



(12)发明专利申请

(10)申请公布号 CN 106537493 A

(43)申请公布日 2017.03.22

(21)申请号 201580031165.8

(22)申请日 2015.09.29

(85)PCT国际申请进入国家阶段日
2016.12.20

(86)PCT国际申请的申请数据
PCT/CN2015/091042 2015.09.29

(71)申请人 深圳市全圣时代科技有限公司
地址 518000 广东省深圳市南山区西丽街
道茶光路1063号一本大楼727

(72)发明人 李强生

(51)Int. Cl.
G10L 15/30(2013.01)
G10L 15/02(2006.01)
G10L 15/06(2013.01)
G10L 15/22(2006.01)

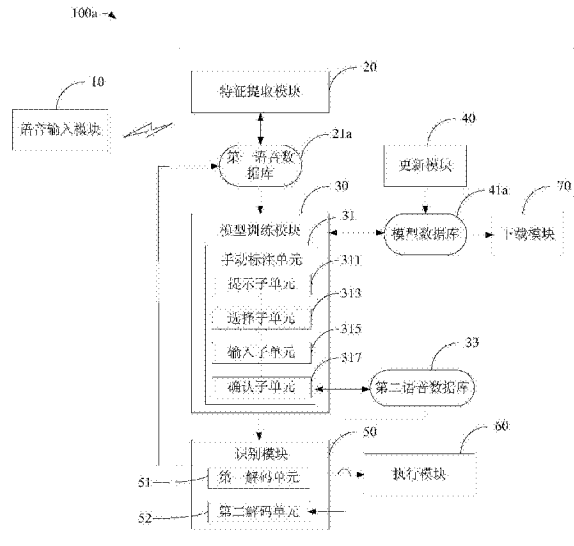
权利要求书2页 说明书6页 附图6页

(54)发明名称

语音识别系统和方法、客户端设备及云端服务器

(57)摘要

本发明公开一种语音识别系统,至少包括:语音输入模块,用于当启用实时通话或语音录入功能时,实时输入用户的语音;特征提取模块,用于从所输入的用户语音中提取语音特征;模型训练模块,用于根据所述语音特征以及预设的规则,建立对应的声学 and 语言模型;以及更新模块,用于保存并更新所述声学 and 语言模型到一个模型数据库中。本发明还提供一种语音识别方法、客户端设备以及云端服务器。



1. 一种语音识别系统,其特征在于,所述系统至少包括:
语音输入模块,用于当启用实时通话或语音录入功能时,实时输入用户的语音;
特征提取模块,用于从所输入的用户语音中提取语音特征;
模型训练模块,用于根据所述语音特征以及预设的规则,建立对应的声学 and 语言模型;
以及
更新模块,用于保存并更新所述声学 and 语言模型到一个模型数据库中。
2. 如权利要求1所述的语音识别系统,其特征在于,所述特征提取模块将提取到的语音特征实时保存在一个第一语音数据库中,所述模型训练模块定时或定量从所述第一语音数据库中提取所述语音特征以进行模型训练。
3. 如权利要求2所述的语音识别系统,其特征在于,所述特征提取模块、模型训练模块以及更新模块集成在一个云端服务器中,所述云端服务器包括对应不同用户的多个私有云模块,所述特征提取模块提取的特定语音特征保存到对应的私有云模块下,并通过所述模型训练模块和更新模块建立模型和更新,所述识别模块则分别识别不同云模块下的语音数据。
4. 如权利要求1所述的语音识别系统,进一步包括:
识别模块,用于根据模型数据库中的所述声学 and 语言模型,判断是否能够识别所述语音特征,如果能够识别,则生成携带控制命令的识别结果,否则,将无法识别的其他语音特征存储到一个第一语音数据库中,以供所述模型训练模块重新进行模型训练。
5. 如权利要求4所述的语音识别系统,其特征在于,至少包括:
第一解码单元,用于将所述语音特征与所述声学 and 语言模型进行匹配度计算,如果匹配度大于等于阈值,则判断能够识别对应的所述语音特征并输出识别结果,否则,判断无法识别所述语音特征;以及
所述模型训练模块进一步包括一手动标注单元,用于根据用户命令,手动将所述匹配度低于所述阈值的无法识别的语音特征与预设的标准语音进行映射匹配,并将所述语音特征与所述标准语音数据及其映射关系保存在一个第二语音数据库中。
6. 如权利要求5所述的语音识别系统,其特征在于,所述手动标注单元包括:
提示子单元,用于周期性提示用户查看存储在所述第一语音数据库中的无法识别的语音特征;
选择子单元,用于供用户选择对应于所述无法识别的语音特征的标准语音,其中所述标准语音预先存储在所述第一语音数据库中;和/或
输入子单元,用于供用户输入对应于所述无法识别的语音特征的标准语音;以及
确认子单元,用于供用户确认所述无法识别的语音特征与所述标准语音之间的映射关系,并存储到所述第二语音数据库。
7. 如权利要求5所述的语音识别系统,其特征在于,所述识别模块还包括第二解码单元,用于根据当前输入的用户语音以及所述第二语音数据库,识别所述用户的语音,并输出对应的标准语音。
8. 如权利要求4所述的语音识别系统,其特征在于,所述识别模块对所述语音特征进行识别的同时,将所述语音特征存储在所述第一语音数据库中,以供模型训练模块从所述第一语音数据库提取所述语音特征,从而进行模型训练。

