

(12) 发明专利申请

(10) 申请公布号 CN 103229233 A

(43) 申请公布日 2013. 07. 31

(21) 申请号 201080070321. 9

(51) Int. Cl.

(22) 申请日 2010. 12. 10

G1OL 15/14 (2006. 01)

(85) PCT申请进入国家阶段日

2013. 05. 24

(86) PCT申请的申请数据

PCT/CN2010/079650 2010. 12. 10

(87) PCT申请的公布数据

W02012/075640 EN 2012. 06. 14

(71) 申请人 松下电器产业株式会社

地址 日本大阪府

(72) 发明人 沈海峰 马龙 张丙奇

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 邸万奎

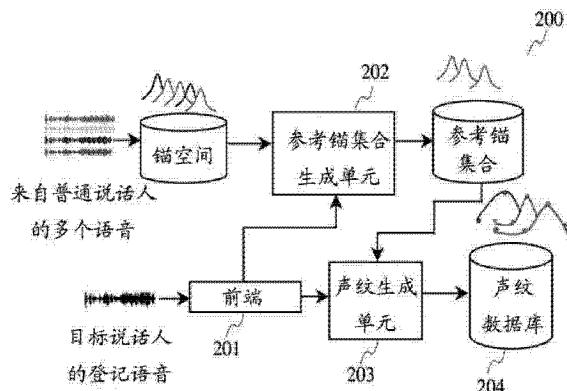
权利要求书3页 说明书10页 附图6页

(54) 发明名称

用于识别说话人的建模设备和方法、以及说话人识别系统

(57) 摘要

本发明实现用于识别说话人的建模设备和方法、以及说话人识别系统。建模设备包括：前端，从各目标说话人取得登记语音数据；参考锚集合生成单元，基于锚空间使用登记语音数据生成参考锚集合；以及声纹生成单元，基于参考锚集合和登记语音数据生成声纹。在本公开中，通过考虑登记语音和说话人自适应技术，能够生成尺寸更小的锚模型，因而能够进行具有尺寸更小的参考锚集合的、可靠性高的鲁棒的说话人识别。这对于进行计算速度的改善以及大幅度的存储器削减是非常有利的。



1. 用于识别说话人的建模设备,其包括 :

前端,从目标说话人取得登记语音 ;

参考锚集合生成单元,基于锚空间使用所述登记语音生成参考锚集合 ;以及声纹生成单元,基于所述参考锚集合和所述登记语音生成声纹。

2. 如权利要求 1 所述的用于识别说话人的建模设备,

所述参考锚集合包括主锚和同伴锚,所述同伴锚基于所述主锚生成。

3. 如权利要求 2 所述的用于识别说话人的建模设备,

所述锚空间包含多个锚模型,通过从所述锚空间中找到距登记语音的所述模型的距离最近的锚模型,生成所述主锚。

4. 如权利要求 2 或 3 所述的用于识别说话人的建模设备,

基于所述主锚将所述锚空间分割为多个集群,找出距不包含主锚的所述多个集群的重心的距离最近的锚模型,由此生成所述同伴锚。

5. 如权利要求 4 所述的用于识别说话人的建模设备,

所述锚空间通过如下步骤分割为所述多个集群,即 :最初,令 N 为与所述主锚的个数相等的值,将所述锚空间分割为 N 个集群 ;找到要进一步进行分割的、在所述 N 个集群中具有最大的类内距离的一个集群,在所述集群具有主锚的情况下,基于所述主锚与所述集群内的距所述主锚最远的锚,将所述集群分割为两个子集群,而在所述集群不具有主锚的情况下,基于相对于距离更远的所述主锚的、所述集群内最远的两个锚,将所述集群分割为两个子集群 ;以及反复执行上述过程,直到得到所述多个集群为止。

6. 如权利要求 3 所述的用于识别说话人的建模设备,

所述距离为似然值。

7. 如权利要求 2 所述的用于识别说话人的建模设备,

所述主锚通过 MAP 或 MLLR 等自适应方法利用所述登记语音进行自适应,所述同伴锚在自适应后基于所述主锚生成。

8. 如权利要求 2 所述的用于识别说话人的建模设备,

所述主锚通过 MAP 或 MLLR 等自适应方法利用所述登记语音进行自适应,所述同伴锚和所述进行了自适应的主锚组合为所述参考锚集合。

9. 用于识别说话人的建模方法,其包括 :

从目标说话人取得登记语音的步骤 ;

基于锚空间使用所述登记语音生成参考锚集合的步骤 ;以及

基于所述参考锚集合和所述登记语音生成声纹的步骤。

10. 如权利要求 9 所述的用于识别说话人的建模方法,

所述参考锚集合包括主锚和同伴锚,该建模方法还包括基于所述主锚生成所述同伴锚的步骤。

11. 如权利要求 10 所述的用于识别说话人的建模方法,

所述锚空间包含多个锚模型,该建模方法还包括从所述锚空间中找到距登记语音的所述模型的距离最近的锚模型从而生成所述主锚的步骤。

12. 如权利要求 10 或 11 所述的用于识别说话人的建模方法,还包括 :

基于所述主锚将所述锚空间分割为多个集群,找出距不包含主锚的所述多个集群的重

心的距离最近的锚模型,由此生成所述同伴锚的步骤。

13. 如权利要求 12 所述的用于识别说话人的建模方法,还包括:

最初,令 N 为与所述主锚的个数相等的值,将所述锚空间分割为 N 个集群的步骤;

找到要进一步进行分割的、在所述 N 个集群中具有最大的类内距离的一个集群的步骤,在该步骤中,在所述集群具有主锚的情况下,基于所述主锚与所述集群内的距所述主锚最远的锚,将所述集群分割为两个子集群,而在所述集群不具有主锚的情况下,基于相对于距离更远的所述主锚的、所述集群内最远的两个锚,将所述集群分割为两个子集群;以及

反复执行上述过程,直到得到所述多个集群为止的步骤。

14. 如权利要求 11 所述的用于识别说话人的建模方法,

所述距离为似然值。

15. 如权利要求 10 所述的用于识别说话人的建模方法,还包括:

使所述主锚通过 MAP 或 MLLR 等自适应方法利用所述登记语音进行自适应的步骤;以及  
在自适应后基于所述主锚生成所述同伴锚的步骤。

16. 如权利要求 10 所述的用于识别说话人的建模方法,还包括:

使所述主锚通过 MAP 或 MLLR 等自适应方法利用所述登记语音进行自适应的步骤;以及  
使所述同伴锚和所述进行了自适应的主锚组合为所述参考锚集合的步骤。

17. 说话人识别系统,包括:

前端,从目标说话人取得登记语音及 / 或测试语音;

参考锚集合生成单元,基于锚空间使用所述登记语音生成参考锚集合;

声纹生成单元,基于所述参考锚集合以及所述登记语音及 / 或测试语音生成声纹;

对比单元,将由所述测试语音生成的所述声纹与由所述登记语音生成的所述声纹进行  
比较;以及

判断单元,基于所述比较结果识别所述目标说话人的同一性。

18. 如权利要求 17 所述的说话人识别系统,

所述参考锚集合包括主锚和同伴锚,所述同伴锚基于所述主锚生成。

19. 如权利要求 18 所述的说话人识别系统,

所述锚空间包含多个锚模型,通过从所述锚空间中找到距登记语音的所述模型的距离  
最近的锚模型,生成所述主锚。

20. 如权利要求 18 或 19 所述的说话人识别系统,

基于所述主锚将所述锚空间分割为多个集群,找出距不包含主锚的所述多个集群的重  
心的距离最近的锚模型,由此生成所述同伴锚。

21. 如权利要求 20 所述的说话人识别系统,

所述锚空间通过如下步骤分割为所述多个集群,即:最初,令 N 为与所述主锚的个数相  
等的值,将所述锚空间分割为 N 个集群;找到要进一步进行分割的、在所述 N 个集群中具有  
最大的类内距离的一个集群,在所述集群具有主锚的情况下,基于所述主锚与所述集群内的  
距所述主锚最远的锚,将所述集群分割为两个子集群,而在所述集群不具有主锚的情况  
下,基于相对于距离更远的所述主锚的、所述集群内最远的两个锚,将所述集群分割为两个  
子集群;以及反复执行上述过程,直到得到所述多个集群为止。

