



(12)发明专利

(10)授权公告号 CN 104185868 B

(45)授权公告日 2017.08.22

(21)申请号 201380016170.2

(22)申请日 2013.01.23

(65)同一申请的已公布的文献号
申请公布号 CN 104185868 A

(43)申请公布日 2014.12.03

(30)优先权数据
2012900256 2012.01.24 AU

(85)PCT国际申请进入国家阶段日
2014.09.23

(86)PCT国际申请的申请数据
PCT/AU2013/000050 2013.01.23

(87)PCT国际申请的公布数据
W02013/110125 EN 2013.08.01

(73)专利权人 澳尔亚有限公司
地址 澳大利亚新南威尔士州

(72)发明人 哈比卜·E·泰欧哈米
阿米特·萨达南德·马勒高恩卡
雷努卡·阿米特·马勒高恩卡
克莱夫·戴维·萨默菲尔德

(74)专利代理机构 北京派特恩知识产权代理有限公司 11270
代理人 浦彩华 武晨燕

(51)Int.Cl.
G10L 17/14(2006.01)

(56)对比文件
WO 01/16940 A1,2001.03.08,
CN 1567431 A,2005.01.19,
CN 102820033 A,2012.12.12,
CN 1543641 A,2004.11.03,
CN 102238189 A,2011.11.09,

审查员 陈红红

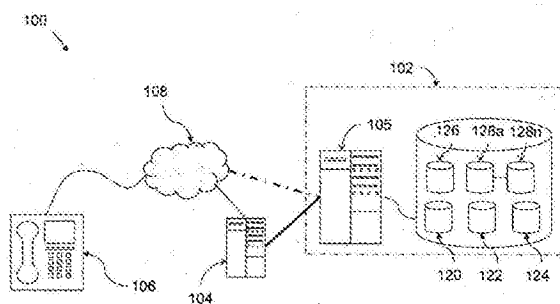
权利要求书2页 说明书8页 附图6页

(54)发明名称

语音认证和语音识别系统及方法

(57)摘要

一种用于配置语音识别系统的方法包括获得一个语音认证系统在一个语音认证过程中所利用的一个语音样本。对该语音样本进行处理以生成与该语音样本相关联的多个语音单元的多个声学模型。对这些声学模型进行存储以便随后由该语音识别系统用作一个语音识别过程的一部分。



1. 一种用于配置语音识别系统的方法,该方法包括:

从一个用户获得一个语音样本,该语音样本用于认证该用户作为一个认证过程的一部分;

对该语音样本进行处理以训练与该语音样本相关联的多个语音单元的一个或多个通用声学模型以创建相应的一个或多个个性化声学模型;

将该一个或多个个性化声学模型存储在该用户的一个个性化声学模型集中;

基于该用户所提供的包含相应的多个语音单元的多个附加语音样本选择性地重新训练该个性化声学模型集中的该一个或多个个性化声学模型;以及

响应于确定该用户已经访问一个语音识别功能,引导一个语音识别过程访问该个性化声学模型集以便识别随后的多种用户言语。

2. 根据权利要求1所述的方法,其中,这些语音单元包括多个三音子、多个双音子、多个聚类状态、多个音素、多个单词或多个短语。

3. 根据权利要求2所述的方法,进一步包括对与该语音样本相关联的语音内容数据进行评估从而为这些语音单元中的每个语音单元确定一个可闻标识符,以及基于所确定的该可闻标识符对所述个性化声学模型进行分类。

4. 根据前述权利要求中任意一项所述的方法,其中,所述个性化声学模型包括这些语音单元的多个语言和/或语法模型。

5. 根据权利要求1至3中任意一项所述的方法,进一步包括确定该一个或多个个性化声学模型中的每个个性化声学模型的一个质量测量,并且其中,基于多个附加语音样本对该一个或多个个性化声学模型进行重新训练,直到相应的质量测量满足一个预先定义的阈值。

6. 根据权利要求4所述的方法,进一步包括确定该一个或多个个性化声学模型中的每个个性化声学模型的一个质量测量,并且其中,基于多个附加语音样本对该一个或多个个性化声学模型进行重新训练,直到相应的质量测量满足一个预先定义的阈值。

7. 根据权利要求1-3中任意一项所述的方法,其中,响应于成功地认证该用户,自动引导该语音识别过程访问该个性化声学模型集。

8. 根据权利要求4所述的方法,其中,响应于成功地认证该用户,自动引导该语音识别过程访问该个性化声学模型集。

9. 根据权利要求5所述的方法,其中,响应于成功地认证该用户,自动引导该语音识别过程访问该个性化声学模型集。

10. 根据权利要求6所述的方法,其中,响应于成功地认证该用户,自动引导该语音识别过程访问该个性化声学模型集。

11. 一种组合式语音识别和话音认证方法,包括:

响应于一个话音认证功能成功地认证一个用户,访问供一个语音识别功能在该用户识别一种或多种言语时使用的一个个性化声学语言和/或语法模型集,该声学语言和/或语法模型集包含多个声学语言和/或语法模型,已经使用源自该用户或者在向该认证功能登记的过程中或者在一次或多次随后的认证过程中所提供的多种言语的话音数据对这些声学语言和/或语法模型进行训练。

12. 根据权利要求11所述的方法,其中,该个性化的声学语言和/或语法模型集包括多

个语音单元的多个声学模型,已经利用一个或多个具有一个共享的话音轮廓的其他用户向该用户提供的多个语音样本对这些声学模型进行了训练。

13. 一种语音识别系统,包括:

一个处理模块,可操作以:

获得一个语音样本,该语音样本由认证系统用于认证一个用户作为一个认证过程的一部分;

对该语音样本进行处理,从而训练与该语音样本相关联的多个语音单元的一个或多个通用声学模型以创建相应的一个或多个个性化声学模型并从而随后将该一个或多个个性化声学模型存储在一个个性化声学模型集中;

基于该用户所提供的包含相应的多个语音单元的多个附加语音样本选择性地重新训练该一个或多个个性化声学模型;以及

响应于确定该用户已经访问一个语音识别功能,进一步将该处理模块安排成用于引导一个语音识别过程以访问该个性化声学模型集以用于识别随后的多种用户言语。

14. 根据权利要求13所述的系统,其中,这些语音单元包括多个三音子、多个双音子、多个聚类状态、多个音素、多个单词或多个短语。

15. 根据权利要求13所述的系统,其中,该处理模块进一步可操作以对与该语音样本相关联的语音内容数据进行评估从而为这些语音单元中的每个语音单元确定一个可闻标识符,以及基于相关标识符对所述个性化声学模型进行分类。

