

## 申请人

郑方

清华大学西 14-4-202 邮编 100084

8289-6531(O)

6277-1030(H)

138-0101-2234(M)

fzheng@tsinghua.edu.cn

## 发明人:

王 军(Wang Jun): wangjun@csl.t.riit.tsinghua.edu.cn

王 东(Wang Dong): wangdong99@mails.tsinghua.edu.cn

邬晓钧(Wu Xiaojun): xjwu@tsinghua.edu.cn

郑 方(Zheng Fang): fzheng@tsinghua.edu.cn

## 摘要

本发明属于计算机及信息服务技术领域,涉及一种实现顺序自适应学习的说话人确认系统的方法。包括设置:一个包含说话人充分统计信息的说话人模型结构模块、一个顺序更新通用背景模型的算法模块、一个顺序更新说话人模型的算法模块。各模块预留有任务相关的操作接口。本发明针对实际应用环境下说话人信道易变或者不匹配的情况,极大提高说话人确认系统性能,高质量地建立相应的说话人确认系统。本发明可广泛用于金融、证券客户身份安全认证,刑事侦听,个性化服务等,使远程身份确认更稳定可靠。

摘要附图见图 3。

## 权利要求书

1、一种实现顺序自适应学习的说话人确认系统的方法,其特征在于,包括以下步骤:

- 1) 设置一个包含说话人充分统计信息的说话人模型结构模块,用于定义和保存每一位说话人充分统计信息,这些信息用于说话人模型的顺序更新;
- 2) 设置一个顺序更新通用背景模型的算法模块,在系统运行时利用顺序录入的说话人语音更新通用背景模型,使其信道逐渐收敛至新的信道;
- 3) 设置一个顺序更新说话人模型的算法模块,在系统运行时利用更新后的通用背景模型及说话人模型中保存的充分统计信息更新说话人模型,使其信道逐渐收敛至新的信道;
- 4) 设置一个顺序更新的说话人确认模块,利用顺序更新后的通用背景模型和说话人模型进行说话人确认;

上述 1) — 4) 中各模块预留有任务相关的操作接口。

2、如权利要求 1 所述的实现顺序自适应学习的说话人确认系统方法,其特征在于,顺序自适应学习的说话人确认实现方法包括以下步骤:

- 1-1) 对于每位新登陆的说话人,训练一个包含说话人充分统计信息的说话人模型;
- 1-2) 对于每位新登陆的说话人,提取其训练语音中的信道信息,利用这些信息更新系统的通用背景模型;
- 1-3) 利用系统中更新后的通用背景模型的均值信息和系统中每一个说话

- 人模型中保存的充分统计信息，更新相应的说话人模型；
- 1-4) 利用更新后通用背景模型和说话人模型进行说话人确认。

3、如权利要求 1 所述实现顺序自适应学习的说话人确认方法，其特征在于，所说的包含说话人充分统计信息的说话人模型，指在传统说话人模型中的均值和方差信息的基础上，定义和保存说话人充分统计信息，包括：说话人语音采样对于高斯混合模型中每一个混合贡献率的累加以及加权和。这些信息用于说话人模型的顺序更新。

4、如权利要求 1 所述的实现顺序自适应学习的说话人确认方法，其特征在于，所说的对话管理模块的实现方法为：

- 2-1) 从新登陆的说话人语音信息中，采用严格的先验和较少的语音信息重新估计通用背景模型参数，使其信道逐渐收敛趋于一致；
- 2-2) 重新训练说话人模型时，采用弱先验和较多的说话人语音信息估计说话人模型的参数，使其信道逐渐收敛趋于一致；
- 2-3) 利用更新过的通用背景模型和说话人模型中保存的充分统计信息，更新相应的说话人模型；

# 说明书

## 基于顺序更新模型的声纹识别系统

### 技术领域

本发明属于计算机及信息服务技术领域，特别涉及一种实现顺序自适应建模的声纹识别系统。

### 背景技术

声纹识别技术属于生物识别技术的一种，是利用声音波形中所包含的表征说话人个性特性的信息进行说话人身份确认的技术。随着技术的推进，声纹识别技术正广泛地应用到各个领域，金融认证、公安刑侦破案、国防监听、个性化服务等多个领域。

