(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111402931 B
(45)授权公告日 2023.05.26

(21)申请号 202010148900.6

(22)申请日 2020.03.05

(65)同一申请的已公布的文献号
申请公布号 CN 111402931 A

(43)申请公布日 2020.07.10

(73)专利权人 云知声智能科技股份有限公司
地址 100000 北京市海淀区西三旗建材城内1幢一层101号
专利权人 厦门云知芯智能科技有限公司

(72)发明人 高扬

(74)专利代理机构 北京冠和权律师事务所 11399
专利代理师 张楠楠

(51)Int.Cl.
G10L 25/78(2013.01)
G10L 25/87(2013.01)
G10L 15/05(2013.01)
G10L 15/06(2013.01)

(56)对比文件
CN 108962283 A,2018.12.07
CN 110047470 A,2019.07.23
CN 110400576 A,2019.11.01
JP 2015161718 A,2015.09.07
谢贵武等.基于语音分段的自适应时长调整.军事通信技术.2008,55-59页.
Thein Htay Zaw et al..The combination of spectral entropy, zero crossing rate, short time energy and linear prediction error for voice activity detection.2017 20th International Conference of Computer and Information Technology (ICCIT).2018,全文.

审查员 王立华

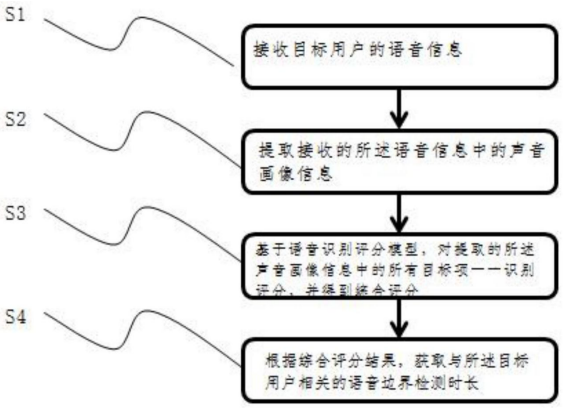
权利要求书2页 说明书6页 附图4页

(54)发明名称

一种利用声音画像辅助的语音边界检测方法及系统

(57)摘要

本发明提供了一种利用声音画像辅助的语音边界检测方法包括以下步骤:S1:接收目标用户的语音信息;S2:提取接收的语音信息中的声音画像信息;S3:基于语音识别评分模型,对提取的声音画像信息中的所有目标项一一识别评分,并得到综合评分;S4:根据综合评分结果,获取与目标用户相关的语音边界检测时长。本实施例提供的一种利用声音画像辅助的语音边界检测方法和设备可以根据不同的用户确定与之相适应的语音边界检测时长,提高语音识别成功率,进而提高用户的体验。



CN 111402931 B

1.一种利用声音画像辅助的语音边界检测方法,其特征在于,包括以下步骤:

S1:接收目标用户的语音信息;

S2:提取接收的所述语音信息中的声音画像信息;

S3:基于语音识别评分模型,对提取的所述声音画像信息中的所有目标项一一识别评分,并得到综合评分;

S4:根据综合评分结果,获取与所述目标用户相关的语音边界检测时长;

获取所述语音识别评分模型的步骤包括:

T1:获取历史用户的若干条待训练语音数据;

T2:基于获取的所述待训练语音数据,对待训练识别模型进行智能训练;

其中,所述待训练语音数据包括每个历史用户的声音画像信息中的每个目标项的评分及对应的所述历史用户的历史边界检测时长;

T3:当对所述待训练识别模型进行智能训练结束后,获得所述语音识别评分模型;

其中,所述历史用户的声音画像信息中的每个目标项的评分所对应的综合评分与历史边界检测时长呈一一对应关系;

S3步骤中,基于语音识别评分模型,对提取的所述声音画像信息中的所有目标项一一识别评分,并得到综合评分的步骤包括:

S31:基于所述语音识别评分模型,对同个所述目标用户的所述声音画像信息中的每个目标项进行单独评分;

S32:对每个目标项的单独评分结果进行综合处理,获得综合评分;

在执行S1步骤之前,包括:

P1:基于目标设备,预先录入所述目标用户的第一语音;

P2:提取P1步骤中的所述第一语音的语音特征并保存;

P3:录入P1步骤中所述目标用户的第一语音中的声音画像信息;

P4: P3步骤录入的声音画像信息经S3、S4步骤得到所述目标用户的第一语音检测时长并保存所述第一语音检测时长;