22. 如权利要求 19 所述的说话人识别系统,

所述距离为似然值。

23. 如权利要求 18 所述的说话人识别系统，

所述主锚通过 MAP 或 MLLR 等自适应方法利用所述登记语音进行自适应，所述同伴锚在自适应后基于所述主锚生成。

24. 如权利要求 18 所述的说话人识别系统，

所述主锚通过 MAP 或 MLLR 等自适应方法利用所述登记语音进行自适应，所述同伴锚和所述进行了自适应的主锚组合为所述参考锚集合。

25. 说话人识别系统，包括建模设备和识别设备，

所述建模设备包括：

第一前端，从目标说话人取得登记语音；

参考锚集合生成单元，基于锚空间使用所述登记语音生成参考锚集合；以及

第一声纹生成单元，基于所述参考锚集合和所述登记语音生成第一声纹，

所述识别设备包括：

第二前端，从所述目标说话人取得测试语音；

第二声纹生成单元，基于所述参考锚集合和所述测试语音生成第二声纹；

对比单元，将所述第一声纹与所述第二声纹进行比较；以及

判断单元，基于比较结果识别所述目标说话人的同一性。

## 用于识别说话人的建模设备和方法、以及说话人识别系统

### 技术领域

[0001] 本公开涉及语音(音频)处理以及语音识别技术,另外涉及说话人对比、电话会议、以及数字网络视听的技术。

### 背景技术

[0002] 说话人识别技术对于许多应用,例如说话人跟踪、语音索引(audio index)、以及分段是非常有用的。近年来,提出了使用多个锚(说话人)模型对说话人进行建模的技术。将说话人语音投影到锚模型上,构成表示说话人的声学特性的向量。

[0003] 图1表示用于进行说话人识别的以往设备的方框图。如图1所示,通过学习来自多个普通说话人的语音,生成锚空间。在参考锚集合生成单元(reference anchor set generation unit)102中,从锚空间(anchor space)选择作为集群(cluster)的重心的多个虚拟锚说话人(virtual anchor speakers)并形成参考锚集合(reference anchor set),或者选择距各集群的重心最近的锚说话人并形成参考锚集合。前端101取得目标说话人的登记语音,将该登记语音转换为特征参数,并将这些特征参数发送至声纹(voice print)生成单元103。声纹生成单元103基于从前端101发送来的特征参数以及由参考锚集合生成单元102生成的参考锚集合,生成声纹。接着,为了进一步用于说话人识别,将生成的声纹存储到声纹数据库104中。

### 发明内容

[0004] 发明要解决的问题

[0005] 根据图1可知,由设备100生成的参考锚集合仅能够反映锚空间自身的分布。因此,为了更好地表现目标说话人,需要更多的锚,因此计算量增大,难以在嵌入型系统中使用。

[0006] 解决问题的方案

[0007] 在本公开的一个方式中,实现用于识别说话人的建模设备,该设备包括:前端,从目标说话人取得登记语音;参考锚集合生成单元,基于锚空间使用登记语音生成参考锚集合;以及声纹生成单元,基于参考锚集合和登记语音生成声纹。

[0008] 在本公开的另一个方式中,提供用于识别说话人的建模方法,该方法包括:从目标说话人取得登记语音的步骤;基于锚空间使用登记语音生成参考锚集合的步骤;以及基于参考锚集合和登记语音生成声纹的步骤。

[0009] 在本公开的又一方式中,实现说话人识别系统,该系统包括:前端,从目标说话人取得登记语音及/或测试语音;参考锚集合生成单元,基于锚空间使用登记语音生成参考锚集合;声纹生成单元,基于参考锚集合以及登记语音及/或测试语音生成声纹;对比单元,将根据测试语音生成的声纹与根据登记语音生成的声纹进行比较;以及判断单元,基于比较结果识别目标说话人的同一性。

[0010] 在本公开的再一方式中,实现说话人识别系统,该系统包括建模设备以及识别设

备,建模设备包括:第一前端,从目标说话人取得登记语音;参考锚集合生成单元,基于锚空间使用登记语音生成参考锚集合;以及第一声纹生成单元,基于参考锚集合和登记语音生成第一声纹,识别设备包括:第二前端,从目标说话人取得测试语音;第二声纹生成单元,基于参考锚集合和测试语音生成第二声纹;对比单元,将第一声纹与第二声纹进行比较;以及判断单元,基于比较结果识别目标说话人的同一性。

[0011] 使用本公开的建模设备、方法、以及说话人识别系统,考虑登记语音以及说话人自适应技术,从而能够生成尺寸更小的锚模型,能够进行具有尺寸更小的参考锚集合的、可靠性高的鲁棒的说话人识别。这对于进行计算速度的改善以及大幅度的存储器削减是非常有利的,因此计算量较少且参考锚集合较小,故更适于嵌入型应用。

[0012] 上述为概要内容,因此当然进行了简化、一般化,并且省略了详细情况,因此本领域技术人员可以理解,概要内容仅为示例,并不意图以任何形式进行限制。本说明书中记载的设备及/或过程及/或其他主题的其他方式、特征、以及优点将通过本说明书中描述的内容而变得明确。本“发明内容”用于导入以下的“具体实施方式”中进一步说明的简化形式的概念的选择。本“发明内容”并不意图明示权利要求的主题的关键特征或本质特征,也并不意图用于辅助决定权利要求的主题的范围。

[0013] 通过参考附图,利用以下的说明以及附属的权利要求可以使本公开的上述特征以及其他特征完全明确。这些附图仅表示基于本公开的多个实施方式,因此应当理解,不应将其认为是限制本公开的范围的内容,并且,使用附图更具体地、详细地说明本公开。

## 附图说明

- [0014] 图1是用于说话人识别的以往设备的方框图。
- [0015] 图2是基于本公开的一实施方式的用于说话人识别的建模设备的方框图。
- [0016] 图3是基于本公开的一实施方式的参考锚集合生成单元的方框图。
- [0017] 图4是基于本公开的一实施方式的锚空间的略图。
- [0018] 图5是基于本公开的另一实施方式的参考锚集合生成单元的方框图。
- [0019] 图6是基于本公开的又一实施方式的参考锚集合生成单元的方框图。
- [0020] 图7是基于本公开的一实施方式的用于说话人识别的建模方法的流程图。
- [0021] 图8是基于本公开的另一实施方式的用于生成参考锚集合的建模方法的流程图。
- [0022] 图9是基于本公开的一实施方式的说话人识别系统的方框图。
- [0023] 图10是表示关于与尺寸不同的参考锚集合的性能比较的实验数据的图。

## 具体实施方式

[0024] 在下面的详细说明中参考构成详细说明的一部分的附图。图中类似的标号典型地标识类似的成分,除非在上下文中另行说明。详细的说明、附图、以及权利要求中说明的例示的实施方式并不意图进行限定。也能够利用其他实施方式,另外在不脱离本说明书提出的主要的精神和范围的情况下,可以增加其他变形。容易理解的是,在本说明书中进行一般的说明并在图中例示的本公开的方式可以通过各种不同结构进行配置、置换、组合以及设计,它们均被明确地考察并构成本公开的一部分。

[0025] 下面介绍本公开中使用的主要用语。

[0026] 1) 锚数据库

[0027] 一般而言,学习体(corpus)中有来自数百或数千个说话人的语音数据。能够学习表示各说话人的声学特性的锚模型(例如高斯混合模型)。若汇总所有锚模型,则构成锚数据库。

[0028] 2) 参考锚集合

[0029] 将参考锚集合定义为用于说话人鉴别 / 识别系统的、按照特定的生成规则从锚数据库生成的集合。

[0030] 3) 锚空间

[0031] 在各锚模型表示空间的一维的情况下,在使用锚数据库内的所有锚模型时可构成锚空间。锚空间的维数等于锚数据库内的锚模型的总数。

[0032] 4) 主锚模型(Principal anchor model)