9. 如权利要求4所述的语音识别系统,其特征在于,通过一个云端服务器的各个私有云模块分别实现所述特征提取模块、模型训练模块、更新模块以及识别模块的功能,其中每一个私有云模块对应一个用户,所述特征提取模块提取的特定语音特征保存到对应的私有云模块下

10. 如权利要求1所述的语音识别系统,进一步包括:

下载模块,用于供用户将对应私有云模块中的声学 and 语言模型下载到本地,以在本地实现语音识别。

11. 一种语音识别方法,包括:

基于启用实时通话或语音录入功能实时输入用户的语音;

从所输入的用户语音中提取语音特征;

根据所述语音特征以及预设的规则,建立对应的声学 and 语言模型;以及

实时保存并更新所述声学 and 语言模型到一个模型数据库中。

12. 一种客户端设备,其包括如权利要求1~9项任一项所述的语音识别系统。

13. 一种云端服务器,其包括对应不同用户的多个私有云主模块,每个云主模块包括:

特征提取模块,用于从来自于正在启用实时通话或语音录入功能的客户端设备所输入的用户语音中提取语音特征;

模型训练模块,用于根据所述语音特征以及预设的规则,建立对应的声学 and 语言模型;以及

更新模块,用于保存并更新所述声学 and 语言模型到一个模型数据库中。

语音识别系统及方法、客户端设备及云端服务器

技术领域

[0001] 本发明涉及语音识别领域,尤其涉及一种语音识别系统及方法和具有语音识别功能的客户端设备及云端服务器。

背景技术

[0002] “大词汇连续语音识别”(Large Vocabulary Continuous Speech Recognition, LVCSR,简称“语音识别”),就是由计算机根据人的连续声音信号中所蕴涵的语言信息,识别出某段语音对应的是哪些文字的过程。

[0003] 大词汇连续汉语语音识别器已经取得了很大的进展,对标准普通话,识别器的准确率可以达到95%以上。但是,汉语的方言问题是汉语语音识别面临的主要问题。由于在中国大部分人的普通话都带有一定的方言背景,在这样的情况下,大部分的语音识别器的性能都会大大下降,甚至无法使用。

[0004] 当前包括苹果公司的Siri、中国的科大讯飞等设备和软件可以提供语音输入功能,但是语音识别受用户个人发音的影响,导致语音识别时准确率受到很大影响,进而影响了语音识别功能的适用。另外,大量的非智能客户端设备,在使用时其自带的语音操控功能,也由于语音输入时识别率的问题,而影响到其语音功能的适用,例如汽车中的语音操作功能、蓝牙耳机、门铃等设备的语音操控等。

[0005] 目前很多识别器对方言背景对语音识别器性能造成的影响是用数据库方法去消除或减弱的,就是说,当已经有一个对标准普通话进行识别的语音识别器,需要对带某种方言背景的普通话进行识别时,采用的方法为:收集大量与该方言有关的第一语音数据库,然后利用已有的声学模型训练方法去重新训练声学模型,或利用已有的说话人自适应方法对声学模型进行自适应。这种方法的缺点是:(1)收集带方言背景的数据库的工作量非常巨大,对于汉语这么多的方言,数据库的收集更是一件巨大的工程。(2)这种方法无法兼顾标准普通话和带发音背景普通话之间的共性,仅是通过数据驱动的方法去解决问题,相当于完全重新构建一个语音识别器,给不同方言背景的语音识别器之间的资源共享和兼容带来困难。

发明内容

[0006] 为了解决上述技术问题,本发明提供一种语音识别系统及方法和具有语音识别功能的客户端设备及云端服务器。

[0007] 本发明一实施例提供一种语音识别系统,至少包括:语音输入模块,用于当启用实时通话或语音录入功能时,实时输入用户的语音;特征提取模块,用于从所输入的用户语音中提取语音特征;模型训练模块,用于根据所述语音特征以及预设的规则,建立对应的声学模型;以及更新模块,用于保存并更新所述声学模型到一个模型数据库中。

[0008] 本发明另一实施例还提供一种语音识别方法,包括:基于启用实时通话或语音录入功能实时输入用户的语音;从所输入的用户语音中提取语音特征;根据所述语音特征以

及预设的规则,建立对应的声学 and 语言模型;以及实时保存并更新所述声学 and 语言模型到一个模型数据库中。

[0009] 本发明又一实施例提供一种客户端设备,其包括上述的语音识别系统。

[0010] 发明再一实施例提供一种云端服务器,其包括对应不同用户的多个私有云主模块。每个云主模块包括:特征提取模块,用于从来自于正在启用实时通话或语音录入功能的客户端设备所输入的用户语音中提取语音特征;模型训练模块,用于根据所述语音特征以及预设的规则,建立对应的声学 and 语言模型;以及更新模块,用于保存并更新所述声学 and 语言模型到一个模型数据库中。

[0011] 本发明的语音识别系统和方法通过实时记录或保存实时通话和录音信息,并作为语音模型训练的样本,从而能够根据用户不同的发音特点持续更新模型数据库。由此,可以满足用户的个性化需求,而且能够支持多种语音,例如英语或者地方方言等,提高了识别度。

附图说明

[0012] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0013] 图1是本发明第一实施例提供的语音识别系统的系统框架图;

[0014] 图2是图1的语音识别系统的功能模块图;

[0015] 图3是本发明第二实施例提供的语音识别系统的功能模块图;

[0016] 图4是本发明一实施例提供的语音识别方法的流程图;

[0017] 图5是本发明另一实施例提供的语音识别方法的流程图;

[0018] 图6是图5中的步骤S409的具体流程图;