16. 根据权利要求14所述的系统,其中,该处理模块进一步可操作以对与该语音样本相关联的语音内容数据进行评估从而为这些语音单元中的每个语音单元确定一个可闻标识符,以及基于相关标识符对所述个性化声学模型进行分类。

17. 根据权利要求13至16中任意一项所述的系统,该处理模块进一步可操作以确定该一个或多个个性化声学模型中的每个个性化声学模型的一个质量测量,以及继续重新生成该一个或多个个性化声学模型直到该质量测量达到一个预先定义的阈值。

18. 根据权利要求13至16中任意一项所述的系统,其中,这些附加语音样本由该用户或者在向该认证系统登记的过程中或者作为随后的由该认证系统所执行的一次认证会话的一部分提供。

19. 根据权利要求17所述的系统,其中,这些附加语音样本由该用户或者在向该认证系统登记的过程中或者作为随后的由该认证系统所执行的一次认证会话的一部分提供。

20. 根据权利要求13至16中任意一项所述的系统,其中,自动地访问该个性化声学模型集以响应于该认证系统成功地认证该用户来执行该语音识别过程。

21. 根据权利要求17所述的系统,其中,自动地访问该个性化声学模型集以响应于该认证系统成功地认证该用户来执行该语音识别过程。

22. 根据权利要求18所述的系统,其中,自动地访问该个性化声学模型集以响应于该认证系统成功地认证该用户来执行该语音识别过程。

23. 根据权利要求19所述的系统,其中,自动地访问该个性化声学模型集以响应于该认证系统成功地认证该用户来执行该语音识别过程。

话音认证和语音识别系统及方法

技术领域

[0001] 本发明涉及作为话音认证系统的一部分而运行的语音识别系统的自动调谐和配置。结果是一种既识别个人又识别其语音的系统。

背景技术

[0002] 制作有效的语音识别系统的关键是创建声学模型、语法和语言模型,这些模型使得底层语音识别技术能够在应用内可靠地识别正在说的内容并且给定语音样本的上下文的情况下能够弄清楚或理解该语音。创建声学模型、语法和语言模型的过程包含收集语音样本(通常还称为话音样本)的数据库,该数据库代表说话者与语音识别系统交互的方式。为了创建这些声学模型、语法和语言模型,需要将数据库中的每个语音样本进行分段并将其标注成其单词或音素组成部分。然后就对所有说话者(比如像所有说单词“二(two)”的说话者)的全部常见组成部分进行编译和处理以创建此组成部分的单词(或音素)声学模型。在基于大词汇音素的系统中,还需要重复该过程以针对此语言学市场创建语言和口音特定的模型和语法。通常,产生可以准确地识别语音的声学模型需要(来自每种性别的)每个单词或音素的约1,000至2,000个示例。

[0003] 针对任何语言学市场开发语音识别系统是一个数据驱动的过程。在不存在代表此市场特定的语言和口音的语音数据的情况下,不能产生适当的声学、语法和语言模型。因此,获得必要的语音数据(假设其是可获得的)以及针对新的语言学市场创建适当的语言和口音特定的模型可能特别耗时并非常昂贵。

[0004] 如果提供了一种可以针对任何语言学市场用有成本效益的方式自动配置的语音识别系统,这将是有益的。

发明内容

[0005] 根据本发明的一个第一方面,提供了一种用于配置语音识别系统的方法,该方法包括:

[0006] 获得由一个话音认证系统在一个话音认证过程中所利用的一个语音样本;

[0007] 对该语音样本进行处理以生成与该语音样本相关联的多个语音单元的多个声学模型;以及

[0008] 对这些声学模型进行存储以便随后由该语音识别系统用作一个语音识别过程的一部分。

[0009] 在一个实施例中,这些语音单元包括三音子、双音子、聚类状态、音素、单词或短语。

[0010] 在一个实施例中,该方法进一步包括:对与该语音样本相关联的语音内容数据进行评估从而为这些语音单元中的每个语音单元确定一个可闻标识符,以及基于所确定的该可闻标识符对这些声学模型进行分类。

[0011] 在一个实施例中,该方法进一步包括基于从多个进一步获得和处理的语音样本所

生成的多个声学模型更新所存储的这些声学模型。

[0012] 在一个实施例中,该方法进一步包括确定所存储的这些声学模型中的每个声学模型的一个质量,以及继续更新这些声学模型直到该质量达到一个预先定义的阈值。

[0013] 在一个实施例中,这些语音样本由该认证系统的不同用户在向其登记的过程中提供。

[0014] 在一个实施例中,该方法进一步包括将这些声学模型存储在一个通用语音识别数据库中。

[0015] 在一个实施例中,该方法进一步包括仅获得与选自包括以下各项的组的一个或多个预先定义的语音轮廓相关联的多个语音样本:语言、性别、信道介质、语法。

[0016] 在一个实施例中,这些语音样本由同一用户或者在向该认证系统登记的过程中或者作为随后的一次认证会话的一部分提供。

[0017] 在一个实施例中,这些声学模型存储在该用户特定的一个数据库中,并且其中,自动地访问该数据库以响应于该用户向该认证系统认证自己来执行该语音识别过程。

[0018] 根据本发明的一个第二方面,提供了一种组合式语音识别和话音认证方法,包括利用一个用户的一次话音认证确定的一个输出设置一种语音识别功能的一个参数,以便随后由该用户识别一种言语。

[0019] 在一个实施例中,利用该输出选择多个声学模型数据库中的一个,以供该语音识别功能在识别该用户的言语时使用,每个声学模型数据库包含一组以不同的方式训练的声学模型。

[0020] 在一个实施例中,该数据库包括多个语音单元的多个声学模型,已经使用源自该用户或者在向该认证系统登记的过程中或者在随后的一次认证会话的过程中所提供的言语的话音数据对这些声学模型进行了训练。

[0021] 在一个实施例中,该数据库包括多个语音单元的多个声学模型,已经利用一个或多个具有一个共享的话音轮廓的其他用户向该用户所提供的多个语音样本对这些声学模型进行了训练。

[0022] 根据本发明的一个第三方面,提供了一种实施计算机程序的计算机可读介质,该计算机程序包括一个或多个用于控制一个计算机系统实施如上文根据该第一方面所述的方法的指令。

[0023] 根据本发明的一个第四方面,提供了一种语音识别系统,包括:

[0024] 一个处理模块,可操作以获得一个话音认证系统在一个话音认证过程中所利用的一个语音样本,该处理模块进一步被安排成用于处理该语音样本以生成与该语音样本相关联的多个语音单元的多个声学模型;以及

