 返回认证失败。

步骤 505: 待识别语音不存在重放, 通过重放检测。

- 10 本实施例, 通过目标用户的预留训练语音建立用户信道模型, 依据用户信道模型计算待识别语音的信任度打分, 将信任度打分与设定的阈值比较, 若信任度打分小于设定阈值, 则认定待识别语音存在重放, 返回待识别语音认证失败, 反之, 则通过重放检测, 即待识别语音认证成功, 通过计算待识别语音在所述用户信道模型上的信任度打分, 从而避免了闯入者重放攻击的问题。为了本领域技术人员更好的理解本申请限定的技术方案,

参见图 6 所述本申请所述一种语音重放检测方法的应用的实例。

在实际应用中, 本发明可作为说话人识别系统的前端, 即增设了一个录音重放检测模块, 具体如下图 6 所示:

- 20 在实际应用中, 将依据目标用户的预留训练语音建立的用户信道模块以及根据信道通用背景模型自适应获取用户信道模型集成在目标用户的录音重放检测模块。在测试阶段, 计算待识别语音在用户信道模型上的信任度打分, 获取待识别语音的信任度打分, 若所述信任度打分小于设定阈值, 则认定待识别语音存在重放, 直接将待识别语音拒绝; 反之, 待识别语音通过录音重放模块进入说话人识别系统, 进行说话人的身份认证和识别。

- 25 基于上述方法实施例的说明, 本申请还提供了相应的一种语音重放检测装置的实施例, 来实现上述方法实施例所述的内容。

实施例三

参见图 7, 示出了本申请实施例三中的一种语音重放装置的结构框图,



具体可以包括：用户信道模块 701，用于根据目标用户的预留训练语音建立用户信道模型。

计算模块 702，用于计算待识别语音在目标用户信道模型上的信任度打分。

- 5 第一判断模块 703，用于若所述信任度打分小于设定阈值，则认定待识别语音存在重放，返回认证失败；反之，通过重放检测。

优选地，用户信道模块包括：第一提取模块，用于提取目标用户的预留训练语音的低能量语音段。

多复合声学特征模块，用于提取所述低能量语音段的多复合声学特征。

- 10 第二判断模块，用于若低能量语音段的时间长度小于设定阈值，则根据信道通用背景模型自适应获取用户信道模型；反之，则采用最大期望算法直接获取用户信道模型。

- 15 优选地，第一提取模块之前还包括：第三判断模块，用于计算当前预留训练语音段的采样值的平方和得到当前预留训练语音段的能量，若所述能量低于设定阈值，则认定为低能量语音段。

优选地，所述多复合声学特征模块包括：拼接模块，用于提取低能量语音段的多类语音声学特征，并拼接成高维度声学特征。

主成分分析模块，用于使用主成分分析 PCA 对高维度声学特征进行处理，获得正交化的声学特征。

- 20 线性判别分析模块，用于使用线性判别分析 LDA 对正交化后的声学特征进行处理，获得低维度声学特征，将所述低维度声学特征作为多复合声学特征。

优选地，第二判断模块，用于若低能量语音段的时间长度小于设定阈值，则根据信道通用背景模型自适应获取用户信道模型包括：

- 25 第二提取模块，用于第二判断模块，用于若低能量语音段的时间长度小于设定阈值，则根据信道通用背景模型自适应获取用户信道模型。

自适应模块，用于根据最大后验概率估计算法，利用所述低能量训练语音数据的多复合特征对所述信道通用背景模型进行模型自适应更新，得到用

户信道模型。

优选地，所述训练所述信道通用背景模型包括：

开发集模块，用于获取系统开发集语音数据的多复合声学特征；

5 后验模块，用于计算多复合声学特征中每一帧声学特征在信道通用背景模型上的后验概率；

概率模块，用于通过最大期望算法计算模型参数的极值，使其在多复合声学特征上的总体概率最大；

10 更新模块，用于通过最大期望算法反复迭代使模型参数不断更新，直到得到收敛的模型参数值，将所述收敛的模型参数值对应的模型作为信道通用背景模型。

优选地，计算模块包括：

第三提取模块，用于提取待识别语音的低能量段的多复合声学特征；

概率似然分模块，用于计算每一帧多复合声学特征在目标用户信道模型上的概率似然分；

15 输出模块，用于计算全部概率似然分的平均值作为待识别语音的录信任度打分。

本实施例，通过目标用户的预留训练语音建立用户信道模型，依据用户信道模型计算待识别语音的信任度打分，将信任度打分与设定阈值的比较，若信任度打分小于设定阈值，则认定待识别语音存在重放，返回待识别语音  
20 认证失败，反之，则通过重放检测，即待识别语音认证成功，通过计算待识别语音在所述用户信道模型上的信任度打分，从而避免了闯入者重放攻击的问题。