[0019] 图7是本发明又一实施例提供的语音识别方法的流程图。

具体实施方式

[0020] 下面结合附图和具体实施方式对本发明的技术方案作进一步更详细的描述。显然,所描述的实施例仅仅是本发明的一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动的前提下所获得的所有其他实施例,都应属于本发明保护的范围。

[0021] 第一实施例

[0022] 请参阅图1,其为本发明第一实施例提供的一种语音识别系统100的系统架构图。在本实施例中,所述语音识别系统100由客户端设备200与云端服务器300共同实现,以能够通过云端服务器300完成识别前端、模型训练和识别后端的全过程,并将最终的语音识别结果下发至客户端设备200。如此,可减轻客户端设备200的数据处理量,部署起来非常方便,且后续升级的大部分工作也都在云端服务器300完成。

[0023] 具体的,请参阅图2,所述语音识别系统100至少包括语音输入模块10、特征提取模块20、模型训练模块30以及更新模块40。本实施例中,所述语音输入模块10设置在客户端设

备200上,例如是麦克风及其处理电路。所述特征提取模块20、模型训练模块30、更新模块40等集成在所述云端服务器300中。

[0024] 所述语音输入模块10用于当客户端设备200启用实时通话或语音录入功能时,实时输入用户的语音。所述客户端设备200可以是手机、车载设备、电脑、手机、智能家居设备以及可穿戴设备等等。所述用户的语音也可进行本地保存或云端保存。

[0025] 所述特征提取模块20用于从所输入的用户语音中提取语音特征。本实施例中,所述特征提取模块20将提取到的语音特征实时保存在一个第一语音数据库21中,所述第一语音数据库21可以是本地数据库,也可以是云端数据库。所述语音特征指所述用户语音的特征数据。

[0026] 所述模型训练模块30用于根据所述语音特征以及预设的规则,建立对应的声学 and 语言模型,以供在后续识别过程中,将提取的语音特征与所述声学 and 语言模型进行匹配与比较,得到最佳的识别结果。本实施例中,所述预设的规则是动态时间规整(Dynamic time warping,简称DTW)、隐形马尔可夫(Hidden Markov Model,HMM)理论、矢量量化(Vector Quantization,简称VQ)技术中的至少其中之一者。此外,本实施例中,所述模型训练模块30定时从所述第一语音数据库21中提取所述语音特征,以进行模型训练。当然,在其他实施例中,所述模型训练模块30也可实时提取第一语音数据库21中的特定语音特征,以进行实时的模型训练,或者定量(例如100条)提取所述特定语音特征,本发明不以此些实施例为限

[0027] 所述更新模块40用于实时保存并更新所述声学 and 语言模型到一个模型数据库41中,由此,能够获取更庞大的声学 and 语言模型数据库41,提高了识别度。

[0028] 此外,为了能够对用户的语音信息进行保密,且针对不同的用户语音特点提供个性化的模型训练,所述云端服务器300包括对应于不同用户的多个私有云主模块,每个私有云主模块包括所述特征提取模块20、模型训练模块30以及更新模块40等等。其中,所述特征提取模块20提取的特定语音特征保存到对应的私有云模块下。同时,所述模型训练模块30对所述特定语音特征进行声学 and 语言模型训练,并通过更新模块40更新所述模型。当用户启用所述语音识别系统100时,可通过帐号鉴权的方式启用所述语音识别功能。

[0029] 可以理解的是,在其他实施例中,所述语音识别系统100还可集成在一个客户端设备200中,例如:车载设备、电脑、手机、智能家居设备以及可穿戴设备等等中,以供用户开启离线语音识别功能。此时,所述第一语音数据库21以及模型数据库41均为本地数据库。采用此种方式,能够在无网络连接的情况下,实现上述语音识别功能。

[0030] 总的来说,在传统语音识别技术中,通常不会将手机实时通话或利用pad(可以是其他设备)录音过程中的语音进行记录或保存,以作为语音模型训练的样本。而本发明通过实时记录或保存实时通话和录音信息,并作为语音模型训练的样本,从而能够根据用户不同的发音特点持续更新模型数据库41。由此,可以满足用户的个性化需求,而且能够支持多种语音,例如英语或者地方方言等,提高了识别度。此外,本发明还提供了针对不同用户的私有云主模块,供用户通过账户鉴权的方式启用语音识别功能,从而能够提高对用户语音信息的保密性能。

[0031] 第二实施例

[0032] 请参阅图3,本发明第二实施例提供的语音识别系统100a与第一实施例的语音识别系统100基本相同,不同之处在于:所述语音识别系统100a进一步包括识别模块50,所述

识别模块50用于根据模型数据库41a中的所述声学 and 语言模型,判断是否能够识别所述语音特征,如果能够识别,则生成携带控制命令的识别结果,否则,将无法识别的其他语音特征存储到第一语音数据库21a中。此时,所述第一语音数据库21a仅需保存无法识别的所述语音特征,节省了占用空间。所述模型训练模块30也进一步包括一个手动标注单元31,用于根据用户命令,手动将所述匹配度低于所述阈值的无法识别的语音特征与预设的标准语音进行映射,并将所述语音特征与所述标准语音数据及其映射关系更新在一个第二语音数据库33中,供所述识别模块50采用。对应的,所述识别模块50还用于根据当前输入的用户语音数据以及所述第二语音数据库33,识别所述语音数据并输出识别结果。