在S1步骤之后包括M步骤:所述M步骤包括:

M1:对S1步骤中接收的语音信息的语音特征与P2中保存的语音特征进行匹配;

如果未匹配成功,则进入S2步骤;

如果匹配成功,则进入M2步骤;

M2:将P4步骤保存的所述目标用户的第一语音检测时长确定为语音边界检测时长。

2.如权利要求1所述的方法,其特征在于,

所述目标项包括年龄项、语速项、表达流畅项。

3.一种利用声音画像辅助的语音边界检测系统,其特征在于,包括:

接收模块,用于接收目标用户的语音信息;

第一提取模块,用于提取所述接收模块接收的所述语音信息中的声音画像信息;

评分模块,用于基于语音识别评分模型,对提取的所述声音画像信息中的所有目标项一一识别评分,并得到综合评分;

第一处理模块,用于根据所述评分模块得到的综合评分结果,获取与所述目标用户相关的语音边界检测时长;

所述检测系统还包括：

第一获取模块，用于获取历史用户的若干条待训练语音数据；

训练模块，用于基于所述获取模块获取的所述待训练语音数据，对待训练识别模型进行智能训练；

其中，所述待训练语音数据包括每个历史用户的声音画像信息中的每个目标项的评分及对应的所述历史用户的历史边界检测时长；

第二获取模块，用于当对所述待训练识别模型进行智能训练结束后，获得所述语音识别评分模型；

其中，所述历史用户的声音画像信息中的每个目标项的评分所对应的综合评分与历史边界检测时长呈一一对应关系；

所述评分模块包括：

第一评分单元，用于基于所述语音识别评分模型，对同个所述目标用户的所述声音画像信息中的每个目标项进行单独评分；

第二评分单元，用于对所述第一评分单元评分得到的每个目标项的单独评分结果进行综合处理，获得综合评分；

所述检测系统还包括：

第一录入模块，用于基于目标设备，预先录入所述目标用户的第一语音；

第二提取模块，用于提取所述录入模块录入的所述第一语音的语音特征并保存；

第二录入模块，用于录入所述目标用户的第一语音中的声音画像信息；

第二处理模块，用于将所述第二录入模块录入的声音画像信息所述评分模块和第一处理模块得到所述目标用户的第一语音检测时长并保存所述第一语音检测时长；

在接收目标用户的语音信息之后，还包括：

匹配模块，用于对所述接收模块接收的语音信息的语音特征与所述第二提取模块保存的语音特征进行匹配；

如果未匹配成功，则控制所述第一提取模块开始工作；

如果匹配成功，则控制确定模块开始工作；

所述确定模块，用于将保存的所述目标用户的第一语音检测时长确定为语音边界检测时长。

4.如权利要求3所述的系统，其特征在于，

所述目标项包括年龄项、语速项、表达流畅项。

一种利用声音画像辅助的语音边界检测方法及系统

技术领域

[0001] 本发明涉及语音边界检测技术领域,特别涉及一种利用声音画像辅助的语音边界检测方法。

背景技术

[0002] 语音边界检测即语音活动检测(Voice Activity Detection,vad)又称语音端点检测。在一般的语音识别过程中,对于如儿童或者语速慢、语言表达不流畅的用户与设备交互的场景下,用户还没有表达完成,就开始进行语音识别,导致语音识别成功率较低。此时,就需要对语音边界检测时长进行检测,从而提高语音识别的成功率。

发明内容

[0003] 为了克服上述问题,本发明提供了一种利用声音画像辅助的语音边界检测方法,具体包括以下步骤:

[0004] S1:接收目标用户的语音信息;

[0005] S2:提取接收的所述语音信息中的声音画像信息;

[0006] S3:基于语音识别评分模型,对提取的所述声音画像信息中的所有目标项一一识别评分,并得到综合评分;

[0007] S4:根据综合评分结果,获取与所述目标用户相关的语音边界检测时长。

[0008] 优选地,获取所述语音识别评分模型的步骤包括:

[0009] T1:获取历史用户的若干条待训练语音数据;

[0010] T2:基于获取的所述待训练语音数据,对待训练识别模型进行智能训练;

[0011] 其中,所述待训练语音数据包括每个历史用户的声音画像信息中的每个目标项的评分及对应的所述历史用户的历史边界检测时长;

[0012] T3:当对所述待训练识别模型进行智能训练结束后,获得所述语音识别评分模型;