25 对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

以上对本申请所提供的一种录音重放检测方法和装置，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于  
5 本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

# 说明书附图

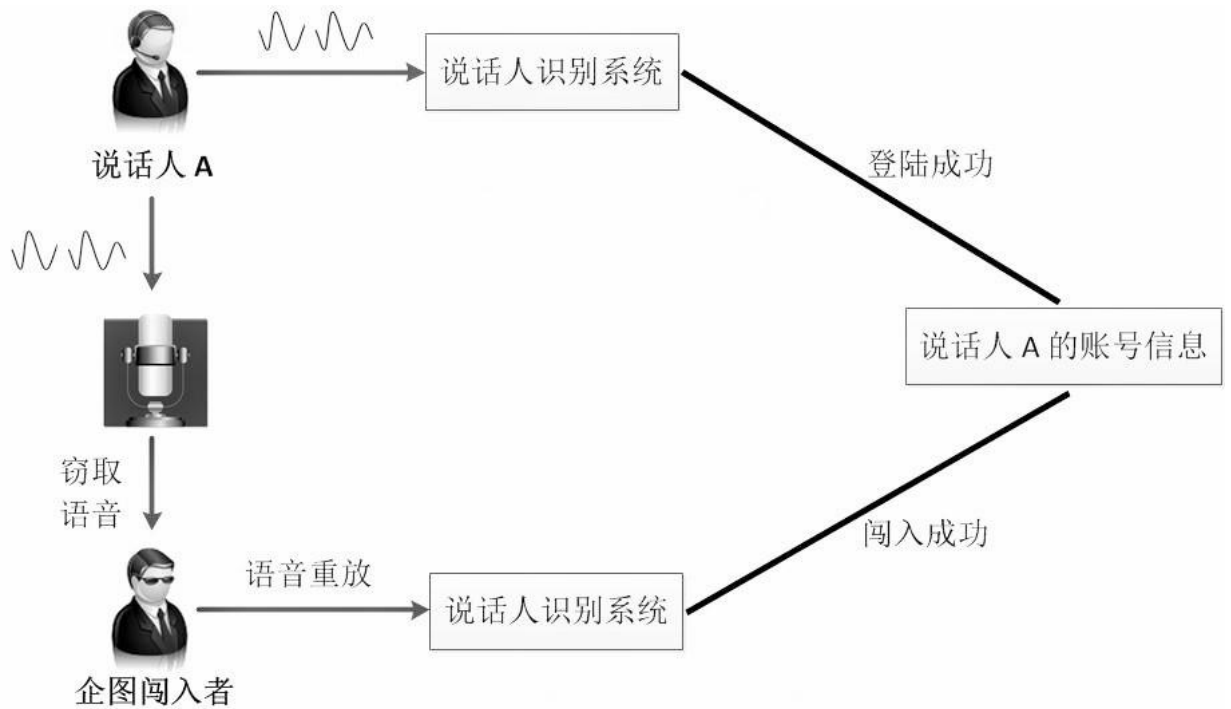


图 1