[0025] 一个存储模块,可操作以存储这些声学模型,以便随后由该语音识别系统用作该处理模块所实施的一个语音识别过程的一部分。

[0026] 在一个实施例中,这些语音单元包括三音子、双音子、聚类状态、音素、单词或短语。

[0027] 在一个实施例中,该处理模块进一步可操作以对与该语音样本相关联的语音内容数据进行评估从而为这些语音单元中的每个语音单元确定一个可闻标识符,以及基于相关标识符对这些声学模型进行分类。

[0028] 在一个实施例中,该处理模块被进一步安排成用于基于从多个进一步获得和处理的语音样本所生成的多个声学模型更新所存储的这些声学模型。

[0029] 在一个实施例中,该处理模块进一步可操作以确定所存储的这些声学模型中的每个声学模型的一个质量,以及继续更新这些声学模型直到该质量达到一个预先定义的阈值。

[0030] 在一个实施例中,这些语音样本由该认证系统的不同用户在向其登记的过程中提供。

[0031] 在一个实施例中,将这些声学模型存储在一个通用语音识别数据库中。

[0032] 在一个实施例中,该处理模块进一步可操作以仅获得与选自包括以下各项的组的一个或多个期望的预先定义的轮廓相关联的多个语音样本:语言、性别、信道介质、语法。

[0033] 在一个实施例中,这些语音样本由同一用户或者在向该认证系统登记的过程中或者作为随后的一次认证会话的一部分提供。

[0034] 在一个实施例中,该系统包括一个数据库,该数据库可操作以存储这些声学模型,并且其中,自动地访问该数据库以响应于该认证系统成功地认证该用户来执行该语音识别过程。

[0035] 根据本发明的一个第五方面,提供了一种组合式语音识别和话音认证系统,该系统包括:

[0036] 一种话音认证功能,可操作以认证一种用户言语;

[0037] 一种语音识别功能,可操作以由该用户响应于该话音认证功能进行的一次肯定认证对随后的多种言语进行评估;以及

[0038] 一个参数设置模块,可操作以基于一个用户标识符将该语音识别功能的一个参数设置为由该话音认证功能所建立。

[0039] 在一个实施例中,利用该标识符选择该语音识别功能在识别该用户的随后的多种言语时所使用的一组声学模型数据库中的一个声学模型数据库。

[0040] 在一个实施例中,所选择的数据库包括多个声学模型,已经使用该用户或者在向该认证系统登记的过程中或者在随后的一次认证确定的过程中所提供的多个语音样本对这些声学模型进行了训练。

[0041] 在一个实施例中,所选择的数据库包括多个声学模型,已经使用该一个或多个具有一个共享的话音轮廓的其他用户提供给该用户的多个语音样本对这些声学模型进行了训练,从该话音认证确定来确定该话音轮廓。

附图说明

[0042] 参照附图,仅通过示例的方式,本发明的特征和优点将从其实施例的下列描述中变得清楚,在附图中:

[0043] 图1是根据本发明的一个实施例的系统的框图;

[0044] 图2是图1的话音处理系统所实施的个体模块的示意图;

[0045] 图3是展示了用于创建声纹的流程的示意图;

[0046] 图4是根据本发明的一个实施例展示了用于为图1的系统提供语音识别能力的流程的示意图;

[0047] 图5是根据一个实施例展示了用于建立语音识别模型和语法的流程的示意图;以及

[0048] 图6是根据一个实施例展示了用于为图1的系统提供用户特定的语音识别能力的流程的示意图。

具体实施方式

[0049] 实施例利用话音认证系统(通常还称为话音生物识别系统)所处理的语音样本自动地创建语音识别模型,可以有利地利用这些语音识别模型来提供附加语音识别能力。由于所生成的模型基于系统的实际用户所提供的样本,为这些用户调谐该系统,并且该系统因此能够为此用户群体提供高水平的语音识别准确度。这种技术还避免了购买“附加”语音识别方案的需要,这些语音识别方案不仅昂贵而且可能难以获得,尤其是对于适合创建语音识别技术所使用的声学模型、语法和语言模型的语音数据库不可用的市场。实施例还涉及个性化语音识别模型,这些个性化语音识别模型用于为系统的个人用户提供甚至更高水平的语音识别准确度。

[0050] 出于说明的目的,并参照附图,下文将在话音处理系统102的上下文中对本发明的实施例进行描述,该话音处理系统为安全服务104(如交互式话音响应(“IVR”)电话银行服务)提供话音认证和语音识别功能两者。在所展示的实施例中,独立于安全服务104(例如,由第三方提供商)实施话音处理系统102。在本实施例中,安全服务104的用户使用电话106形式的输入设备(例如,基于标准电话、移动电话或互联网协议(IP)的电话服务,如Skype™)与安全服务104通信。

[0051] 图1展示了用于实施本发明的一个实施例的示例系统配置100。如上所述,用户使用电话106与电话银行服务104通信。安全服务104进而连接至话音处理系统102,以便初始地认证这些用户并且此后在电话银行会话过程中为用户话音命令提供语音识别能力。根据所展示的实施例,话音处理系统102通过一个公共交换电话网108形式的通信网连接至安全服务104。

[0052] 系统配置的进一步细节

[0053] 参照图2,话音处理系统102包括一个服务器计算机105,该服务器计算机包括典型的服务器硬件,该服务器硬件包括处理器、主板、随机存取存储器、硬盘和电源。服务器105还包括一个操作系统,该操作系统与该硬件合作以提供可以在其中执行软件应用的环境。就这一点而言,服务器105的硬盘装载有处理模块114,该处理模块在该处理器的控制下可操作以实施各种话音认证和语音识别功能。如所展示的,处理模块114由各种用于实施前述功能的个体模块/组件构成,即,话音生物识别训练器115、话音生物识别引擎116、自动语音识别训练器117和自动语音识别引擎118。

[0054] 处理器模块114通信地耦合至许多数据库,包括身份管理数据库120、话音文件数据库122、声纹数据库124和语音识别模型和语法数据库126。还可以提供许多个性化语音识别模型数据库128a至128n以用于存储各自为特定用户的话音所定制的模式和语法。提供了一个规则存储器130以用于存储处理模块114所实施的各种规则,如在随后的段落中将更详细描述。

[0055] 服务器105包括用于与安全服务提供商系统104通信的适当软件和硬件。可以通过

任何合适的通信链路进行该通信,如互联网连接、无线数据连接或公共网络连接。在一个实施例中,通过安全服务提供商104路由用户语音数据(即,代表用户在登记、认证和随后与安全服务提供商系统104交互的过程中所提供的语音样本的数据)。可替代地,可以直接向服务器105提供语音数据(在这种情况下,服务器105还将实施合适的呼叫应答服务)。