[0033] 将主锚模型定义为相对于一个说话人的登记语音最近的模型。

[0034] 5) 同伴锚模型(Associate anchor model)

[0035] 将除了主锚以外的、参考锚集合内的剩余的锚定义为同伴锚模型。

[0036] 图 2 表示基于本公开的一实施方式的用于说话人识别的建模设备的方框图。

[0037] 如图 2 所示,基于本公开的一实施方式的用于说话人识别的建模设备 200 包括前端 201、参考锚集合生成单元 202、声纹生成单元 203、以及声纹数据库 204。例如,在上述结构中,声纹生成单元 203 连接于前端 201、参考锚集合生成单元 202、以及声纹数据库 204。前端 201 也与参考锚集合生成单元 202 连接。

[0038] 根据本公开的一实施方式,在建模设备 200 中,前端 201 从目标说话人取得登记语音,参考锚集合生成单元 202 基于锚空间使用登记语音生成参考锚集合,声纹生成单元 203 基于参考锚集合以及登记语音生成声纹。

[0039] 以下说明基于本公开的实施方式的建模设备 200 的上述各个单元的操作。

[0040] 最初,通过学习来自普通说话人的多个语音来生成锚空间,锚空间包含表示这些普通说话人的声学特性的多个锚模型。锚空间能够以锚数据库的形式存储在数据库内。前端 201 取得目标说话人的登记语音,将该登记语音转换为特征参数,并将这些特征参数发送至参考锚集合生成单元 202 以及声纹生成单元 203。参考锚集合生成单元 202 基于目标说话人的登记语音,从锚空间中生成参考锚集合。声纹生成单元 203 将从登记语音中提取出、并从前端 201 发送来的特征参数适用于由参考锚集合生成单元 202 生成的参考锚集合的锚模型,由此生成声纹,为了进一步用于识别说话人,将生成的声纹存储到声纹数据库 204 中。

[0041] 前端 201 以及声纹生成单元 203 的操作是本领域技术人员所熟知的,因而为了不与本公开的要点相混淆,省略其详细说明。以下,详细说明参考锚集合生成单元 202 的操作。

[0042] 根据本公开的另一实施方式,由参考锚集合生成单元 202 生成的参考锚集合包括至少一个主锚和至少一个同伴锚,参考锚集合生成单元 202 基于锚空间使用登记语音生成主锚,并基于主锚生成至少一个同伴锚。

[0043] 假定在存在“n”个目标说话人的登记语音的情况下,生成的主锚的数量能够设为 1 至 n,较为理想的是,基于来自 n 个目标说话人的登记语音生成 n 个主锚。

[0044] 图 3 表示基于本公开的一实施方式的参考锚集合生成单元的方框图。参考锚集合生成单元 302 的功能与图 2 所示的参考锚集合生成单元 202 的功能相同。如图 3 所示，参考锚集合生成单元 302 包括主锚生成单元 3021、同伴锚生成单元 3022、以及组合单元 3023。

[0045] 具体而言，主锚生成单元 3021 从锚空间中找到与目标说话人的登记语音的距离最近的锚模型，由此生成主锚。在此，登记语音的模型既可以是表示目标说话人的声学特性的 GMM(高斯混合模型)，也可以使用目标说话人的登记语音的特征参数而由前端 201 或者参考锚集合生成单元 302 进行计算。根据目标说话人的人数生成至少一个主锚。同伴锚生成单元 3022 基于主锚将锚空间分割为多个集群，找出距不包含主锚的多个集群的重心的距离最近的锚模型，由此生成同伴锚。在此，距离也可以是似然值。组合单元 3023 组合所生成的主锚与同伴锚，以作为参考锚集合。

[0046] 以下说明基于本公开的一实施方式的参考锚集合生成单元 302 的上述各个单元的操作。

[0047] 最初，在主锚生成单元 3021 中，通过使用目标说话人的登记语音来生成主锚。具体而言，找到锚空间内与目标说话人的登记语音的 GMM 模型的距离最近的锚模型，由此能够得到对于一个目标说话人的主锚。更具体而言，根据本公开的一实施方式，能够使用似然距离来作为上述距离。

[0048] 例如，使用来自目标说话人的登记语音，计算目标说话人与锚空间内存在的各普通说话人之间的似然距离。最初，假定锚空间内的各说话人  $i$  通过以  $\{p_{ij}, \mu_{ij}, \Sigma_{ij}\}$ ,  $j=1, 2, \dots, M_i$  为参数的 GMM (高斯混合模型)  $\Lambda_i$  等概率模型进行建模，式中， $M_i$  表示混合成分的个数， $p_{ij}$  表示混合增益， $\mu_{ij}, \Sigma_{ij}$  分别是各高斯混合成分的平均向量和对角协方差矩阵。目标说话人的登记语音  $X$  具有语音帧  $[x_1, x_2, x_3, \dots, x_T]$ ，式中， $T$  是总帧数。因此，似然距离可作为下式 (1) 得到。

$$[0049] P(X | \Lambda_i) = \prod_{l=1}^T \sum_{j=1}^{M_i} p_{ij} \frac{1}{\sqrt{2\pi\Sigma_{ij}}} \exp \left[ -\frac{1}{2} (x_l - \mu_{ij}) \Sigma_{ij}^{-1} (x_l - \mu_{ij})^\top \right] \quad (1)$$

[0050] 随后，例如通过计算  $\arg \max_i P(X | \Lambda_i)$ ，求出普通说话人的最近说话人模型作为主锚。需要注意的是，在存在多个目标说话人的情况下，可以反复上述操作以得到多个主锚。因此，生成的主锚的数量并不是本公开的范围的限制因素。另外，距离并不限定于似然距离。

[0051] 接着，在同伴锚生成单元 3022 中，基于由主锚生成单元 3021 生成的主锚，生成同伴锚。

[0052] 根据本公开的一实施方式，将锚空间分割为多个集群的过程如下。最初，将锚空间分割为  $N$  个集群，在此， $N$  等于主锚的个数。 $N$  个集群中具有最大的类内距离的一个集群被进一步分割，在此，在集群具有主锚的情况下，基于主锚与集群内的距主锚最远的锚，将集群分割为两个子集群，而在集群不具有主锚的情况下，基于相对于距离更远的主锚的、集群内最远的两个锚，将集群分割为两个子集群。反复上述过程，直到得到多个集群。

[0053] 具体而言，最初，按照锚空间内的各锚与所生成的主锚之间的距离，将锚空间分割为  $N$  个集群( $N$  等于主锚的个数)。得到的  $N$  个集群分别包含由主锚生成单元 3021 生成的一个主锚。在此，距离也可以是似然值。例如，可以使用 Kullback-Leibler (KL) 散度距离或者欧氏距离作为上述距离。以欧氏距离为例， $\Lambda_1$  以及  $\Lambda_2$  这两个 GMM 之间的欧氏距离作

为下式(2)给出。

$$[0054] \quad e = \int \Lambda_1^2 dx + \int \Lambda_2^2 dx - 2 \int \Lambda_1 \Lambda_2 dx$$

[0055]

$$= \sum_{i=1}^I \sum_{j=1}^J p_i p_j Q_{i,j,1,1} + \sum_{i=1}^I \sum_{j=1}^J p_i p_j Q_{i,j,2,2} - 2 \sum_{i=1}^I \sum_{j=1}^J p_i p_j Q_{i,j,1,2} \quad (2)$$

[0056] 式中,

[0057]

$$Q_{i,j,k,m} = \int N_1(x; \mu_{ki}, \Sigma_{ki}) N_1(x; \mu_{mj}, \Sigma_{mj}) dx \quad (3)$$

[0058] 在此, k, m=1, 2 仅表示两个 GMM 的索引。

[0059] 接着, 反复执行找到要进一步分割的具有最大类内距离的一个集群的操作, 直到满足全部集群为止。集群的个数可根据实际应用的条件预先确定。进行分割的步骤基于以下规则。