[0013] 其中,所述历史用户的声音画像信息中的每个目标项的评分所对应的综合评分与历史边界检测时长呈一一对应关系。

[0014] 优选地,所述目标项包括年龄项、语速项、表达流畅项。

[0015] 优选地,S3步骤中,基于语音识别评分模型,对提取的所述声音画像信息中的所有目标项一一识别评分,并得到综合评分的步骤包括:

[0016] S31:基于所述语音识别评分模型,对同个所述目标用户的所述声音画像信息中的每个目标项进行单独评分;

[0017] S32:对每个目标项的单独评分结果进行综合处理,获得综合评分。

[0018] 优选地,在执行S1步骤之前,包括:

[0019] P1:基于目标设备,预先录入所述目标用户的第一语音;

[0020] P2:提取P1步骤中的所述第一语音的语音特征并保存;

[0021] P3:录入P1步骤中所述目标用户的第一语音中的声音画像信息;

[0022] P4：P3步骤录入的声音画像信息经S3、S4步骤得到所述目标用户的第一语音检测时长并保存所述第一语音检测时长；

[0023] 在S1步骤之后包括M步骤：所述M步骤包括：

[0024] M1：对S1步骤中接收的语音信息的语音特征与P2中保存的语音特征进行匹配；

[0025] 如果未匹配成功，则进入S2步骤；

[0026] 如果匹配成功，则进入M2步骤；

[0027] M2：将P4步骤保存的所述目标用户的第一语音检测时长确定为语音边界检测时长。

[0028] 本发明实施例提供一种利用声音画像辅助的语音边界检测系统，包括：

[0029] 接收模块，用于接收目标用户的语音信息；

[0030] 第一提取模块，用于提取所述接收模块接收的所述语音信息中的声音画像信息；

[0031] 评分模块，用于基于语音识别评分模型，对提取的所述声音画像信息中的所有目标项一一识别评分，并得到综合评分；

[0032] 第一处理模块，用于根据所述评分模块得到的综合评分结果，获取与所述目标用户相关的语音边界检测时长。

[0033] 优选地，还包括：

[0034] 第一获取模块，用于获取历史用户的若干条待训练语音数据；

[0035] 训练模块，用于基于所述获取模块获取的所述待训练语音数据，对待训练识别模型进行智能训练；

[0036] 其中，所述待训练语音数据包括每个历史用户的声音画像信息中的每个目标项的评分及对应的所述历史用户的历史边界检测时长；

[0037] 第二获取模块，用于当对所述待训练识别模型进行智能训练结束后，获得所述语音识别评分模型；

[0038] 其中，所述历史用户的声音画像信息中的每个目标项的评分所对应的综合评分与历史边界检测时长呈一一对应关系。

[0039] 优选地，所述目标项包括年龄项、语速项、表达流畅项。

[0040] 优选地，所述评分模块包括：

[0041] 第一评分单元，用于基于所述语音识别评分模型，对同个所述目标用户的所述声音画像信息中的每个目标项进行单独评分；

[0042] 第二评分单元，用于对所述第一评分单元评分得到的每个目标项的单独评分结果进行综合处理，获得综合评分。

[0043] 优选地，还包括：

[0044] 第一录入模块，用于基于目标设备，预先录入所述目标用户的第一语音；

[0045] 第二提取模块，用于提取所述录入模块录入的所述第一语音的语音特征并保存；

[0046] 第二录入模块，用于录入所述目标用户的第一语音中的声音画像信息；

[0047] 第二处理模块，用于将所述第二录入模块录入的声音画像信息所述评分模块和第一处理模块得到所述目标用户的第一语音检测时长并保存所述第一语音检测时长；

[0048] 在接收目标用户的语音信息之后，还包括：

[0049] 匹配模块，用于对所述接收模块接收的语音信息的语音特征与所述第二提取模块

保存的语音特征进行匹配；

[0050] 如果未匹配成功，则控制所述第一提取模块开始工作；

[0051] 如果匹配成功，则控制确定模块开始工作；

[0052] 所述确定模块，用于将保存的所述目标用户的第一语音检测时长确定为语音边界检测时长。

[0053] 本实施例提供的一种利用声音画像辅助的语音边界检测方法和系统可以根据不同的用户确定与之相适应的语音边界检测时长，提高语音识别成功率，进而提高用户的体验。

[0054] 本发明的其它特征和优点将在随后的说明书中阐述，并且，部分地从说明书中变得显而易见，或者通过实施本发明而了解。本发明的目的和其他优点可通过在所写的说明书、权利要求书、以及附图中所特别指出的结构来实现和获得。