[0056] 如所讨论的,所展示的实施例的通信系统108是公共交换电话网的形式。然而,在可替代的实施例中,通信网可以是数据网,如互联网。在这种实施例中,用户可以使用联网的计算设备使用网络协议(如TCP/IP协议)与服务器105交换数据(在一个实施例中,XML代码和分组的话音消息)。在国际专利申请PCT/AU 2008/000070中概述了这种实施例的进一步细节,该国际专利申请的内容通过引用结合在此。在另一可替代的实施例中,通信系统可以附加地包括连接至分组交换网的第三或第四代(“3G”)CDMA或GPRS使能的移动电话网,可以利用该移动电话网访问服务器105。在这种实施例中,用户输入设备102包括用于将语音样本作为数据传输的无线能力。该无线计算设备可以包括例如移动电话、具有无线卡的个人计算机和任何其他有利于语音备案功能的移动通信设备。在另一实施例中,本发明可以采用基于802.11的无线网或某种其他个人虚拟网络。

[0057] 根据所展示的实施例,安全服务提供商系统104是电话银行服务器的形式。安全服务提供商系统104包括一个收发器,该收发器包括一个用于与处理系统102通信的网卡。该服务器还包括用于提供应答服务的适当硬件和/或软件。在所展示的实施例中,安全服务提供商104通过公共交换电话网108利用收发器模块与用户通信。

[0058] 声纹登记

[0059] 在以任何细节描述用于创建语音识别模型的技术之前,将首先参照图3对一种用于登记语音样本和生成声纹的基本流程进行描述。在步骤302,由语音处理系统102接收语音样本并将其以合适的文件存储格式(例如,wav文件格式)存储在语音文件数据库122中。语音生物识别训练器115在步骤304处理所存储的语音文件以用于生成声纹,该声纹与提供该语音样本的用户的标识符相关联。系统102可以从用户请求附加语音样本,直到已经接收到足够数量的样本以用于创建准确的声纹。通常,对于文本相关的实施方式(即,其中,用户说出的文本必须是相同的以用于登记和验证)而言,请求和处理对相同单词或短语的三次重复,从而生成准确的声纹。在文本无关的实施方式(即,其中,用户可以出于验证目的提供任何言语)的情况下,请求超过30秒的语音以用于生成准确的声纹。可以例如使用授权给同一申请人的澳大利亚专利2009290150中所描述的过程对声纹质量进行测量,该澳大利亚专利的内容通过引用结合在此。在步骤306,将声纹加载至声纹数据库124中以便随后由语音生物识别引擎116在用户认证过程中使用(步骤308)。用户在认证过程中所提供的验证样本(这些验证样本可以例如是口令短语、账号等)也存储在语音文件数据库122中,以便使用本领域技术人员所熟知的技术更新或“调谐”所存储的与此用户相关联的声纹时使用。

[0060] 创建通用语音识别模型

[0061] 参照图4,示出了登记过程的扩展,该扩展有利地允许基于已登记的语音文件针对语音识别能力自动创建通用语音识别模型。在步骤402,将所存储的语音文件(该语音文件可以或者是登记过程中所提供的语音文件或者是后成功认证所提供的语音文件)传递至ASR训练器117,该训练器处理该语音文件以生成与该语音文件相关联的多个语音单元的多个声学模型,如将在随后的段落中更详细描述。在步骤404,随后将这些声学模型存储在

语音识别模型数据库126中,这些声学模型各自优选地由从语音文件数据库122所获得的多个语音文件所生成。这些模型随后可以在步骤406用于提供自动语音识别能力以供用户访问安全服务104。

[0062] 更详细地,并附加地参照图5,声学模型生成步骤402包括使用分段器模块502将这些语音文件分开成期望语音单元类型的语音单元(还称为成分)(502)。根据所展示的实施例,分段器模块502可处理的不同类型的语音单元包括三音子、双音子、聚类状态、音素、单词和短语,尽管将理解任何合适的语音单元可以是取决于期望的实施方式可处理的。分段器模块502为语音单元指定一个起始点并为语音单元指定一个结束点。可以对分段器模块502进行编程以将该结束点标识为下一语音单元的起始点。同等地,可以对分段器模块502进行编程以识别一个语音单元的结束与下一语音单元的起始之间的间隙。该间隙中的波形在此被称为“垃圾”并且可以代表沉默、背景噪声、通信信道所引入的噪声或说话者所产生但与语音不相关联的声音(如,呼吸噪声、“嗯”、“啊”、犹豫等)。训练器506使用此类声音来产生特殊模型,这种特殊模型在本领域内通常被称为“垃圾模型”。随后识别引擎126使用这些垃圾模型来识别语音样本中所听到的声音,但是这些语音样本不是预先定义的语音单元。在步骤504与可闻标识符(下文中“分类符”)相关联地存储已分段的非垃圾语音单元,该可闻标识符源自与原始语音样本相关联的语音内容数据。例如,语音处理系统可以存储包含用户在登记(例如,其账号等)过程中所说的单词或短语的元数据。分段器502可以对语音学查找字典进行评估以确定组成所登记的单词/短语的语音单元(三音子、双音子、聚类状态或音素)。语音单元的通用或原型声学模型存储在分段器502中并被其用于将用户所提供的语音分段成其三音子、双音子、聚类状态或音素组成部分。获得、分段并存储(步骤504)进一步的话音文件,直到已经获得了每个语音单元的足够数量的样本从而为已分类的语音单元创建通用语音模型。在具体实施例中,需要每个三音子、双音子、聚类状态或音素部分的500和2,000之间个样本以产生适合识别的那部分的通用声学模型。根据所展示的实施例,当新的语音文件存储在数据库122中时,ASR训练器117自动地对其进行处理,以用于创建和/或更新模型数据库126中所存储的声学模型。通常在生成一个模型之前获得并处理500和2,000之间个语音文件,以便提供一个将足够反映已登记的用户语言和口音的模型。随后由训练器模块506对语音单元进行处理。训练器模块506对已登记的说话者所说的已分段的语音单元进行处理以使用本领域中已知的模型生成技术创建语音识别系统所需的这些语音单元中的每个语音单元的声学模型。类似地,训练模块506还对来自与语音识别所使用的语音单元相关联的语音文件的语法和语言模型进行编译。从语音样本中的三音子、双音子、聚类状态、音素、单词和/或短语序列的统计分析来计算语法和语言模型,该统计分析代表特定三音子、双音子、聚类状态、音素、单词和/或短语后跟着另一特定三音子、双音子、聚类状态、音素、单词和/或短语的几率。这样,特定于说话者在系统中进行登记的方式并因此特定于所登记的说话者所说的口音和语言实施声学模型、语法和语言模型。将所生成的模型和内含的语法存储在数据库126中以便随后在向安全服务104的用户提供自动语音识别时使用。