[0060] 在被选作分割对象的集群具有主锚的情况下, 找到该集群内的距主锚最远的锚, 为了进行分类, 将该集群内的剩余的锚分别与两个锚(距离最远的锚与主锚)进行比较, 基于比较结果将该集群分割为两个子集群。在与此不同的情况下

[0061] 在被选作分割对象的集群不具有主锚的情况下, 找到相对于更远的主锚最远的、集群内的两个锚, 即分割集群的更远集群中包含的主锚, 为了进行分类, 将该集群内的剩余的锚分别与两个锚进行比较, 基于比较结果将该集群分割为两个子集群。

[0062] 例如, 如上所述, 按照两个锚之间的距离执行比较。如上所述, 反复执行上述操作后, 将锚空间分割为期望数量的集群。

[0063] 随后, 作为同伴锚, 找到距包含主锚的集群以外的集群的重心最近的锚。根据本公开的一实施方式, 能够按照带比例尺的巴特查理亚(Bhattacharyya)距离得到重心。例如, 为了在各集群内得到各重心, 组合 GMM 的距离尺度基于以下作为式(4)给出的带比例尺的巴特查理亚距离。

[0064]

$$B_{SD} = B_{scale} B_{distance} = B_{scale} \left[ -\log \int \sqrt{\Lambda_i \Lambda_j} dx \right] \quad (4)$$

[0065] 在距重心最近的锚的选择中, 例如可以使用上述的 KL/ 欧氏距离。

[0066] 最后, 在组合单元 3023 中, 为了进一步用于说话人识别, 作为参考锚集合, 组合由主锚生成单元 3021 生成的主锚以及由同伴锚生成单元 3022 生成的同伴锚。

[0067] 图 4 表示基于本公开的一实施方式的锚空间 400 的略图。在图 4 中, 以两个目标说话人的情况为例, 图示了图 3 的参考锚集合生成单元 302 的处理之后的结果。如图 4 所示, 在锚空间 400 内, 共有 6 个分割出的集群 401、402、403、404、405 以及 406。如上述那样基于说话人 1 和说话人 2 的登记语音数据 4012 和 4042, 由主锚生成单元 3021 生成分别配置在集群 401 以及 404 内的两个主锚 4011 以及 4041。通过基于两个主锚 4011 以及 4041 的上述分割操作, 同伴锚生成单元 3022 生成分别配置在不包含主锚的集群 402、403、405 以及 406 中的 4 个同伴锚 4021、4031、4051 以及 4061。两个主锚 4011、4041 和 4 个同伴锚 4021、4031、4051 以及 4061 的组合是由参考锚集合生成单元 302 生成的参考锚集合。需要注意的

是,分割集群的总数并不限于 6 个,另外,本领域技术人员应当理解,集群的总数能够适当设定为其他任意数值。

[0068] 根据本公开的一实施方式,参考锚集合生成单元还可以包括自适应单元。图 5 表示基于本公开的一实施方式的建模设备 200 内的参考锚集合生成单元 502 的方框图。

[0069] 根据图 5 可知,参考锚集合生成单元 502 与图 3 的参考锚集合生成单元 302 的不同之处在于,在主锚生成单元 3021 与同伴锚生成单元 3022 之间增加了自适应单元 5024。

[0070] 具体而言,由主锚生成单元 3021 生成的主锚输入到自适应单元 5024 中。在自适应单元 5024 中,可以将 MLLR (最大似然线性回归)/MAP (最大后验概率) 等说话人自适应技术应用于生成的主锚,并将进行了自适应后的精细化了的主锚上输出到同伴锚生成单元 3022。在同伴锚生成单元 3022 中,由精细化了的主锚进行引导,分割锚空间,并按照前面说明的操作找到同伴锚。由于自适应技术是一般性技术,所以其详细内容不在下面进行说明。

[0071] 即,可以通过自适应单元 5024 中的 MAP 或 MLLR 等自适应方法利用登记语音使主锚进行自适应,并且可以在同伴锚生成单元 3022 中的自适应后基于主锚生成同伴锚。

[0072] 需要注意的是,并不限于在生成的主锚上应用自适应,也可以将其替代而在生成的参考锚集合上应用自适应。

[0073] 图 6 表示基于本公开的另一实施方式的建模设备 200 内的参考锚集合生成单元的方框图。自适应单元可以直接配置在主锚生成单元与参考锚集合生成单元的组合单元之间。

[0074] 根据图 6 可知,参考锚集合生成单元 602 与图 5 的参考锚集合生成单元 502 的不同之处在于,在主锚生成单元 3021 与组合单元 3023 之间,并列配置了自适应单元 5024 以及同伴锚生成单元 3022。

[0075] 具体而言,图 6 所示的各个单元的操作如下。由主锚生成单元 3021 生成主锚后,得到的主锚分别输入到自适应单元 5024 以及同伴锚生成单元 3022 中。在自适应单元 5024 中,为了进行自适应处理,可以在主锚上应用 MLLR/MAP 等若干说话人自适应技术。在同伴锚生成单元 3022 中,由所生成的主锚进行引导,分割锚空间,并按照参考图 3 在前面说明的操作找到同伴锚。接着,从自适应单元 5024 输出的进行了自适应的主锚以及从同伴锚生成单元 3022 输出的同伴锚为了进行组合处理而输入到组合单元 3023,通过组合进行了自适应的主锚以及同伴锚得到精细化的参考锚集合,为了生成目标说话人的声纹,或者为了进一步用于说话人识别,将精细化的参考锚集合输入声纹生成单元 203。

[0076] 图 7 表示基于本公开的一实施方式的用于说话人识别的方法的流程图。

[0077] 如图 7 所示,基于本公开的实施方式的用于说话人识别的方法可以包含以下步骤。在步骤 S701 中,从目标说话人取得登记语音。在步骤 S702 中,基于锚空间使用登记语音,从而生成参考锚集合。在步骤 S703 中,基于参考锚集合以及登记语音,生成目标说话人的声纹。根据本发明的实施方式,步骤 S701 可以由前端 201 执行,步骤 S702 如上所述可以由参考锚集合生成单元 202、302、502、以及 602 中的任一者执行,步骤 S703 可以由声纹生成单元 203 执行。

[0078] 图 8 表示基于本公开的一实施方式的用于生成参考锚集合的方法的流程图。如上所述,参考锚集合包括主锚和同伴锚,同伴锚基于主锚生成。

[0079] 具体而言,图 7 所示的生成参考锚集合的步骤 S702 可以进一步包含图 8 所示的子

步骤。如图 8 所示,在步骤 S801 中,基于锚空间使用登记语音,从而生成主锚。在步骤 S802 中,基于生成的主锚生成同伴锚。在步骤 S803 中,将主锚以及同伴锚组合为参考锚集合。根据实施方式,步骤 S801 至步骤 S803 分别可以如上所述由主锚生成单元 3021、同伴锚生成单元 3022、以及组合单元 3023 执行。

[0080] 根据本公开的另一实施方式,步骤 S801 进一步包含从锚空间中找到与登记语音的模型的距离最近的锚模型,由此生成主锚的步骤。

[0081] 根据本公开的另一实施方式,步骤 S802 进一步包含如下的步骤,即:基于主锚将锚空间分割为多个集群,找出距不包含主锚的多个集群的重心的距离最近的锚模型,由此生成同伴锚的步骤。

[0082] 根据本公开的另一实施方式,步骤 S802 进一步包含:最初,令 N 为与主锚的个数相等的值,将锚空间分割为 N 个集群的步骤;找到要进一步进行分割的、在 N 个集群中具有最大的类内距离的一个集群的步骤,在该步骤中,在集群具有主锚的情况下,基于主锚与集群内的距主锚最远的锚,将集群分割为两个子集群,而在集群不具有主锚的情况下,基于相对于距离更远的主锚的、集群内最远的两个锚,将集群分割为两个子集群的步骤;以及反复执行上述过程直到得到多个集群为止的步骤。