[0055] 下面通过附图和实施例，对本发明的技术方案做进一步的详细描述。

附图说明

[0056] 附图用来提供对本发明的进一步理解，并且构成说明书的一部分，与本发明的实施例一起用于解释本发明，并不构成对本发明的限制。在附图中：

[0057] 图1: 一种利用声音画像辅助的语音边界检测方法的示意图；

[0058] 图2:智能训练的示意图；

[0059] 图3:利用声音画像辅助的语音边界检测方法进行综合评分的示意图；

[0060] 图4 声音画像信息的语音边界检测方法的示意图；

[0061] 图5:一种利用声音画像辅助的语音边界检测系统的结构图。

具体实施方式

[0062] 以下结合附图对本发明的优选实施例进行说明，应当理解，此处所描述的优选实施例仅用于说明和解释本发明，并不用于限定本发明。

[0063] 本实施例提供了一种利用声音画像辅助的语音边界检测方法，如图1所示，包括以下步骤：

[0064] S1:接收目标用户的语音信息。

[0065] S2:提取接收的所述语音信息中的声音画像信息。本实施例中，根据用户语音提取的声音画像信息为年龄、语速、表达流畅度信息，其中语速分为快、中、慢，表达流畅度分为好、中、次。

[0066] S3:基于语音识别评分模型，对提取的所述声音画像信息中的所有目标项一一识别评分，并得到综合评分。本实施例中，用户年龄在7-60岁之间评为9分，年龄在3-6岁之间评分为2分，年龄在60岁以上评分为6分；用户语速快(>150字/分)评分为9，语速中(120-150字/分)评分为7，语速慢(<120字/分)评分为3；表达流畅度好评分为9分，表达流畅度中(语音信号有偶尔不连续的)评分为6分，表达流畅度次(语音信号断断续续，且不稳定)评分为2分。

[0067] S4:根据综合评分结果，获取与所述目标用户相关的语音边界检测时长。本实施例中具体为根据S3步骤中的三个维度的评分得出适合不同分数段的语音边界检测时长。

[0068] 上述技术方案的工作原理为：

[0069] 本实施例中通过对用户语音声音画像信息提取、评分来确定语音边界检测时长。

[0070] 上述技术方案的有益效果为：可以根据不同的用户确定与之相适应的语音边界检测时长，提高语音识别成功率，进而提高用户的体验。

[0071] 在一个实施例中，如图2所示

[0072] 获取所述语音识别评分模型的步骤包括：

[0073] T1：获取历史用户的若干条待训练语音数据；

[0074] T2：基于获取的所述待训练语音数据，对待训练识别模型进行智能训练；

[0075] 其中，所述待训练语音数据包括每个历史用户的声音画像信息中的每个目标项的评分及对应的所述历史用户的历史边界检测时长；

[0076] T3：当对所述待训练识别模型进行智能训练结束后，获得所述语音识别评分模型；

[0077] 其中，所述历史用户的声音画像信息中的每个目标项的评分所对应的综合评分与历史边界检测时长呈一一对应关系。

[0078] 上述技术方案的工作原理为：T2步骤的智能训练，是通过对每个待训练语音数据进行年龄、语速和表达流畅的预先标志评分，都是提前设定好的训练样本，通过该样本对待训练识别模型进行智能训练，可以使得评分与历史边界检测时长的对应关系更加得准确。

[0079] 上述技术方案的有益效果为：有助于更准确地确定语音边界检测时长。

[0080] 在一个实施例中，如图3所示

[0081] S3步骤中，基于语音识别评分模型，对提取的所述声音画像信息中的所有目标项一一识别评分，并得到综合评分的步骤包括：

[0082] S31：基于所述语音识别评分模型，对同个所述目标用户的所述声音画像信息中的每个目标项进行单独评分；

[0083] S32：对每个目标项的单独评分结果进行综合处理，获得综合评分。

[0084] 具体，本实施例是取三项评分求和后取平均值。

[0085] 本实施例可以为根据S32步骤得到的平均值确定语音边界检测时长。

[0086] 具体为：评分平均值在0-4之间的，语音边界检测时长设为600ms；评分平均值在5-7之间的，语音边界检测时长设为400ms；评分平均值在8-10之间的，语音边界检测时长设置为100ms。