[0063] 在一个实施例中,由处理模块114实施某些规则,这些规则指定了模型创建所必须处理的语音单元样本的最小数量。这些规则还可以在处理模块114将可利用所存储的模型对语音进行识别之前指定所存储的模型的质量。在具体实施例中,对于每个分类器而言,可

以存在一个男性和女性模型。根据这种实施例,这些规则可以规定只选择来自男性用户的语音样本来创建男性模型,并且只选择来自女性用户的语音样本来创建女性模型。可以从所存储的与已知用户相关联的元数据或通过样本进行评估的方式(该评估包含使用女性和男性模型两者声学地处理该样本,以及基于所产生的认证分数确定性别,即,使用男性模型的更高分表示男性说话者,同时使用女性模型的更高分表示女性说话者)对此进行确定。可以同等地为不同语言、信道介质(例如,移动电话、陆上通信线等)和语法轮廓创建附加的或可替代的模型,这样使得将基于所检测到的呼叫者轮廓选择具体的模型集。可以例如基于该呼叫可获得的数据(如可以指示哪个轮廓最接近地匹配当前呼叫的电话线路号码或IP地址),或通过并行地使用许多不同的模型以及选择生成最佳结果或合适的模型(例如,通过评估所产生的认证分数)处理语音来确定所检测到的轮廓。

[0064] 创建个性化语音识别模型

[0065] 一旦已经成功地认证了用户,认为他们是系统102‘已知的’。在具体实施例中,一旦用户已知,可以创建并随后访问个性化模型集从而为此用户提供更高的语音识别准确度。

[0066] 根据这种实施例,并附加地参照图6,为系统已知的每个用户提供了个性化声纹和语音识别数据库128(见步骤602至606)。可以初始地从用户在登记过程中所提供的语音样本(例如,在某些实例中,可以要求用户提供多个登记语音样本,例如陈述其账号、姓名、pin号码等,可以对这些语音样本进行处理以创建有限数量的模型)、从如之前所述的通用模型、或从这两者的组合来配置这些模型。当用户提供新的语音样本时,可以创建新的模型并更新现有模型,如果需要的话。将理解的是,可以或者在成功认证用户的过程中或之后提供新的样本(例如,由用户在电话银行会话过程中所发布的话音命令所导致)。系统102还可以提示用户不时地发出特定单词、短语等(即,在步骤602)以帮助针对此用户建立更完整的模型集。再次,可以用规则存储器130中所存储的规则来控制此过程。

[0067] 尽管之前的段落中所描述的实施例描述了“第三方”、或集中式系统形式的处理系统102,但它仍将是提供商系统104。

[0068] 可替代的配置和方法可以包括说话者使用第三方语音识别功能(如“Siri”私人助理)收集语音样本(如转让给苹果公司(Apple Inc.)的已公开的美国专利申请号20120016678中所描述的),或者集成到手机或其他计算设备中的“龙(Dragon)”语音识别软件(从美国马萨诸塞州伯灵顿市的Nuance公司(Nuance Communications, Inc.)可获得),结合本文所述的话音认证系统使用所述手机或其他计算设备。在这种情况下,来自“已知”说话者的语音样本可以存储在语音文件数据库122中并且然后被分段器模块502和训练器模块506用于使用上文所述的过程为此说话者创建语音识别模型。

[0069] 可替代地,主机服务或云服务(如与话音认证系统结合使用的托管IVR服务或基于云的话音处理系统)所收集的语音样本还可以用于使用本文所述的方法创建语音识别模型。

[0070] 虽然已经参照本实施例描述了本发明,但本领域的技术人员将理解,在不背离本发明的范围的情况下可以做出更改、改变和改善并且可用等效物来替代其元件和其步骤。此外,可以进行许多修改以使本发明适应一种具体状况或使材料适应本发明的教导,而不背离其中心范围。然而,此类更改、改变、修改和改善(尽管未在上文专门地描述)旨在并被

暗示处于本发明的范围和精神内。因此,本发明旨在不限于在此所描述的具体实施例,并且将包括所有落在独立权利要求的范围内的实施例。

[0071] 在下面的权利要求书和本发明的前续描述中,除了由于明确的语言或必要的暗示,上下文另作要求以外,单词“包括(comprise)”或变形形式如“包括(comprises)”或“包括(comprising)”是以包含在内的意义来使用的,即,指明了所述特征的存在但并不排除在本发明的不同实施例中进一步特征的存在或添加。