[0083] 根据本公开的一实施方式,图 8 所示的生成参考锚集合的方法可以在步骤 S801 之后进一步包含追加自适应过程的步骤。一方面,步骤 S801 中生成的主锚可以在步骤 S802 中用于生成同伴锚之前,使用 MAP 以及 MLLR 等自适应技术利用登记语音进行自适应,同伴锚在步骤 S802 的自适应后基于主锚生成。另一方面,MAP 以及 MLLR 等自适应技术可以在由步骤 S801 生成的主锚上应用,同伴锚不进行步骤 S802 的自适应而基于主锚生成,因此组合单元 3023 将同伴锚与进行了自适应的主锚组合在一起,从而得到参考锚集合。根据本公开的实施方式,自适应的步骤如上所述可以由自适应单元 5024 执行。

[0084] 该方法的上述步骤的执行并不限定于上述顺序,这些步骤可以逐一执行及 / 或并列执行。还有可能无须执行图示的所有步骤。

[0085] 图 9 表示基于本公开的一实施方式的说话人识别系统的方框图。

[0086] 如图 9 所示,整个说话人识别系统 900 包括两个阶段,一个是登记阶段,另一个是测试阶段。登记阶段的结构与参考图 2 说明的相同,因此以下省略其详细说明。测试阶段包含前端 901、声纹生成单元 902、对比单元 903、以及判断单元 904。前端 901 取得目标说话人的测试语音,从测试语音中提取特征参数,并将特征参数发送至声纹生成单元 902。前端 901 以及前端 201 可以作为一体进行动作,并不限定于上述结构。声纹生成单元 902 基于如上所述从前端 901 发送来的特征参数以及由参考锚集合生成单元 202 生成的参考锚集合,生成声纹。与前端 901 以及 201 同样,声纹生成单元 902 以及声纹生成单元 203 也可以作为一体进行动作,并不限定于上述结构。对比单元 903 将由测试阶段生成的声纹与由登记阶段生成的声纹进行比较,并将比较结果发送至判断单元 904。判断单元 904 基于结果识别目标说话人的同一性,即,在比较结果大于规定阈值的情况下,识别出目标说话人的同一性,在比较结果小于规定阈值的情况下,否认目标说话人的同一性。

[0087] 据此,基于本公开的实施方式的说话人识别系统 900 可构成为包括如下装置,即:前端 201 或 901,从目标说话人取得登记语音及 / 或测试语音;参考锚集合生成单元 202,基于锚空间使用登记语音,由此生成参考锚集合;声纹生成单元 203 或 902,基于参考锚集合

以及登记语音及 / 或测试语音生成声纹 ; 对比单元 903 , 将根据测试语音生成的声纹与根据登记语音生成的声纹进行比较 ; 以及判断单元 904 , 基于比较结果识别目标说话人的同一性。

[0088] 根据本公开的另一实施方式 , 说话人识别系统 900 可构成为包括建模设备以及识别设备 , 建模设备包括 : 第一前端 201 , 从目标说话人取得登记语音 ; 参考锚集合生成单元 202 , 基于锚空间使用登记语音 , 由此生成参考锚集合 ; 以及第一声纹生成单元 203 , 基于参考锚集合以及登记语音生成第一声纹 ; 识别设备包括 : 第二前端 901 , 从目标说话人取得测试语音 ; 第二声纹生成单元 902 , 基于参考锚集合以及测试语音生成第二声纹 ; 对比单元 903 , 将第一声纹与第二声纹进行比较 ; 以及判断单元 904 , 基于比较结果识别目标说话人的同一性。

[0089] 这样 , 通过考虑登记语音以及说话人自适应技术 , 能够生成尺寸更小的锚模型 , 因而能够进行具有尺寸更小的参考锚集合的、可靠性高的鲁棒的说话人识别。这对于进行计算速度的改善以及大幅的存储器削减是非常有利的 , 因此计算量较少 , 参考锚集合较小 , 因而更适于嵌入型应用。为了确认本公开的有效性 , 进行若干实验。在第一实验中 , 使用对特定尺寸的参考锚集合的说话人识别以及对比的实验。如表 1 所示 , 本公开能够优于以往方法。

[0090] [ 表 1 ]

[0091]

	识别	对比
以往方法	56. 58	65. 46
有自适应的以往方法	60. 48	72. 09
实施方式	73. 62	74. 00
有自适应的实施方式	75. 08	77. 29

[0092] 在下面的实验中 , 检验参考锚集合的尺寸对说话人识别系统的影响。从图 10 中可以看出 , 本公开在参考锚集合的尺寸较小这一点上显著优于以往方法 , 并且说话人自适应可以提高性能。此外还显示出 , 有两个因素对系统性能产生影响。一个因素是锚模型关于目标说话人的辨别能力。以往方法的初始性能由于该辨别能力弱而较差。根据实施方式 , 在锚生成过程中考虑辨别能力 , 因此性能得到改善。此外 , 锚上的说话人自适应也能够提高该能力 , 性能进一步得到改善。除了主锚 , 同伴锚尤其以否决为目的还具有追加的辨别能力。这是因为 , 在无自适应时 , 锚的尺寸增大时性能得到改善 , 从图 10 的以往方法以及实施方式的曲线中可以看出这一点。改善辨别能力的自适应的正面效果在锚较小时达到最大。作为其他因素 , 还可以举出生成的说话人声纹向量的稳定程度 , 该稳定程度受到登记数据的尺寸的影响。根据图 10 的有自适应的实施方式的曲线 , 性能发生降低 , 这是因为在锚的尺寸增大时 , 为了生成稳定的更高维数的说话人声纹向量需要相当多的登记数据。

[0093] 在锚的大小较小的情况下 , 说话人声纹生成所需的登记数据较少即可 , 因此该自适应居于主导地位。随着该锚的尺寸变大 , 为了生成稳定的说话人声纹向量所需的登记数

据的量变得相当大,自适应的效果减少。总之,声纹的维数越多,则否决能力越大,声纹的维数越少,则所需的登记数据可以越少。维数较少的情况下效果优于维数较多的情况下效果,这是因为在较少维数的状态下附加了这种性能提高的效果。

[0094] 上述详细说明通过使用方框图、流程图及 / 或实施例,描述了设备及 / 或过程的各种实施方式。本领域技术人员应当理解,只要这种方框图、流程图及 / 或实施例包含一个或多个功能及 / 或运算,这种方框图、流程图或实施例内的各功能及 / 或运算可以通过各种硬件、软件、固件、或者它们的实质上的任意组合,单独及 / 或集成地实现。在一实施方式中,本说明书中记载的主题的若干部分可以通过专用集成电路(Application Specific Integrated Circuits, ASIC)、场可编程门阵列(Field Programmable Gate Arrays, FPGA)、数字信号处理器(Digital Signal Processor, DSP)、或者其他集成电路方式实现。但是,本领域技术人员应当理解,本说明书中公开的实施方式的若干形式的全部或者部分可以作为一个或多个计算机上执行的一个或多个计算机程序(例如,作为一个或多个计算机系统上执行的一个或多个程序)、作为一个或多个处理器上执行的一个或多个程序(例如,作为一个或多个微处理器上执行的一个或多个程序)、作为固件、或者作为它们的实质上的任意组合,在集成电路内以等价的结构实现,并且根据本公开,本领域技术人员应当能够为软件及 / 或固件设计电路及 / 或编写代码。此外,本领域技术人员应当理解,本说明书中记载的主题的机制可以作为各种形式的程序产品分发,另外本说明书中记载的主题的例示的一实施方式的应用与用于实际执行分发的特定种类的信号传播介质无关。作为信号传播介质的例子,可举出且不限于如下介质:软盘、硬盘驱动器、紧凑式光盘(CD)、数字视频光盘(DVD)、数字磁带、计算机存储器等可记录型介质,以及数字及 / 或模拟通信介质等发送型介质(例如光导纤维电缆、波导管、有线通信链路、无线通信链路等)。