[0087] 当然，在不同的应用场景中对语音边界检测时长可以做出相应的调整，在本实施例中不再详述。

[0088] 本实施例给出了具体的一种评分的方法。

[0089] 在一个实施例中，如图4所示

[0090] 在S1步骤之前还包括P步骤，所述P步骤包括：

[0091] P1：基于目标设备，预先录入所述目标用户的第一语音；

[0092] P2：提取P1步骤中的所述第一语音的语音特征并保存；

[0093] 所述的语音特征是指用户的声音特征，包括振幅、频率、音色，其中的音色具体表现为声音声音的频率表现在波形方面总是有与众不同的特性。本实施例具体为保存语音的振幅、频率和时间的分布关系，用于后续的通过三维语图分析进行匹配。

[0094] P3：录入P1步骤中所述目标用户的第一语音中的声音画像信息；

[0095] 具体的，本实施例是通过手动录入声音画像信息的每个目标项。具体为：录入用户年龄；录入用户的快、中、慢；表达流畅度的好、中、次。

[0096] P4：P3步骤录入的声音画像信息经S3、S4步骤得到所述目标用户的第一语音检测时长并保存所述第一语音检测时长；具体可见第一个实施例。

[0097] 在S1步骤之后包括M步骤：所述M步骤包括：

[0098] M1：对S1步骤中接收的语音信息的语音特征与P2中保存的语音特征进行匹配；

[0099] 如果未匹配成功，则进入S2步骤；

[0100] 如果匹配成功，则进入M2步骤；

[0101] M2：将P4步骤保存的所述目标用户的第一语音检测时长确定为语音边界检测时长。

[0102] 上述技术方案的工作原理为：本实施例的方案是通过P3-P4步骤预先设置用户的第一语音检测时长。当接收用户语音时，首先进行识别，如果接收到的语音与经P 1-P2步骤保存的语音特征匹配，则直接调取该语音边界检测时长。如果未设置，则通过S2-S4步骤确认语音边界检测时长。

