



(12) 发明专利申请

(10) 申请公布号 CN 118871985 A

(43) 申请公布日 2024. 10. 29

(21) 申请号 202380030965.2

(22) 申请日 2023.03.02

(30) 优先权数据

2022-053033 2022.03.29 JP

(85) PCT国际申请进入国家阶段日

2024.09.26

(86) PCT国际申请的申请数据

PCT/JP2023/007820 2023.03.02

(87) PCT国际申请的公布数据

W02023/189173 JA 2023.10.05

(71) 申请人 松下电器(美国)知识产权公司

地址 美国

(72) 发明人 土井美沙贵

(74) 专利代理机构 中科专利商标代理有限责任公司 11021

专利代理师 韩丁

(51) Int.Cl.

G10L 17/22 (2006.01)

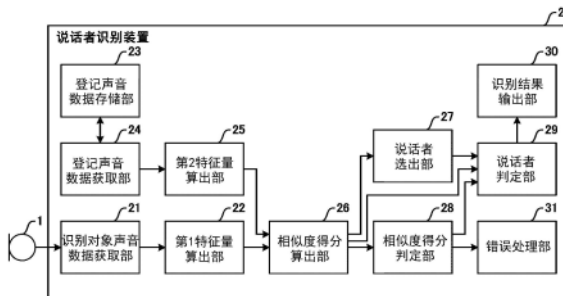
权利要求书2页 说明书15页 附图9页

(54) 发明名称

说话者识别方法、说话者识别装置以及说话者识别程序

(57) 摘要

说话者识别装置获取识别对象声音数据,获取预先登记的多个登记声音数据,算出识别对象声音数据与多个登记声音数据各自的相似度,选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者,基于所算出的多个相似度来判定识别对象声音数据是否适于说话者识别,在判定为识别对象声音数据适于说话者识别的情况下,基于最高的相似度来判定是否将所选出的登记说话者识别为识别对象声音数据的识别对象说话者,输出识别结果。



1. 一种说话者识别方法,是计算机中的说话者识别方法,包括:
获取识别对象声音数据;
获取预先登记的多个登记声音数据;
算出所述识别对象声音数据与所述多个登记声音数据各自的相似度;
选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者;
基于所算出的所述多个相似度来判定所述识别对象声音数据是否适于说话者识别;
在判定为所述识别对象声音数据适于所述说话者识别的情况下,基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者;
输出识别结果。
2. 根据权利要求1所述的说话者识别方法,其中,
在所述识别对象声音数据是否适于所述说话者识别的判定中,判定所算出的所述多个相似度当中的最高的相似度是否比第1阈值高,在判定为最高的所述相似度比所述第1阈值高的情况下,判定为所述识别对象声音数据适于所述说话者识别。
3. 根据权利要求1所述的说话者识别方法,其中,
在所述识别对象声音数据是否适于所述说话者识别的判定中,将所算出的所述多个相似度的方差值算出,判定所算出的所述方差值是否比第1阈值高,在判定为所述方差值比所述第1阈值高的情况下,判定为所述识别对象声音数据适于所述说话者识别。
4. 根据权利要求2或3所述的说话者识别方法,其中,
在是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者的判定中,判定所算出的所述多个相似度当中的最高的相似度是否比高于所述第1阈值的第2阈值高,在判定为最高的所述相似度比所述第2阈值高的情况下,将所选出的所述登记说话者识别为所述识别对象声音数据的所述识别对象说话者。
5. 根据权利要求1所述的说话者识别方法,其中,
所述多个登记声音数据包含对识别对象的多个登记说话者所发声的声音预先进行登记而得到的多个第1登记声音数据、和对所述识别对象的所述多个登记说话者以外的多个其他登记说话者所发声的声音预先进行登记而得到的多个第2登记声音数据,
在所述相似度的算出中,算出所述识别对象声音数据与所述多个第1登记声音数据各自的第1相似度,并且算出所述识别对象声音数据与所述多个第2登记声音数据各自的第2相似度,
在所述登记说话者的选出中,选出与所算出的多个第1相似度当中的最高的第1相似度对应的第1登记声音数据的登记说话者,
在所述识别对象声音数据是否适于所述说话者识别的判定中,判定所算出的所述多个第1相似度以及所述多个第2相似度当中的最高的第1相似度或第2相似度是否比第1阈值高,在判定为最高的所述第1相似度或所述第2相似度比所述第1阈值高的情况下,判定为所述识别对象声音数据适于所述说话者识别。
6. 根据权利要求5所述的说话者识别方法,其中,
所述多个第2登记声音数据不包含噪声,仅包含所述其他登记说话者所发声的所述声音。