[0095] 本说明书中记载的主题有时例示收容于其他不同的组件内或者连接于其他组件的组件。应当理解的是,所示出的这种结构仅为例示,实际上可以实现具有相同功能的多种其他结构。在概念意义上,为了实现相同功能,无论以任何方式配置组件,都是为了实现期望功能而实际进行“关联”。因此,为了实现特定功能而组合的本说明书的任意两个组件可以视为与结构或中间组件无关,为了实现期望功能而相互“关联”。同样,以此方式关联的任意两个组件还可以视为为了实现期望功能而相互“可动作地连接”或者“可动作地结合”,进而,能够以此方式关联的任意两个组件还可以视为为了实现期望功能而相互“能够可动作地连接”。作为能够可动作地结合的特定例子,可举出且不限于如下例子:物理上可啮合、及 / 或物理上相互作用的组件、及 / 或能够以无线方式相互作用、及 / 或能够以无线方式相互作用的组件、及 / 或逻辑上相互作用、及 / 或逻辑上能够相互作用的组件。

[0096] 关于本说明书中的实质上的多数形及 / 或单数形的用语的使用,本领域技术人员能够根据背景情况及 / 或用途,适当地将多数形改变为单数形、及 / 或将单数形改变为多数形。关于各种单数形 / 多数形的置换,在本说明书中为了易于理解有时明确进行描述。

[0097] 本领域技术人员应当理解的是,一般而言,本说明书中使用的、尤其是附属的权利要求书(例如附属的权利要求书的正文)中使用的说法一般是“非限定性的”说法(例如,说法“包括有”应解释为“包括有但不限于”,说法“具有”应解释为“至少具有”,说法“包括”应解释为“包括但不限于”,等等)。此外,本领域技术人员应当理解,在意图所导入的权利要求列举的特定数量的情况下,该意图在权利要求内会明确记载,在没有这种列举的情况下,不

存在这种意图。例如,为了便于理解,在以下的附属的权利要求中,可以加入导入句“至少一个”以及“一个或多个”以导入权利要求列举。但是,即使使用了这种语句,通过基于不定冠词“a”或者“an”的权利要求列举的导入,即使该权利要求包含导入句“至少一个”或“一个或多个”,以及“a”或“an”等不定冠词,包含这种导入的权利要求列举的特定权利要求也不应解释为限定于仅包含一个这种列举的公开(例如,“a”及 / 或“an”典型地应解释为指“至少一个”或“一个或多个”),关于用于导入权利要求列举的定冠词的使用,这一点也成立。此外,本领域技术人员应当理解,即使明确记载了特定数量的导入的权利要求列举,这种列举典型地也应该理解为指至少记载的数量(例如,不加其他修饰语的“两个列举”这一不加修饰的列举典型地指至少两个列举,或者两个以上列举)。在使用类似于“A、B、或 C 等中至少一者”的惯用说法的情况下,一般而言,该结构是指本领域技术人员对该惯用说法所理解的意义(例如,“具有 A、B、或 C 中至少一者的系统”包括但不限于仅具有 A、仅具有 B、仅具有 C、具有 A 及 B、具有 A 及 C、具有 B 及 C、及 / 或具有 A、B 及 C 等的系统)。此外,本领域技术人员应当理解,不管是说明书、权利要求书、还是附图中,表示两个以上代替词语的实质上的任意分离词语及 / 或语句,都应当理解为有可能包含多个词语中之一、多个词语中的任一者、或者词语双方。例如,语句“A 或 B”应理解为包含“A”或“B”、或者“A 及 B”的可能性。

[0098] 本说明书中公开了各种形式以及实施方式,但本领域技术人员应当显而易见地想到其他形式以及实施方式。本说明书中公开的各种形式以及实施方式旨在进行例示,并不意图进行限定,真正的范围和精神由权利要求书示出。