声纹识别过程主要包括声纹建模和识别。目前声纹建模通常采用高斯混合模型和通用背景模型（GMM-UBM）框架：利用数百人、男女均衡、信道匹配的说话人语音训练通用背景模型（UBM）做为基准；利用单个说话人数十秒的语音在通用背景模型上进行适应，得到反映该说话人特征的高斯混合模型（GMM）；通用背景模型（UBM）性能的优劣很大程度上决定了声纹识别系统的性能。

在实际声纹识别应用中，用户语音采集面临着信道（channel）变化的问题。比如用户由电话信道转而采用手机信道，移动基站调整信道参数等等。新登录的说话人语音应用于旧的通用背景模型直接导致系统性能的下降，而重新训练新的通用背景模型又面临着重新采集大量信道匹配的说话人数据的难题，不仅费用高，而且在大多数情况下是不可行的。目前，针对信道补偿的声纹识别的技术主要有：本征信道、因子分析等。这些方法均需要大量预知的信道数据进行计算。

本发明的目的是克服现有技术的不足，针对实际使用中说话人顺序录入的特点，实现基于顺序更新模型的声纹识别系统。可让系统无须重新训练通用背景模型，通过自适应自动更新至新的信道，从而保持系统性能。

### 发明内容

本发明的目的是克服现有技术的不足，针对实际使用中，说话人顺序录入的特点，通过研究说话人语音中包含的表征说话人个性特征的信息和表征信道特性的信息，建立顺序更新模型方法的框架和流程，设计了顺序自适应声纹识别算法。本发明可广泛用于构建声纹识别系统，具有较高的系统识别率和稳定性。

本发明提出一种基于顺序更新模型的声纹识别系统，其特征在于，包括以下步骤：

（1）对于每位新登陆的说话人，训练一个包含说话人充分统计信息的说话人模型；

(2) 对于每位新登陆的说话人，提取其训练语音中的信道信息，利用这些信息更新系统的通用背景模型；

(3) 利用系统中更新后的通用背景模型的均值信息和系统中每一个说话人模型中保存的充分统计信息，更新相应的说话人模型；

(4) 利用更新后通用背景模型和说话人模型进行说话人确认。

其中(1)训练一个包含说话人充分统计信息的说话人模型，其特征在于：

(1-1) 在目前通用的说话人模型结构的基础上，设置一个包含说话人充分统计信息的说话人模型结构模块，用于定义和保存每一位说话人充分统计信息，这些信息用于说话人模型的顺序更新；

(1-2) 所说的包含说话人充分统计信息的说话人模型，指在传统说话人模型中的均值和方差信息的基础上，定义和保存说话人充分统计信息，包括：说话人语音采样对于高斯混合模型中每一个混合贡献率的累加以及加权和。这些信息用于说话人模型的顺序更新。

(1-3) 所说的说话人充分统计信息定义如下：

$$r_i(c) = \frac{N(x_i; \mu_c, \sigma_c)}{\sum_m N(x_i; \mu_c, \sigma_c)} \quad (1)$$

$$r_c = \sum_i r_i(c) \quad (2)$$

$$z_c = \sum_i r_i(c) x_i \quad (3)$$

其中 $N(x_i; \mu_c, \sigma_c)$ 表示采样 $x_i$ 在单个高斯混合上的后验概率， $r_c$ 表示采样 $x_i$ 中每一个混合贡献率的累加， $z_c$ 表示采样 $x_i$ 中每一个混合贡献率的加权。