[0103] 上述技术方案的有益效果为：预先对特定的用户设置语音边界检测时长，则其信息更准，有助于提高语音识别成功率，进而提高用户的体验。

[0104] 本实施例提供了一种利用声音画像辅助的语音边界检测系统，如图5所示，包括：

[0105] 接收模块，用于接收目标用户的语音信息；

[0106] 第一提取模块，用于提取所述接收模块接收的所述语音信息中的声音画像信息；

[0107] 评分模块，用于基于语音识别评分模型，对提取的所述声音画像信息中的所有目标项一一识别评分，并得到综合评分；

[0108] 第一处理模块，用于根据所述评分模块得到的综合评分结果，获取与所述目标用户相关的语音边界检测时长。

[0109] 上述技术方案的有益效果为：可以根据不同的用户确定与之相适应的语音边界检测时长，提高语音识别成功率，进而提高用户的体验。

[0110] 在一个实施例中，还包括：

[0111] 第一获取模块，用于获取历史用户的若干条待训练语音数据；

[0112] 训练模块，用于基于所述获取模块获取的所述待训练语音数据，对待训练识别模型进行智能训练；

[0113] 其中，所述待训练语音数据包括每个历史用户的声音画像信息中的每个目标项的评分及对应的所述历史用户的历史边界检测时长；

[0114] 第二获取模块，用于当对所述待训练识别模型进行智能训练结束后，获得所述语音识别评分模型；

[0115] 其中，所述历史用户的声音画像信息中的每个目标项的评分所对应的综合评分与历史边界检测时长呈一一对应关系。

[0116] 上述技术方案的有益效果为：有助于语音边界检测时长确定模块根据评分模块的评分更准确地确定语音边界检测时长。

[0117] 在一个实施例中，评分模块包括：

[0118] 第一评分单元，用于基于所述语音识别评分模型，对同个所述目标用户的所述声

音画像信息中的每个目标项进行单独评分；

[0119] 第二评分单元，用于对所述第一评分单元评分得到的每个目标项的单独评分结果进行综合处理，获得综合评分。

[0120] 所述评分模块分别对所述声音画像信息中的每个目标项进行单独评分，并根据每个单独评分结果，进而进行相应的综合评分；

[0121] 具体的，本实施例中的综合评分是所述声音画像信息中每个目标项的评分的平均分。

[0122] 给出了一种评分模块的评分的方案。

[0123] 在一个实施例中，还包括：

[0124] 第一录入模块，用于基于目标设备，预先录入所述目标用户的第一语音；

[0125] 第二提取模块，用于提取所述录入模块录入的所述第一语音的语音特征并保存；

[0126] 第二录入模块，用于录入所述目标用户的第一语音中的声音画像信息；

[0127] 第二处理模块，用于将所述第二录入模块录入的声音画像信息所述评分模块和第一处理模块得到所述目标用户的第一语音检测时长并保存所述第一语音检测时长；

[0128] 在接收目标用户的语音信息之后，还包括：

[0129] 匹配模块，用于对所述接收模块接收的语音信息的语音特征与所述第二提取模块保存的语音特征进行匹配；

[0130] 如果未匹配成功，则控制所述第一提取模块开始工作；

[0131] 如果匹配成功，则控制确定模块开始工作；

[0132] 所述确定模块，用于将保存的所述目标用户的第一语音检测时长确定为语音边界检测时长。

[0133] 上述进行保存，一般是将其数据保存到了存储器中。

[0134] 上述技术方案的有益效果为：预先对特定的用户设置语音边界检测时长，则其信息更准，有助于提高语音识别成功率，进而提高用户的体验。

[0135] 显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

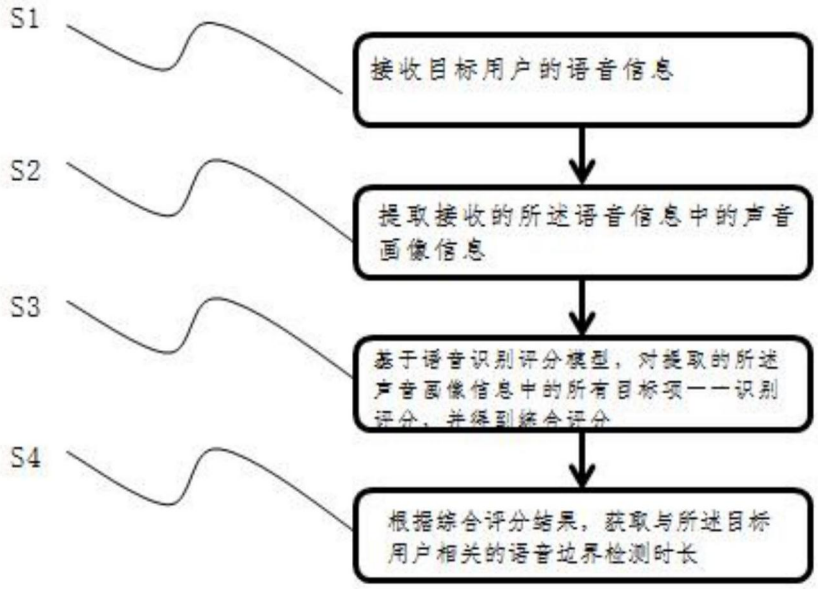

S1
接收目标用户的语音信息
S2
提取接收的所述语音信息中的声音画像信息
S3
基于语音识别评分模型，对提取的所述声音画像信息中的所有目标项一一识别评分，并得到综合评分
S4
根据综合评分结果，获取与所述目标用户相关的语音边界检测时长