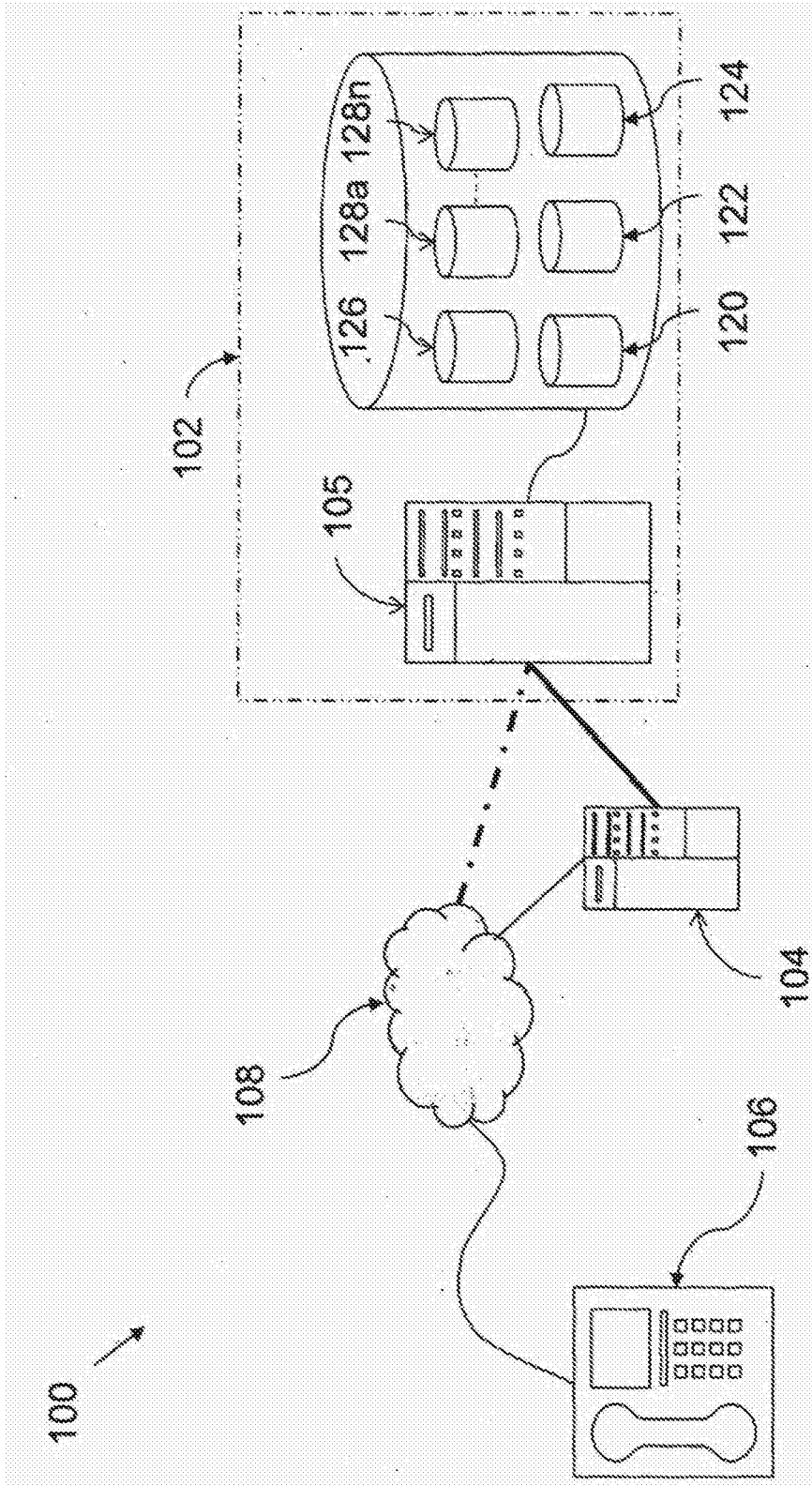


图1

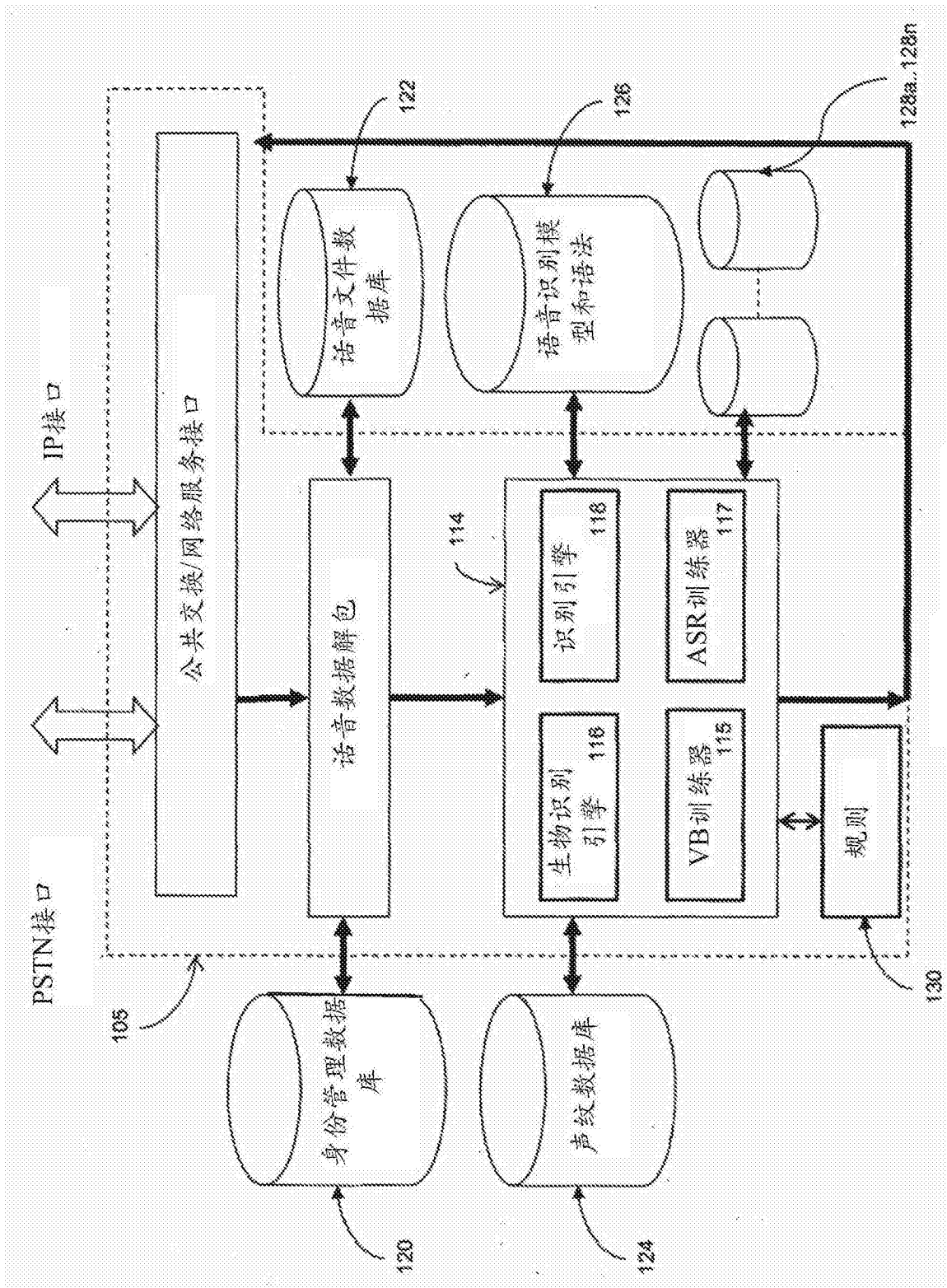


图2