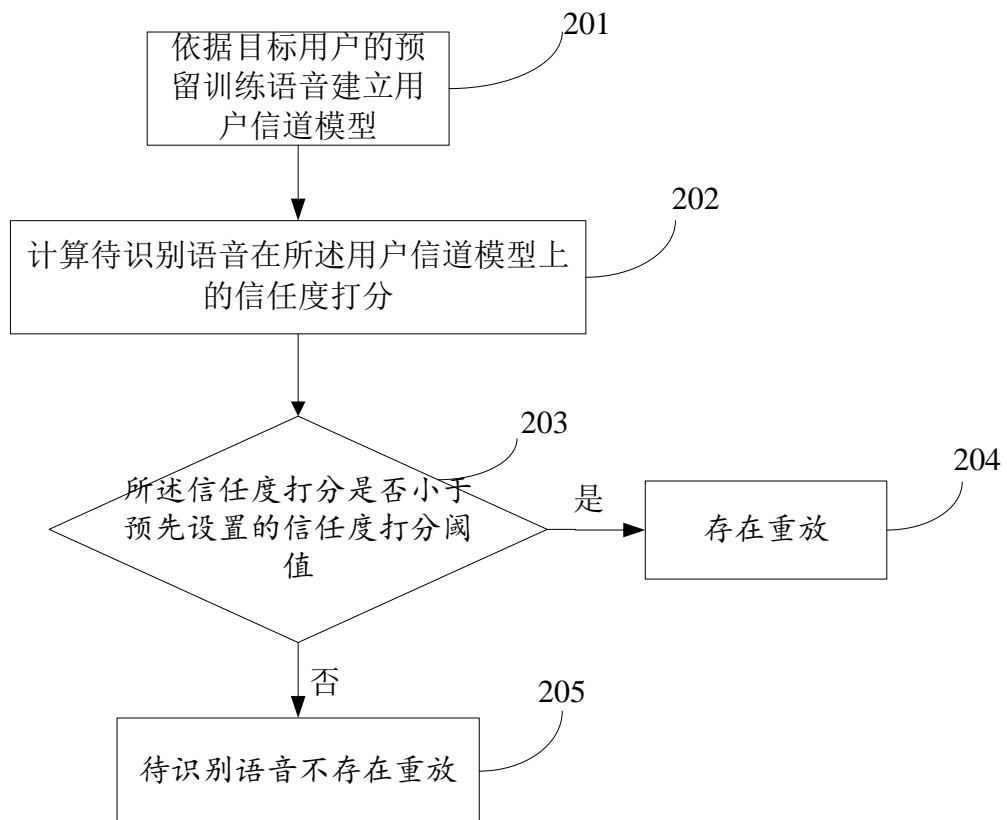


图 2

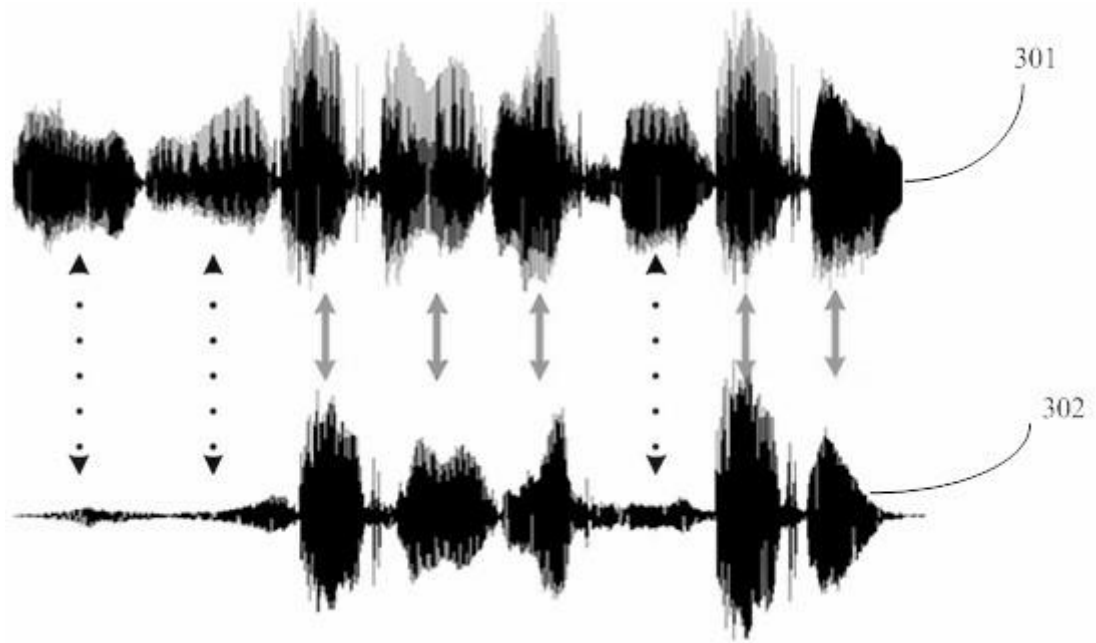


图 3

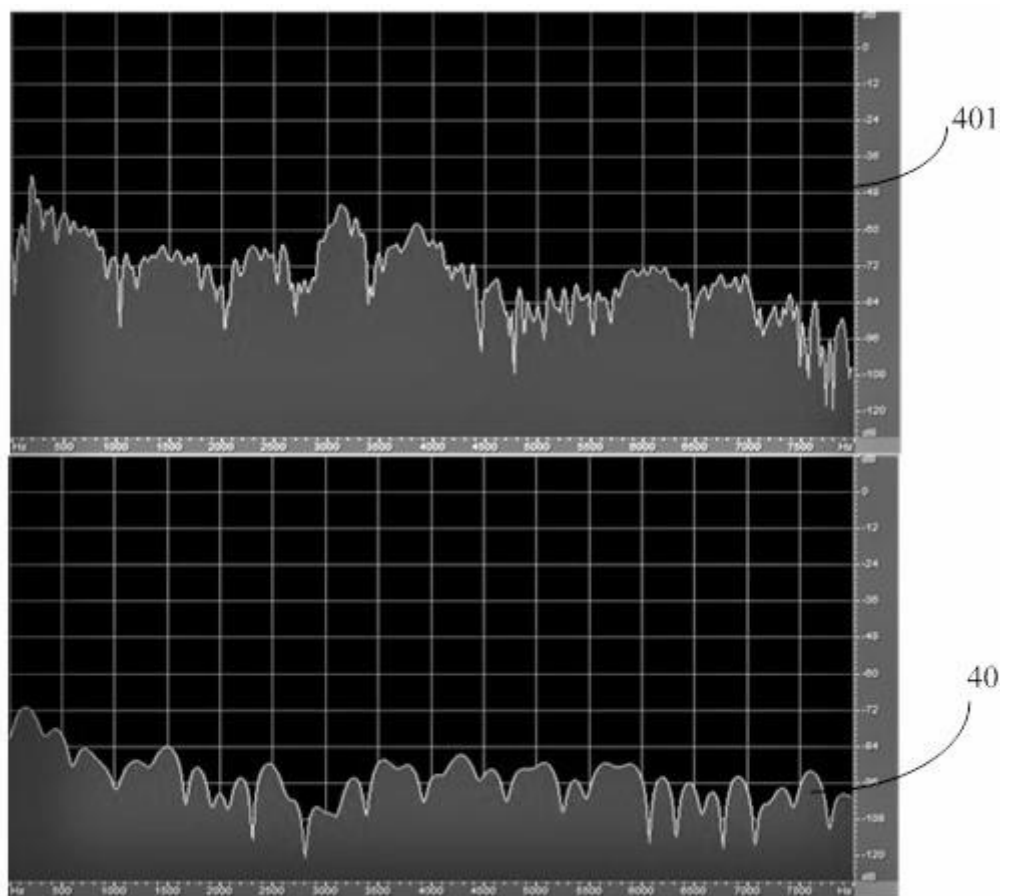


图 4