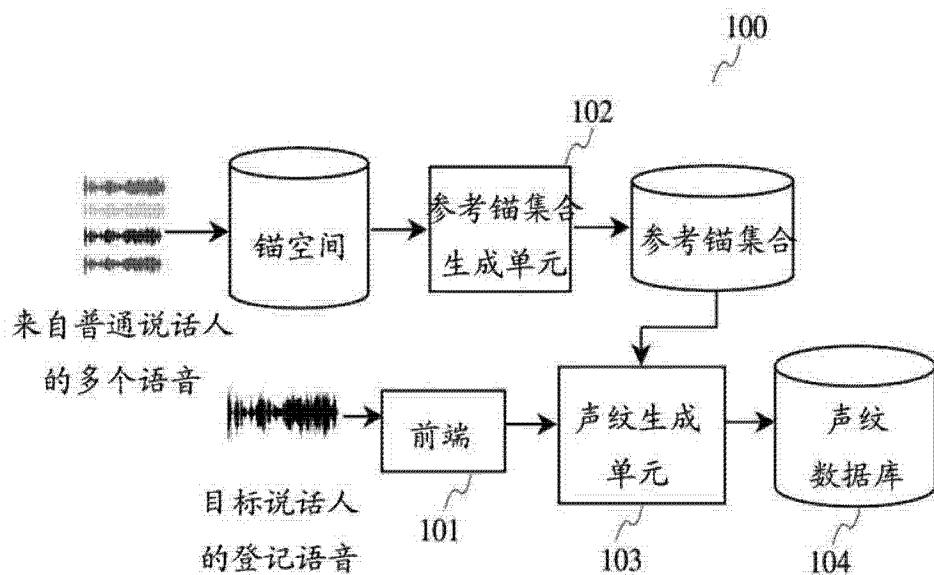


图 1

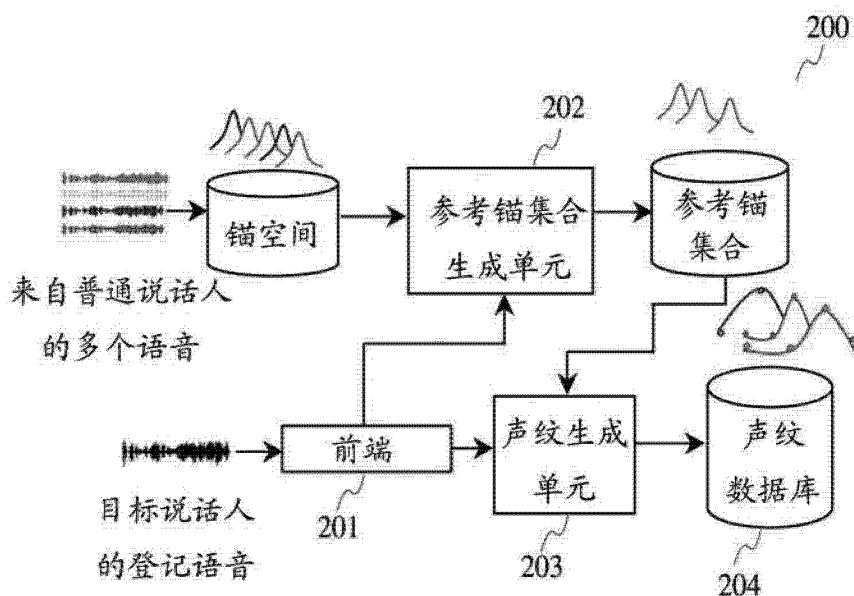


图 2

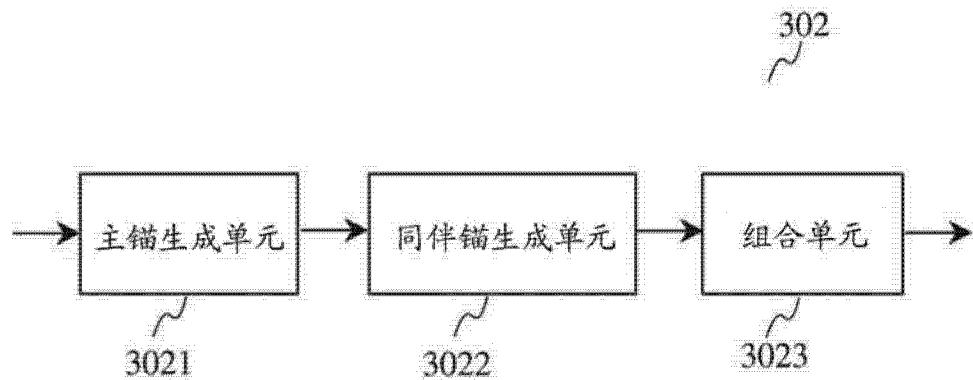


图 3

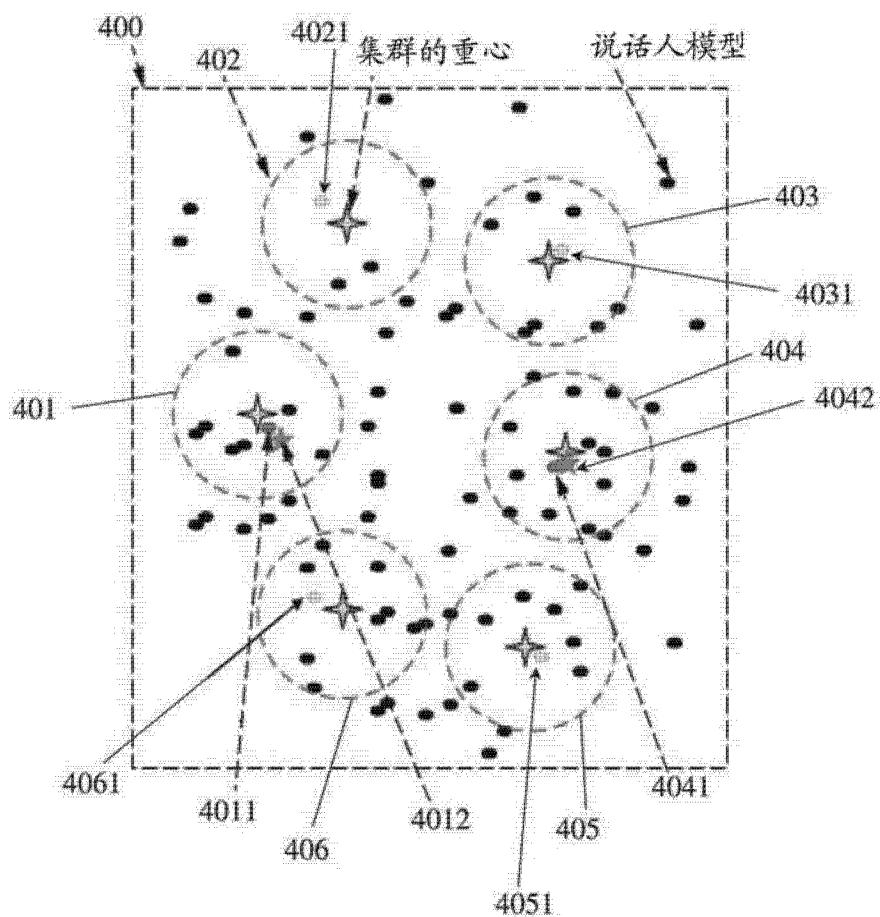
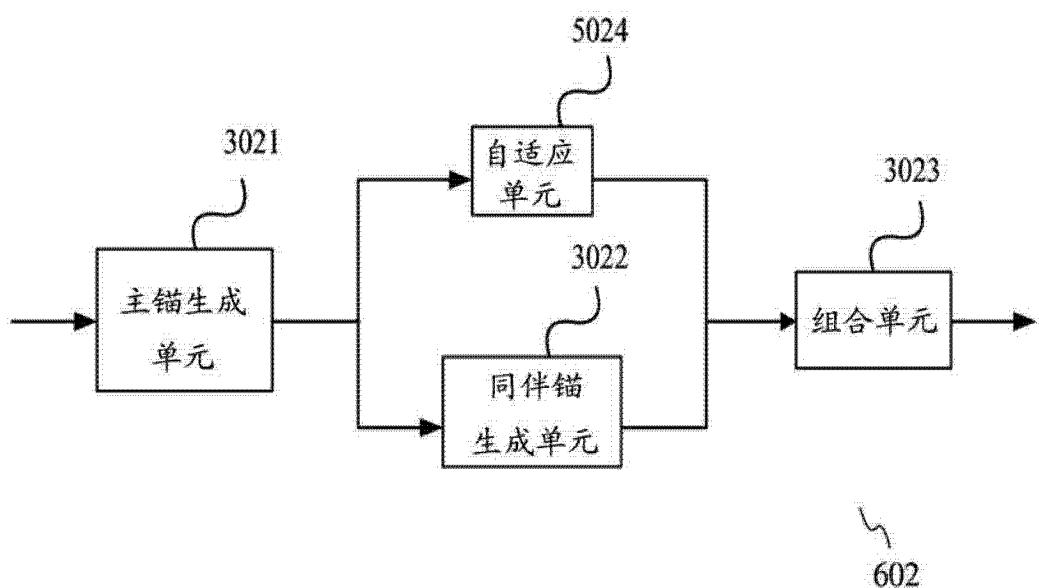
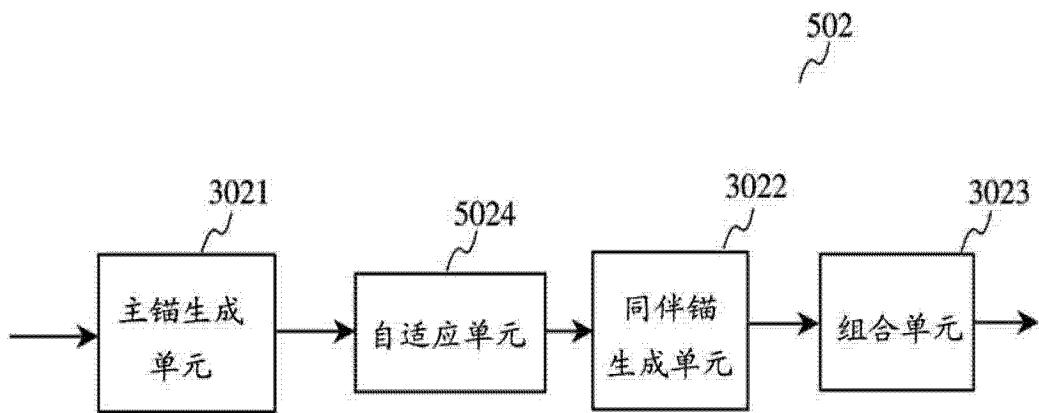