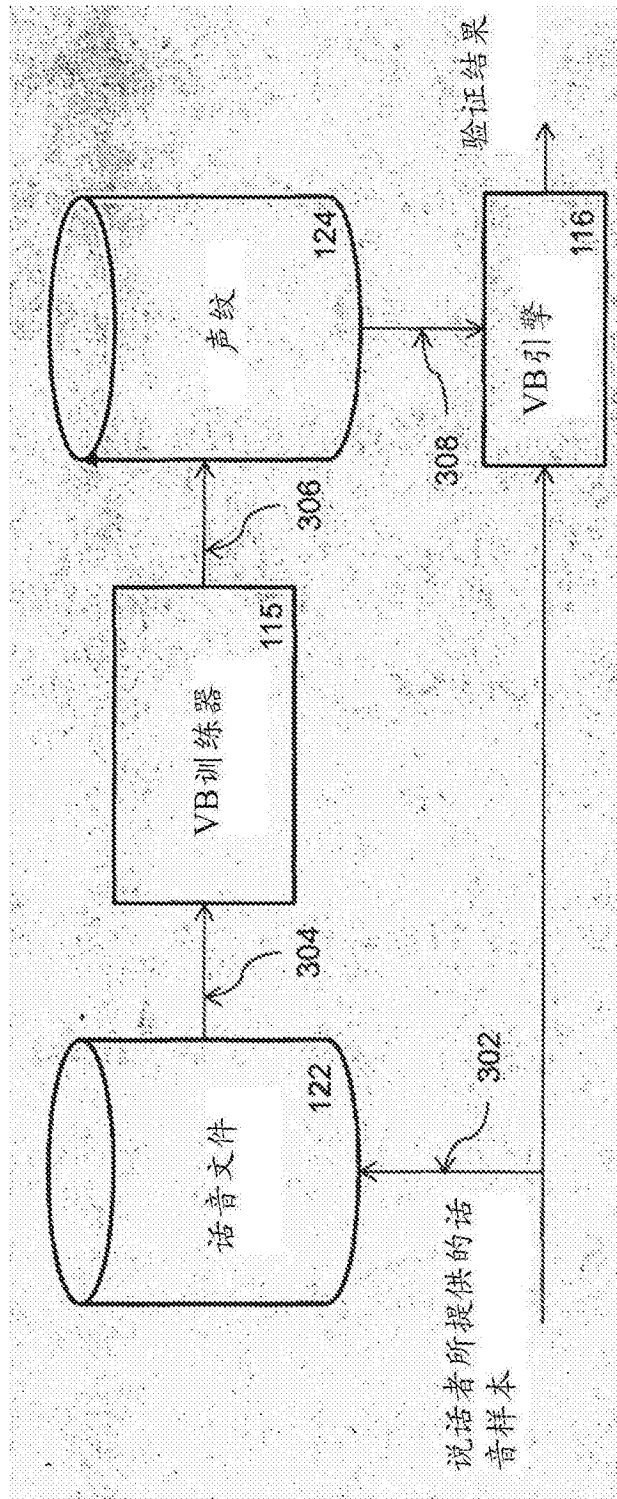


图3

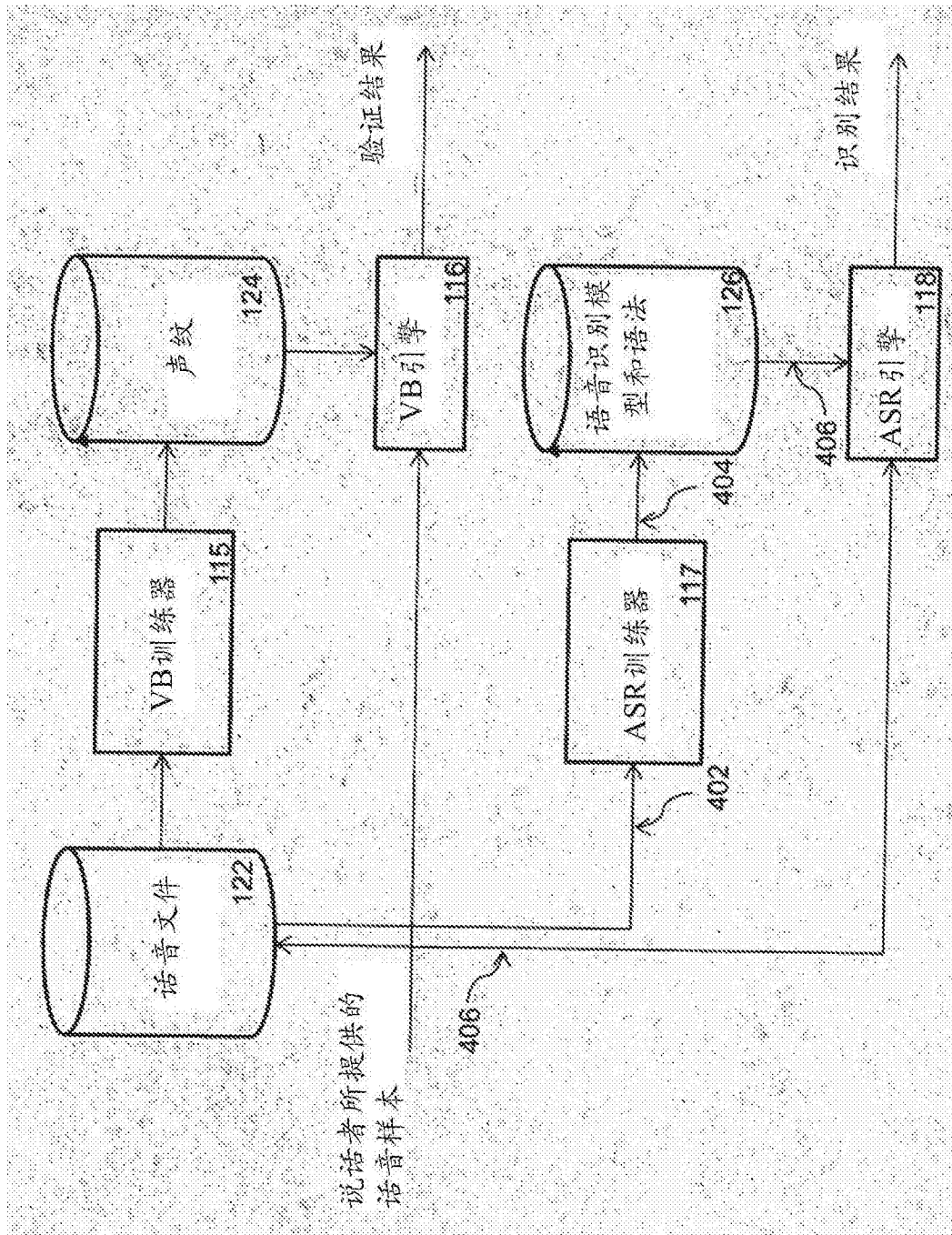


图4