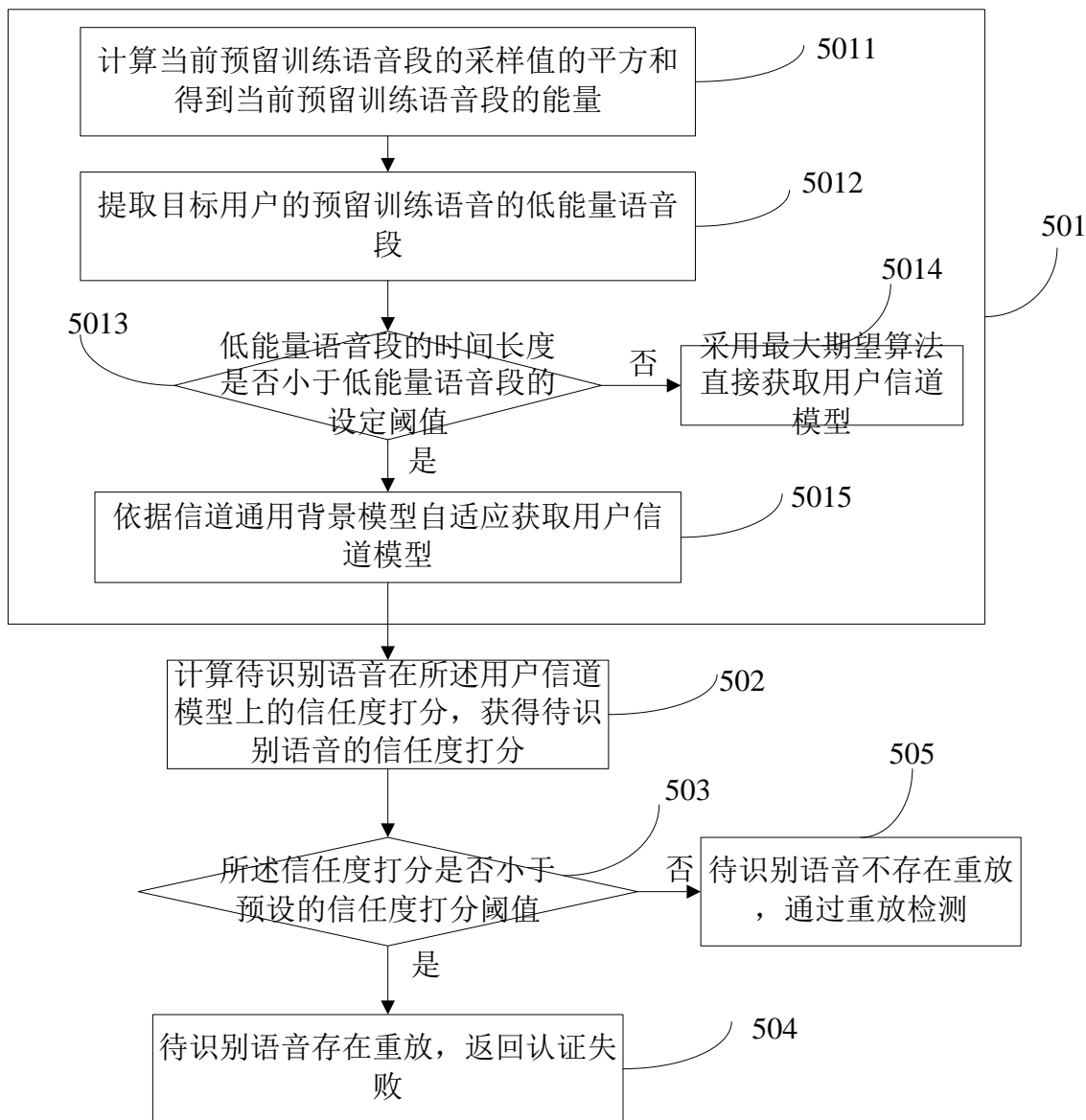


图 5

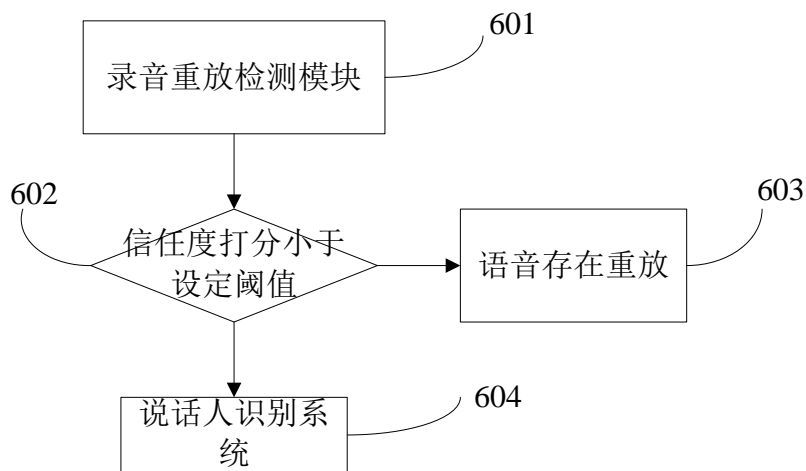


图 6

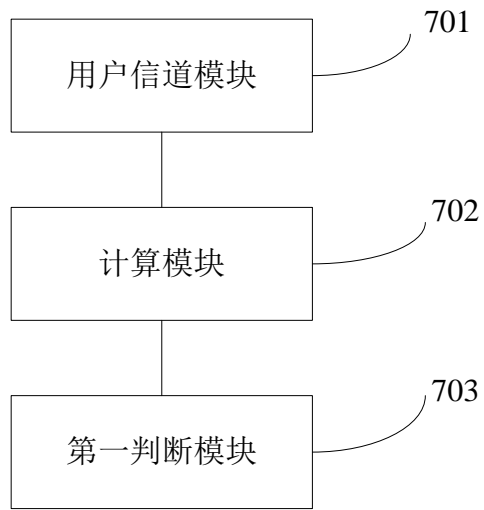


图 7