[0033] 更具体的,所述识别模块50包括第一解码单元51以及第二解码单元52,所述第一解码单元51用于将当前提取的语音特征与所述声学 and 语言模型进行匹配度计算。如果匹配度大于等于阈值,则判断能够识别对应的所述语音特征并输出识别结果,否则,判断无法识别所述语音特征。所述第二解码单元52用于根据当前输入的用户语音以及所述第二语音数据库33,识别所述用户的语音,并输出对应的标准语音。

[0034] 本实施例中,所述手动标注单元31包括提示子单元311、选择子单元313、输入子单元315以及确认子单元317。所述提示子单元311用于周期性提示用户查看存储在所述第一语音数据库21中的无法识别的语音特征。所述选择子单元313用于供用户选择对应于所述无法识别的语音特征的标准语音,其中所述标准语音预先存储在所述第一语音数据库21中。例如,用户可以通过听取所述无法识别的特定的语音,然后根据所提供的标准语音,选择与所述语音特征相匹配的标准语音。所述输入子单元315,用于供用户输入对应于所述无法识别的语音特征的标准语音。可以理解的是,可仅选择所述选择子单元313以及所述输入子单元315其中之一者进行设置,当标准语音中无对应的选项时,可通过语音输入的方式,确定对应的标准语音。所述确认子单元317用于供用户确认所述语音特征与所述标准语音之间的映射关系,并于确认后,将所述映射关系存储到所述第二语音数据库33中。

[0035] 在第二实施例中,所述特征提取模块20、模型训练模块30、更新模块40以及识别模块50等集成在所述云端服务器300a中,所述识别模块50分别识别不同云模块下的语音数据。

[0036] 第二实施例提供的语音识别系统100a仅对无法识别的语音数据进行再次模型训练,能够减少数据冗余度,提高了识别速度和效率。

[0037] 此外,所述语音识别系统100a(或100)可进一步包括执行模块60,用于根据所述识别结果,生成特定格式的文本或播放对应的标准语音,并根据所述控制命令控制对应的客户端设备。而为了能够在不同的客户端设备200中运行所述语音识别系统100a,所述语音识别系统100a还可进一步包括下载模块70,用于供用户将对应私有云模块中的更新后的声学 and 语言模型下载到本地,以在本地实现语音识别。

[0038] 可以理解是,在其他实施例中,所述识别模块50对所述语音特征进行识别的同时,还可将全部所述语音特征存储在所述第一语音数据库21中,以供模型训练模块30定时从所述第一语音数据库21提取所述语音特征,从而进行模型训练。

[0039] 请参阅图4,本发明的一个实施例提供一种语音识别方法,所述方法包括以下步骤:

[0040] 步骤S401,基于启用实时通话或语音录入功能,实时输入用户的语音。具体的,所

述实时通话或语音录入功能通过手机、车载设备、电脑、手机、智能家居设备以及可穿戴设备等实现。同时,所述用户的语音也可进行实时保存,供后续调用。

[0041] 步骤S403,从所输入的用户语音中提取语音特征。本实施例中,提取到的语音特征被实时保存在一个第一语音数据库21中。其中,所述第一语音数据库21可以是本地数据库,也可以是云端数据库,所述语音特征指所述用户语音的特征数据。

[0042] 步骤S405,根据所述语音特征以及预设的规则,建立对应的声学 and 语言模型,以供在后续识别过程中,将提取的语音特征与所述声学 and 语言模型进行匹配与比较,得到最佳的识别结果。

[0043] 步骤S407,实时保存并更新所述声学 and 语言模型到一个模型数据库41中,由此,能够获取更庞大的声学 and 语言模型数据库41,提高了识别度。

[0044] 在本实施例中,步骤S401在客户端设备上执行,例如是通过麦克风及其处理电路进行语音输入。所述步骤S403、步骤S405、步骤S407在云端服务器300中执行。而为了能够对用户的语音信息进行保密,且针对不同的用户语音特点提供个性化的模型训练,所述云端服务器还包括对应于不同用户的多个私有云账户,每个私有云主账户可分别执行所述步骤S403~S407,当用户启用所述语音识别功能时,可通过帐号鉴权的方式进行。

[0045] 可以理解的是,在其他实施例中,所述步骤S401~S407均可在客户端设备200上执行,且所述第一语音数据库21和模型数据库41为本地数据库。

[0046] 请参阅图5,在又一实施例中,除了上述步骤S401~S407,所述语音识别方法进一步包括:

[0047] 步骤S409,根据模型数据库41中的所述声学 and 语言模型,判断是否能够识别所述语音特征,如果能够识别,则执行步骤S411,生成携带控制命令的识别结果,否则,执行步骤S413,将无法识别的其他语音特征存储到所述第一语音数据库21中。

[0048] 具体的,请参阅图6,所述步骤S409包括以下子步骤:

[0049] 子步骤S409a,将所述语音特征与所述声学 and 语言模型进行匹配度计算,如果匹配度大于等于阈值,则执行子步骤S409b,判断能够识别对应的所述语音特征并输出识别结果,否则,执行子步骤S409c,判断无法识别所述语音特征。

[0050] 子步骤S409d,根据用户命令,手动将所述匹配度低于所述阈值的无法识别的语音特征与预设的标准语音进行映射,并将所述语音特征与所述标准语音数据及其映射关系更新在一个第二语音数据库33中。

[0051] 此时,所述第一语音数据库21仅保存无法识别的所述语音特征,因此语音识别系统100仅需要对无法识别的语音数据进行再次模型训练,能够减少数据冗余度,提高了识别速度和效率。

[0052] 请参阅图7,在又一实施例中,结合步骤S401~S413,所述方法进一步包括:

[0053] 步骤S415,根据所述识别结果,生成特定格式的文本或播放对应的标准语音,并根据所述控制命令控制对应的客户端设备;