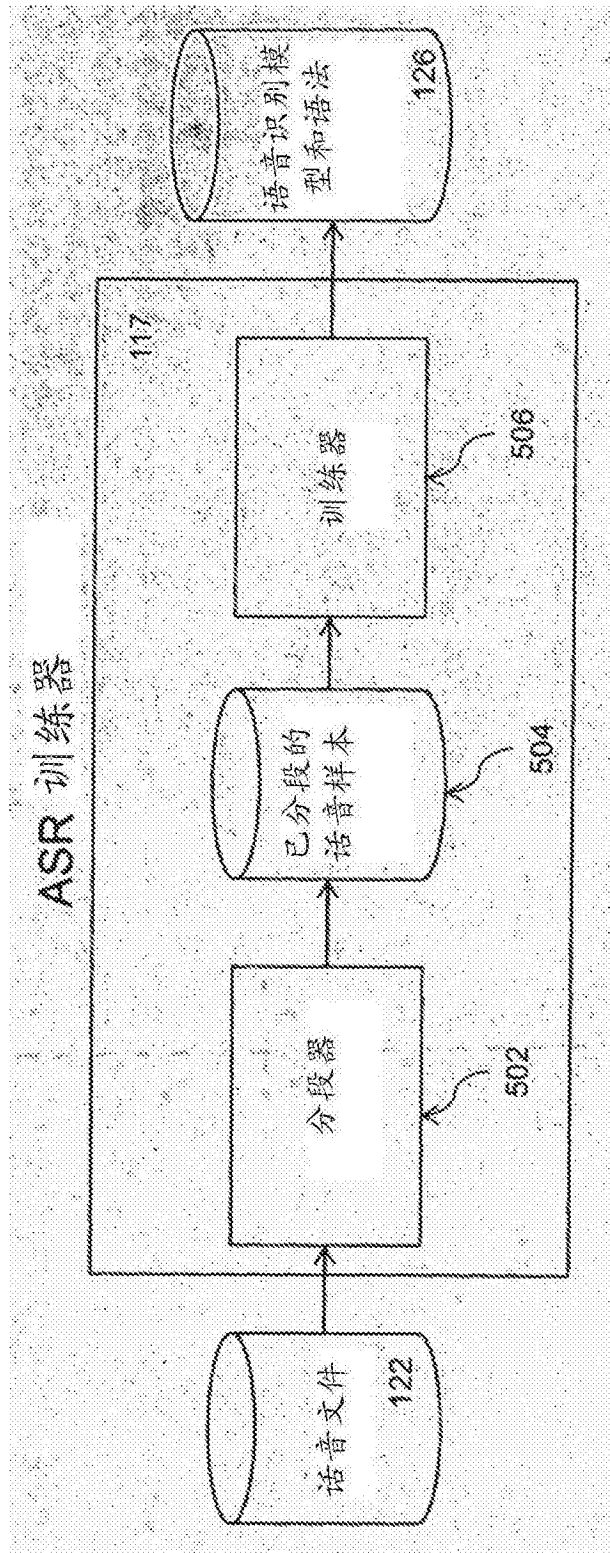


图5

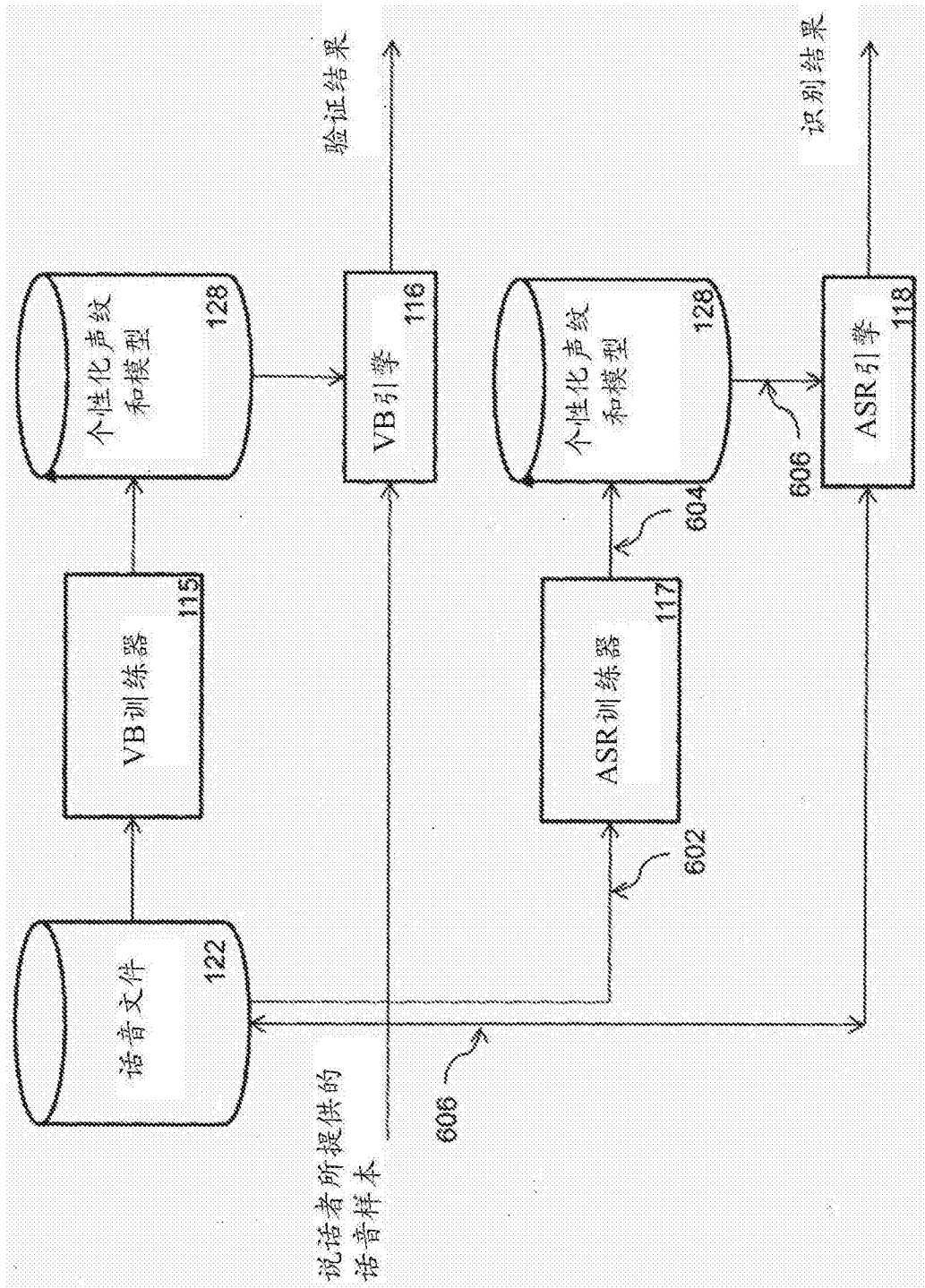


图6