(1-4) 传统说话人模型结构图如附图 2(a) 所示，包含说话人充分统计信息的说话人模型结构如附图 2(b) 所示。

其中(2)对于每位新登陆的说话人，提取其训练语音中的信道信息，利用这些信息更新系统的通用背景模型，其特征在于：

(2-1) 设置一个顺序更新通用背景模型的算法模块，在系统运行时利用顺序录入的说话人语音更新通用背景模型，使其信道逐渐收敛至新的信道；

(2-2) 从新登陆的说话人语音信息中，采用严格的先验和较少的语音信息重新估计通用背景模型参数，使其信道逐渐收敛趋于一致；

(2-3) 重新训练说话人模型时，采用弱先验和较多的说话人语音信息估计说话人模型的参数，使其信道逐渐收敛趋于一致；

(2-4) 顺序更新的算法定义如下：

$$\mu_c = \frac{z_c + \frac{\sigma}{\hat{\sigma}} \hat{\mu}_c}{r_c + \frac{\sigma}{\hat{\sigma}}} \quad (4)$$

其中 $\mu_c$ 为适应后的模型均值参数， $\hat{\mu}_c$ 为初始模型均值参数。 $\hat{\sigma}$ 为(2-2)中所述的先验，严格的先验指 $\hat{\sigma}$ 取值较小，弱先验指 $\hat{\sigma}$ 取值较大。

其中(3)利用系统中更新后的通用背景模型的均值信息和系统中每一个说话人模型中保存的充分统计信息，更新相应的说话人模型，其特征在于：

(3-1) 设置一个顺序更新说话人模型的算法模块, 在系统运行时利用更新后的通用背景模型及说话人模型中保存的充分统计信息更新说话人模型, 使其信道逐渐收敛至新的信道;

(3-2) 利用更新过的通用背景模型和说话人模型中保存的充分统计信息, 更新相应的说话人模型;

其中(4)利用更新后通用背景模型和说话人模型进行说话人确认, 其特征在于:

(4-1) 设置一个顺序更新的声纹识别模块, 利用顺序更新后的通用背景模型和说话人模型进行说话人确认;

(4-2) 顺序更新模型的声纹识别系统系统流程图如图3所示。

本发明可用于构建声纹识别系统, 广泛应用于如金融认证、公安刑侦、个性服务等, 可有效提高声纹识别系统的识别率和可靠性。

#### 附图说明

图1为已有的典型基于通用背景模型-高斯混合模型框架的声纹识别系统的模块组成。

图2(a)为传统说话人模型结构图。

图2(b)包含说话人充分统计信息的说话人模型结构图。

图3为本发明的基于顺序更新模型的声纹识别系统的模块组成。

图4为本发明顺序更新通用背景模型的实现流程框图。

附图:

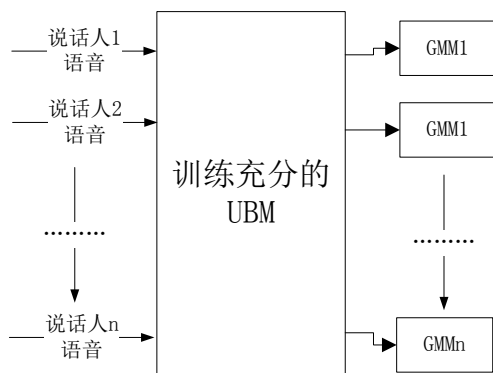


图 1

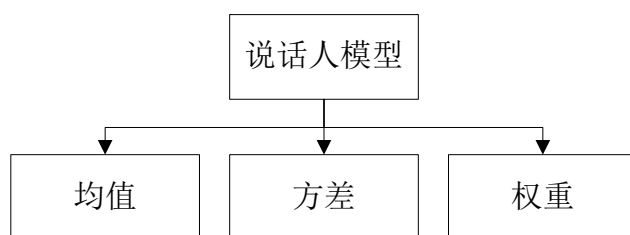


图 2 (a)

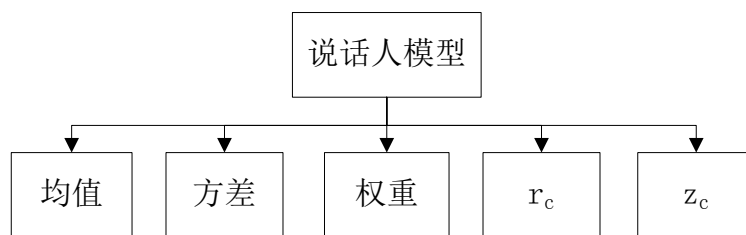


图 2 (b)