[0054] 步骤S417,将对应私有云模块中的更新后的声学 and 语言模型下载到本地,以在本地实现语音识别。

[0055] 又,在其他实施例中,对所述语音特征进行识别的同时,还可将全部所述语音特征存储在所述第一语音数据库21中,以定时、实时或定量从所述第一语音数据库21提取所述

语音特征,从而进行模型训练。

[0056] 本发明的语音识别系统和方法通过实时记录或保存实时通话和录音信息,并作为语音模型训练的样本,从而能够根据用户不同的发音特点持续更新模型数据库41。由此,可以满足用户的个性化需求,而且能够支持多种语音,例如英语或者地方方言等,提高了识别度。此外,本发明还提供了针对不同用户的私有云主模块(账户),供用户通过账户鉴权的方式启用语音识别功能,从而能够提高对用户语音信息的保密性能。

[0057] 需要说明的是,通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到本发明可借助软件加必需的硬件平台的方式来实现,当然也可以全部通过硬件来实现。基于这样的理解,本发明的技术方案对背景技术做出贡献的全部或者部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例或者实施例的某些部分所述的方法。

[0058] 以上所揭露的仅为本发明实施例中的较佳实施例而已,当然不能以此来限定本发明之权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

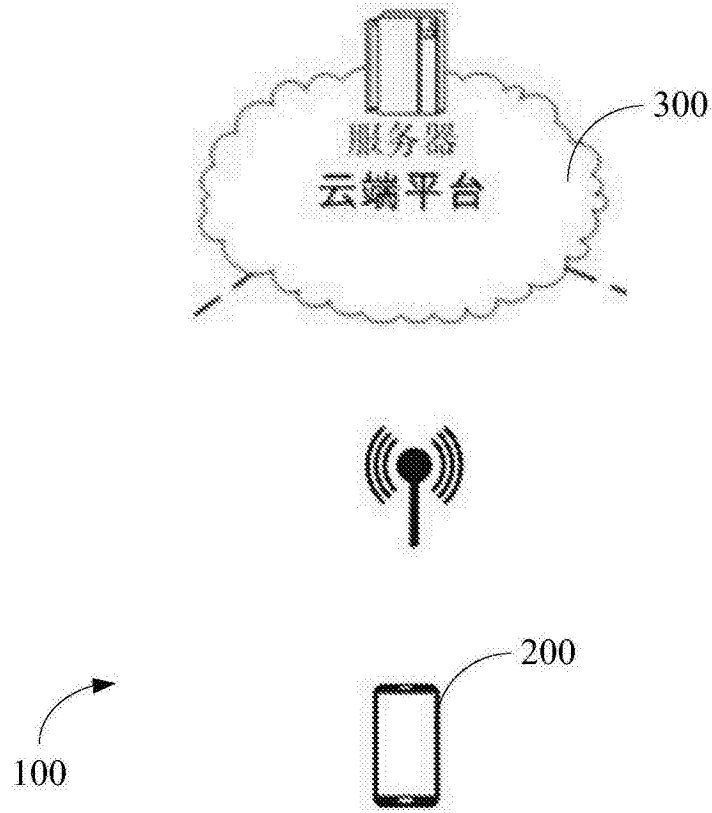


图1

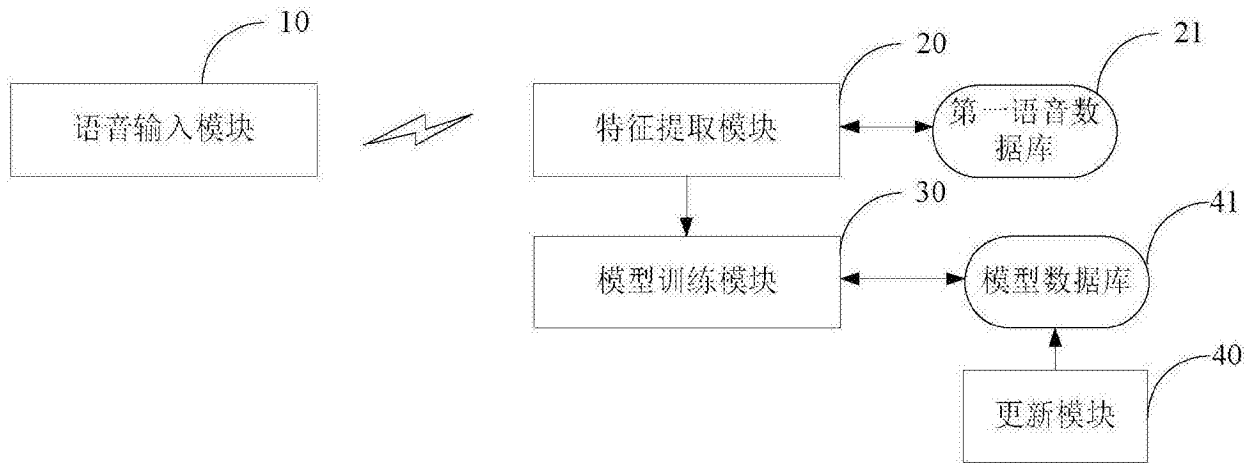