图1

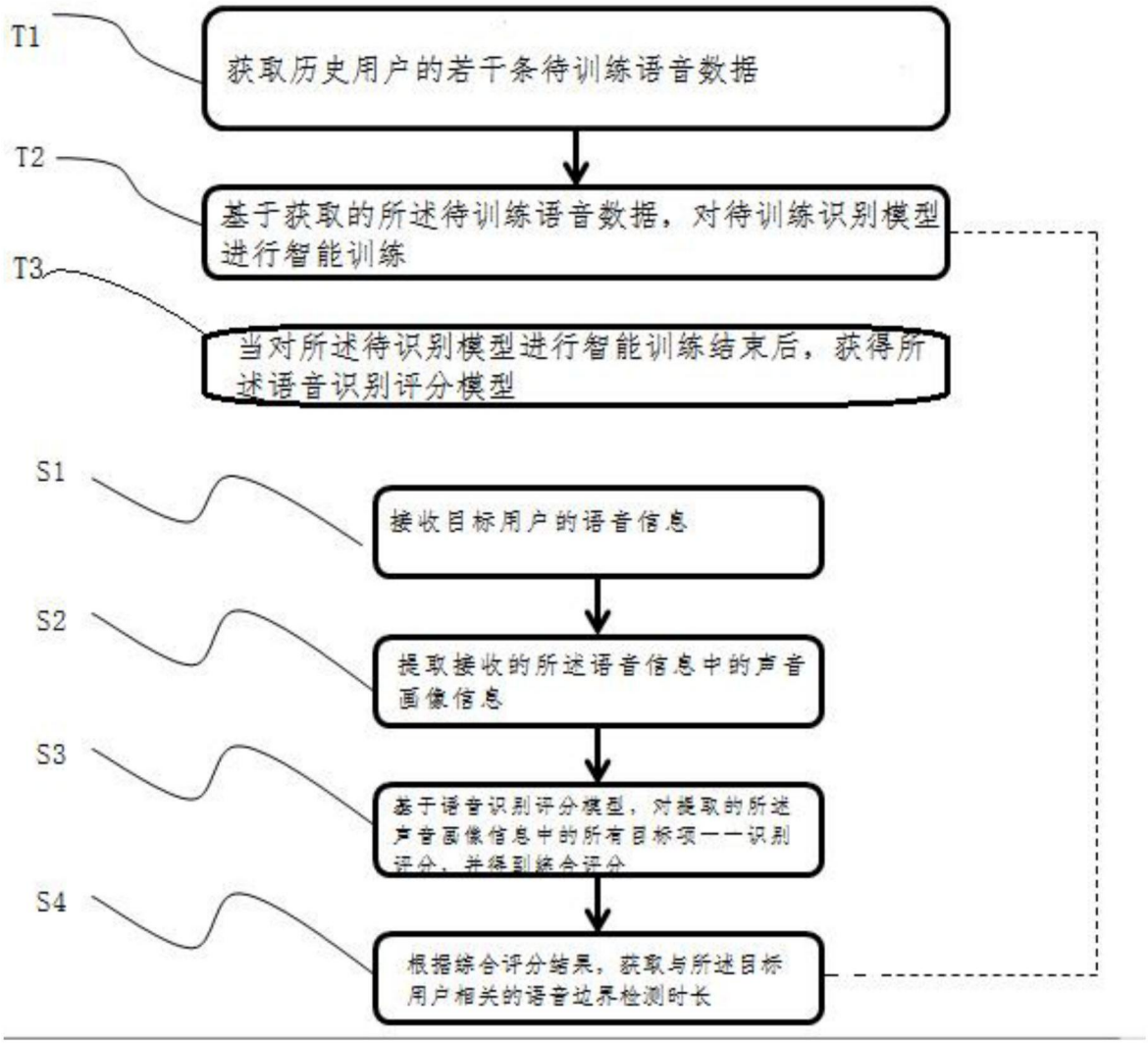

T1
获取历史用户的若干条待训练语音数据
T2
基于获取的所述待训练语音数据，对待训练识别模型进行智能训练
T3
当对所述待识别模型进行智能训练结束后，获得所述语音识别评分模型
S1
接收目标用户的语音信息
S2
提取接收的所述语音信息中的声音画像信息
S3
基于语音识别评分模型，对提取的所述声音画像信息中的所有目标项一一识别评分，并得到综合评分
S4
根据综合评分结果，获取与所述目标用户相关的语音边界检测时长