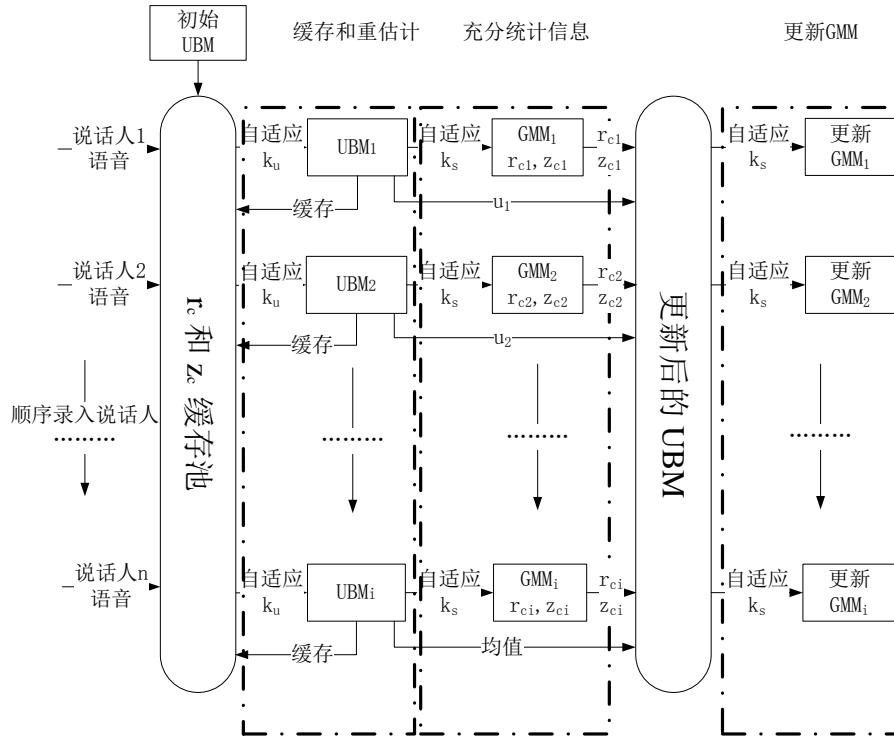


图 3

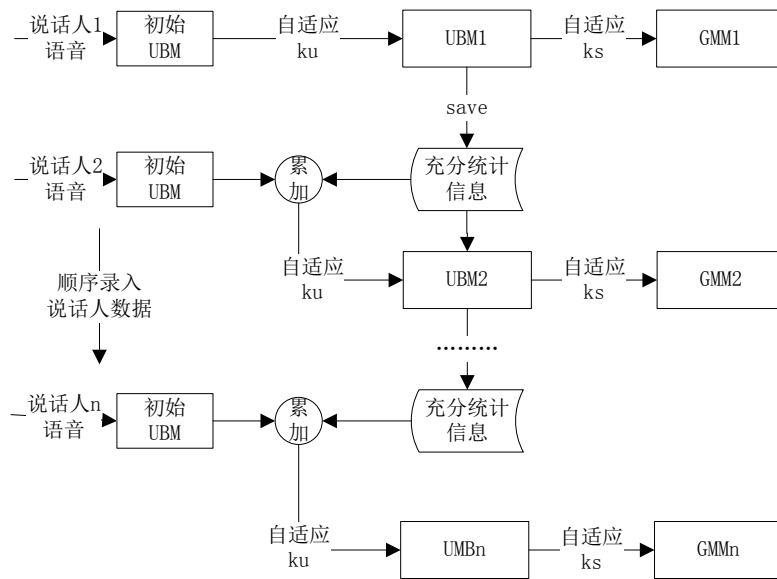


图 4