图2

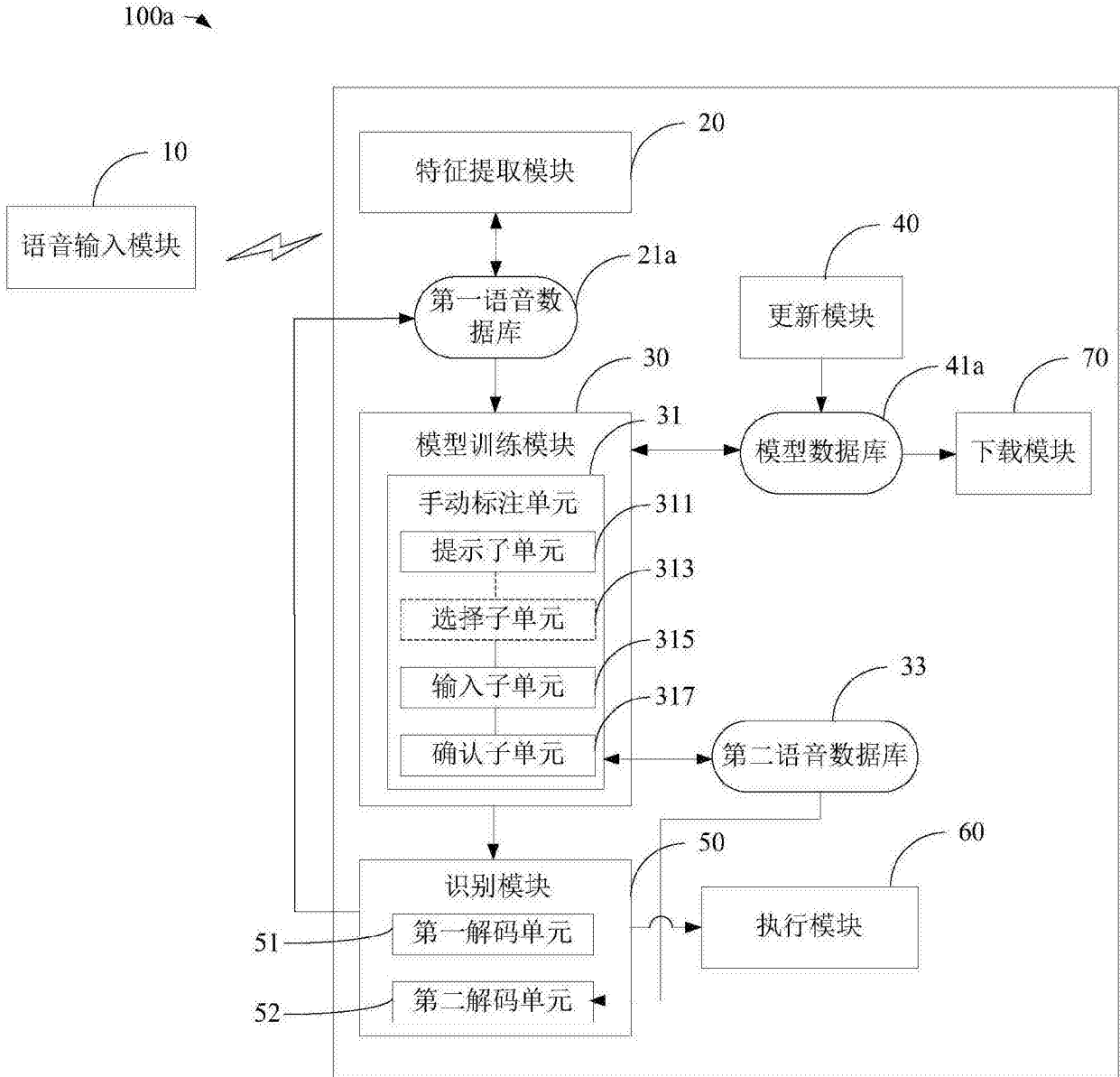


图3

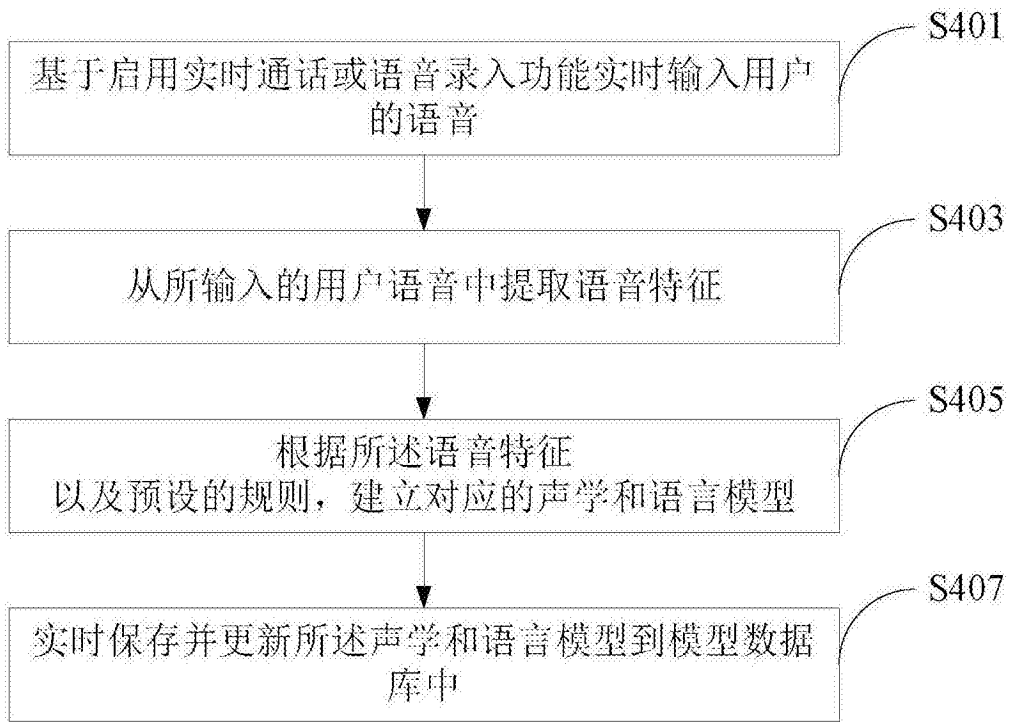


图4

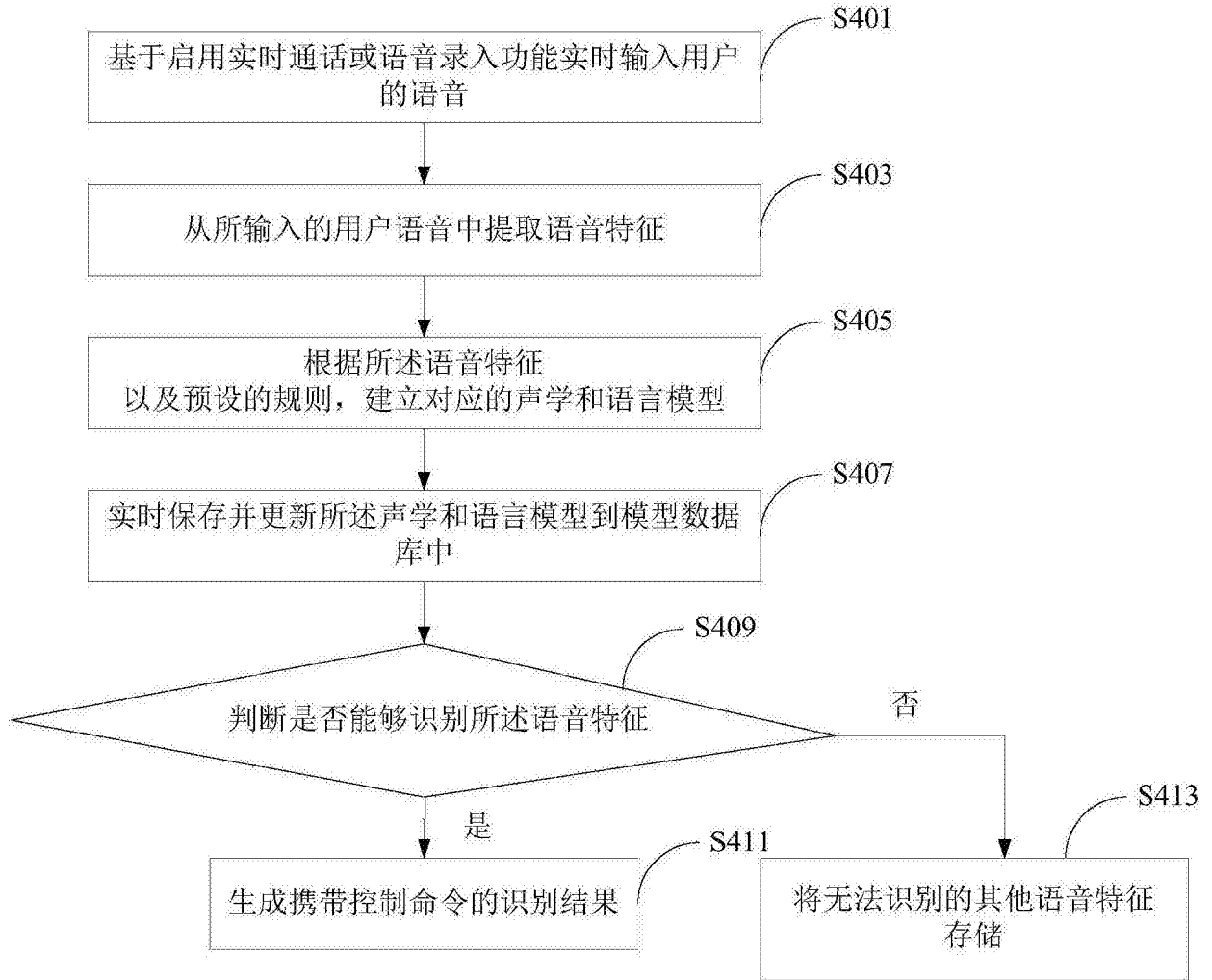


图5

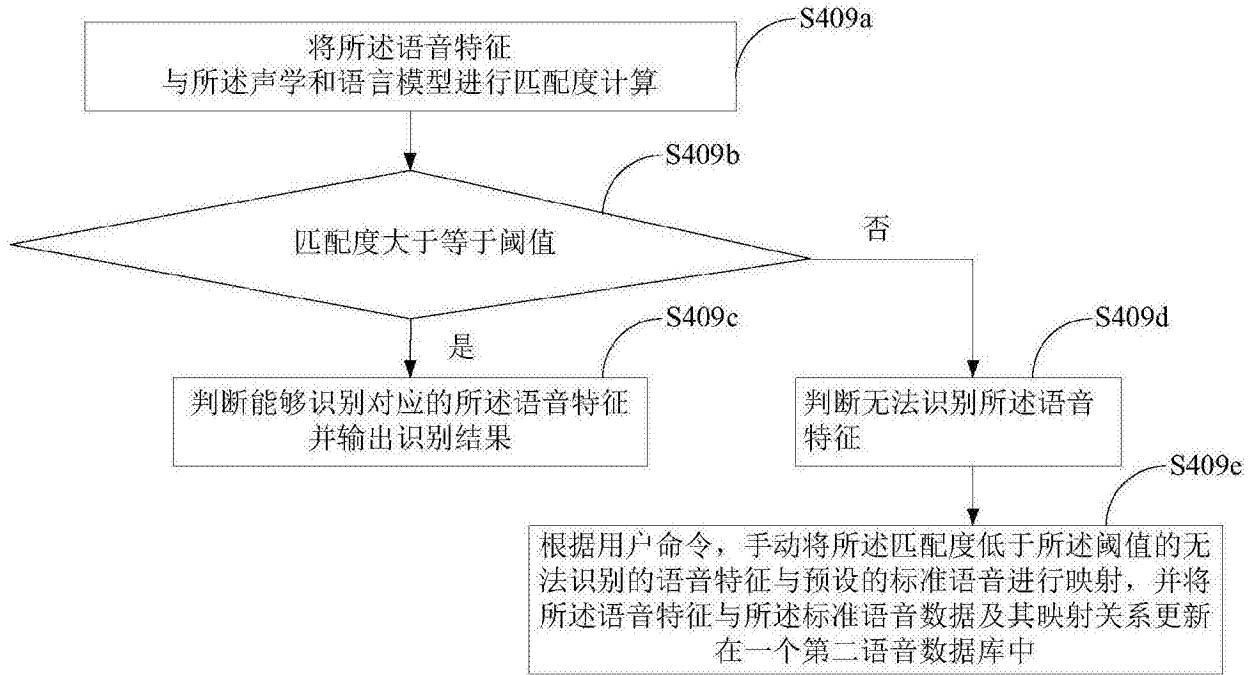


图6

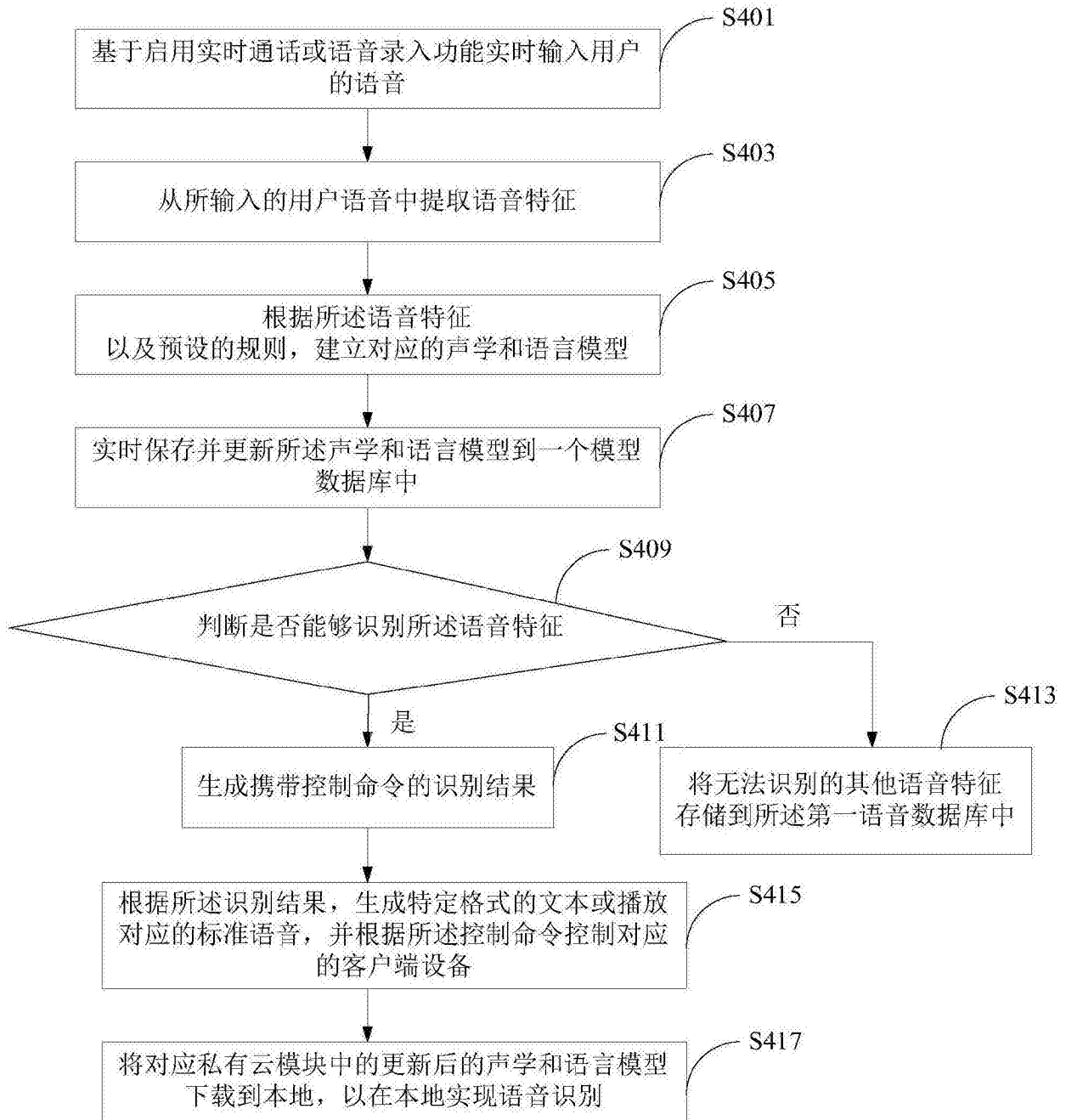


图7