图 4



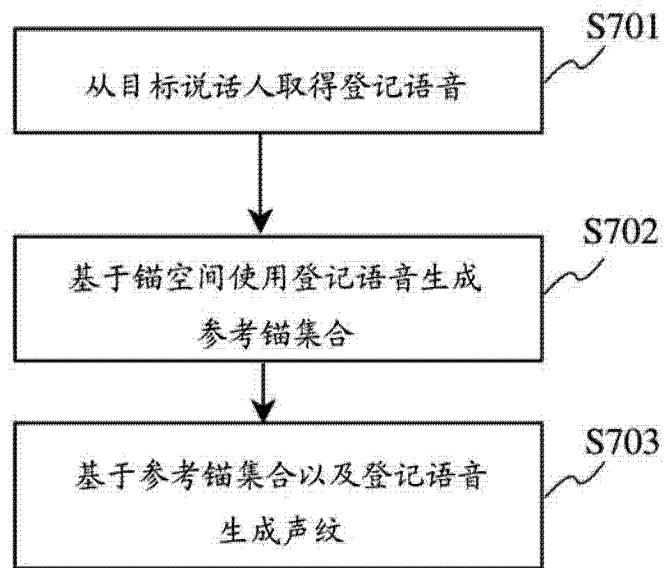


图 7

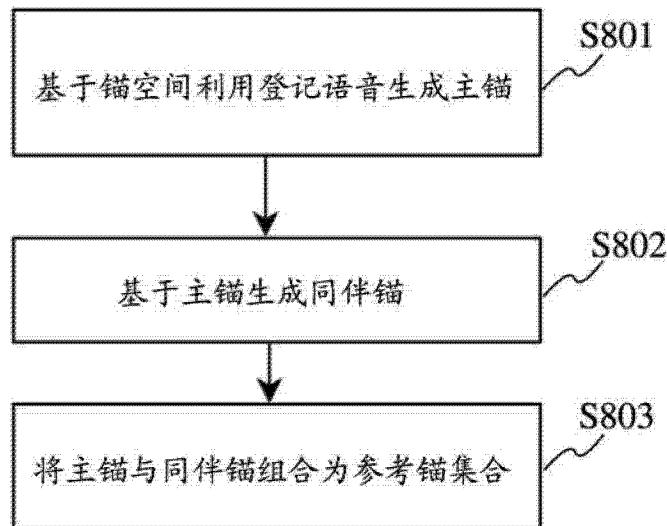


图 8

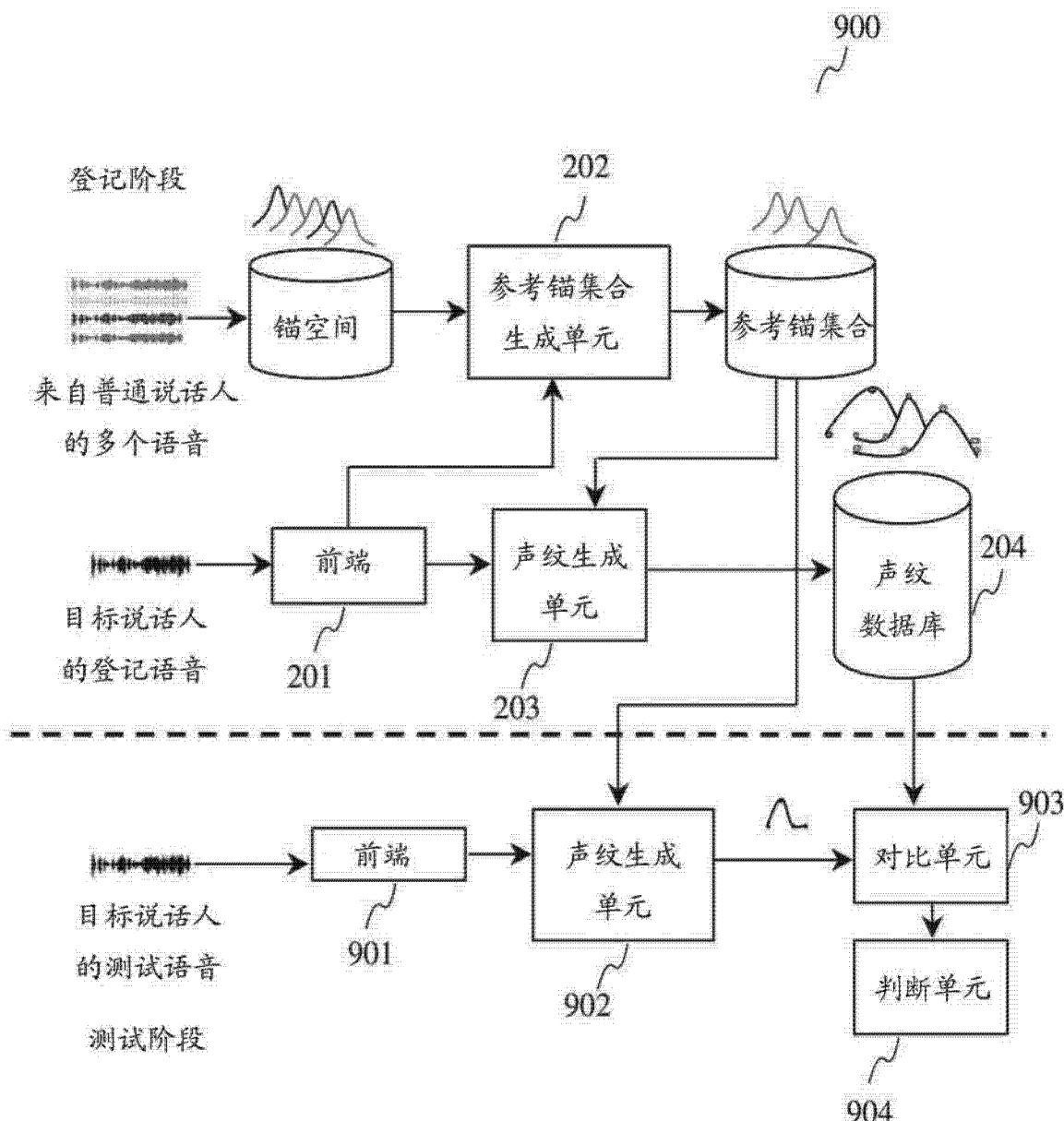


图 9

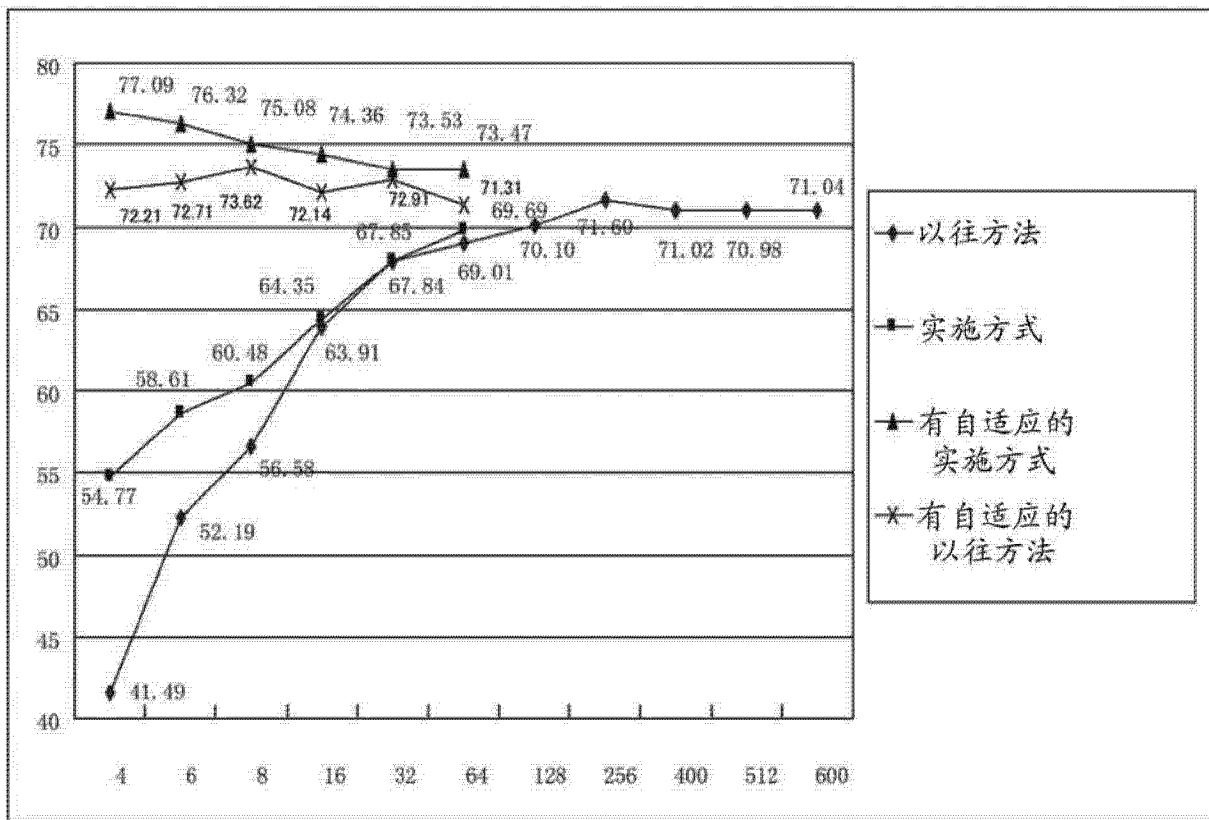


图 10