图2

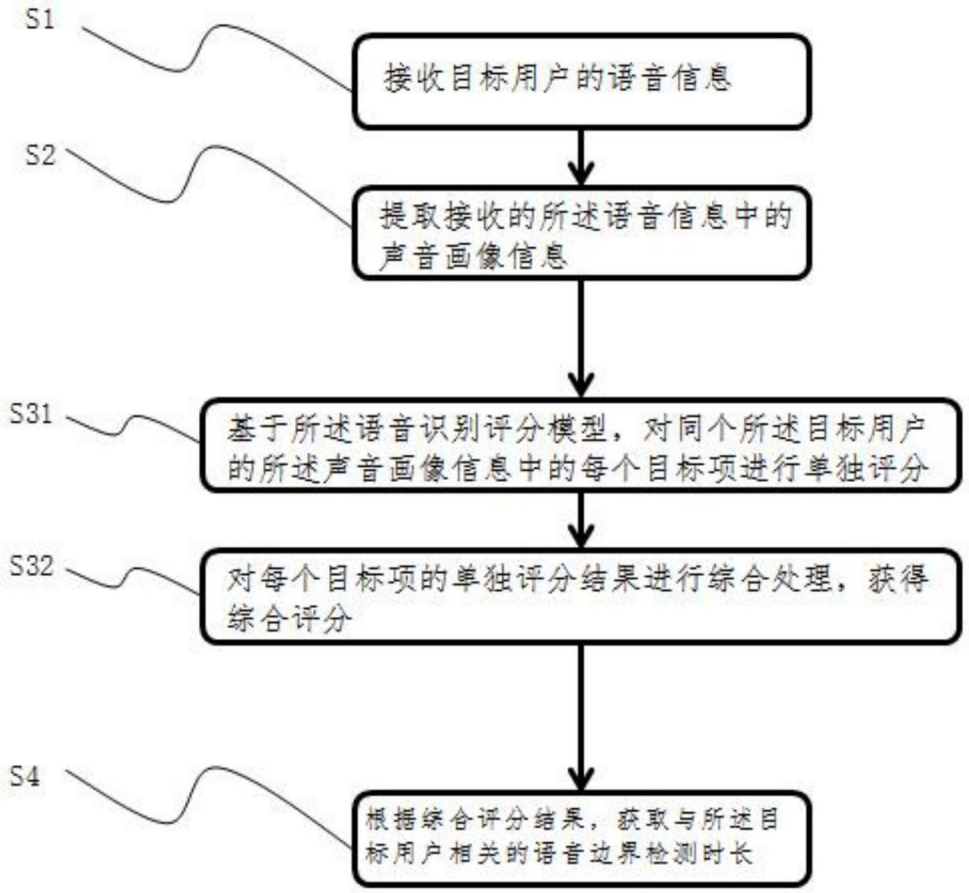

S1
接收目标用户的语音信息
S2
提取接收的所述语音信息中的声音画像信息
S31
基于所述语音识别评分模型，对同个所述目标用户的所述声音画像信息中的每个目标项进行单独评分
S32
对每个目标项的单独评分结果进行综合处理，获得综合评分
S4
根据综合评分结果，获取与所述目标用户相关的语音边界检测时长

图3

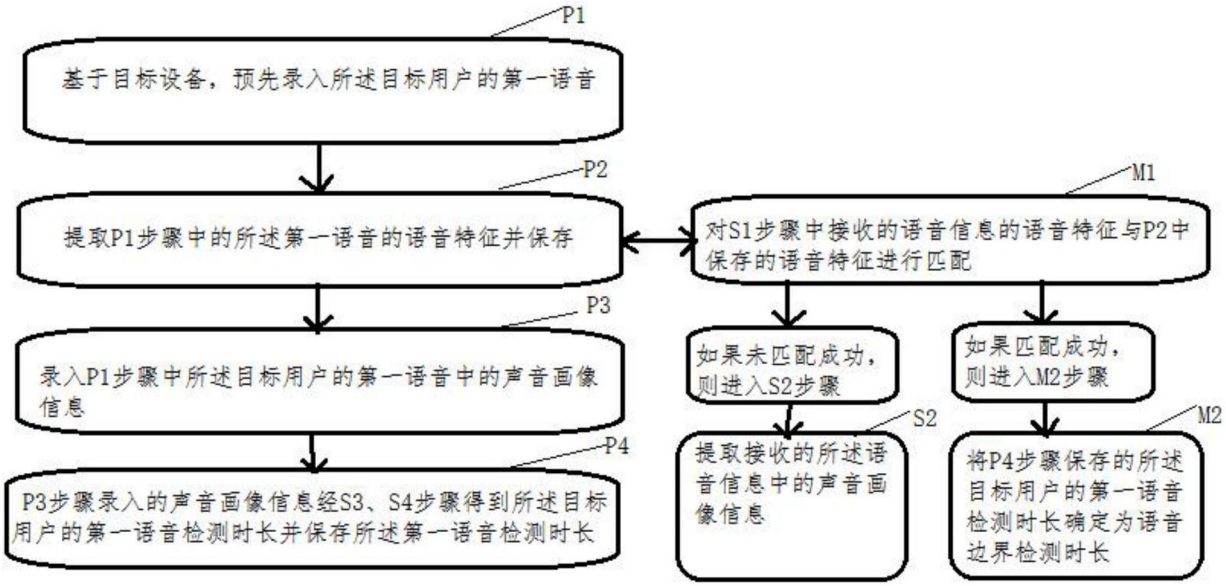

P1
基于目标设备，预先录入所述目标用户的第一语音
P2
提取P1步骤中的所述第一语音的语音特征并保存
P3
录入P1步骤中所述目标用户的第一语音中的声音画像信息
P4
P3步骤录入的声音画像信息经S3、S4步骤得到所述目标用户的第一语音检测时长并保存所述第一语音检测时长
M1
对S1步骤中接收的语音信息的语音特征与P2中保存的语音特征进行匹配
如果未匹配成功，则进入S2步骤
如果匹配成功，则进入M2步骤
S2
提取接收的所述语音信息中的声音画像信息
M2
将P4步骤保存的所述目标用户的第一语音检测时长确定为语音边界检测时长

图4

图5