7. 根据权利要求5或6所述的说话者识别方法, 其中,

在是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者的判定中, 判定所算出的所述多个第1相似度当中的最高的第1相似度是否比高于所述第1阈值的第2阈值高, 在判定为最高的所述第1相似度比所述第2阈值高的情况下, 将所选出的所述登记说话者识别为所述识别对象声音数据的所述识别对象说话者。

8. 根据权利要求1~3中任一项所述的说话者识别方法, 其中,

进一步地在判定为所述识别对象声音数据不适于所述说话者识别的情况下, 对催促所述识别对象说话者再输入所述识别对象声音数据的错误消息进行输出。

9. 根据权利要求1~3中任一项所述的说话者识别方法, 其中,

在所述识别对象声音数据的获取中, 获取从所述识别对象说话者所发声的声音数据截取给定的区间而得到的所述识别对象声音数据,

进一步地在判定为所述识别对象声音数据不适于所述说话者识别的情况下, 获取从所述声音数据截取与所述给定的区间不同的区间而得到的其他识别对象声音数据。

10. 一种说话者识别装置, 具备:

识别对象声音数据获取部, 其获取识别对象声音数据;

登记声音数据获取部, 其获取预先登记的多个登记声音数据;

算出部, 其算出所述识别对象声音数据与所述多个登记声音数据各自的相似度;

选出部, 其选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者;

相似度判定部, 其基于所算出的所述多个相似度来判定所述识别对象声音数据是否适于说话者识别;

说话者判定部, 其在判定为所述识别对象声音数据适于所述说话者识别的情况下, 基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者; 和

输出部, 其输出识别结果。

11. 一种说话者识别程序, 使计算机如下地发挥功能:

获取识别对象声音数据;

获取预先登记的多个登记声音数据;

算出所述识别对象声音数据与所述多个登记声音数据各自的相似度;

选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者;

基于所算出的所述多个相似度来判定所述识别对象声音数据是否适于说话者识别;

在判定为所述识别对象声音数据适于所述说话者识别的情况下, 基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者;

输出识别结果。

说话者识别方法、说话者识别装置以及说话者识别程序

技术领域

[0001] 本公开涉及识别说话者的技术。

背景技术

[0002] 例如,在专利文献1中公开了一种噪声抑制声音辨识装置,该噪声抑制声音辨识装置从输入声音中以帧为单位提取音频特征量,检测输入声音的声音区间,检测噪声的每个种类的噪声区间,选择噪声抑制方法,生成用所选择的噪声抑制方法抑制了噪声的音频特征量的音频特征量,通过所生成的音频特征量来进行声音辨识。

[0003] 但在上述现有的技术那样通过信号处理来抑制输入声音的噪声的情况下,有可能使说话者的个人的特征失真,其结果,说话者辨识的精度有可能会降低。因此,在上述现有的技术中,需要进一步的改善。

[0004] 在先技术文献

[0005] 专利文献

[0006] 专利文献1:JP特开2016-180839号公报

发明内容

[0007] 本公开为了解决上述的问题而提出,目的在于,提供不使计算量增加就能够提升对识别对象的说话者是预先登记的多个说话者中的哪一个进行识别的精度技术。

[0008] 本公开所涉及的说话者识别方法是计算机中的说话者识别方法,所述说话者识别方法获取识别对象声音数据,获取预先登记的多个登记声音数据,算出所述识别对象声音数据与所述多个登记声音数据各自的相似度,选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者,基于所算出的所述多个相似度来判定所述识别对象声音数据是否适于说话者识别,在判定为所述识别对象声音数据适于所述说话者识别的情况下,基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者,并输出识别结果。

[0009] 根据本公开,不使计算量增加就能够提升对识别对象的说话者是预先登记的多个说话者中的哪一个进行识别的精度。

附图说明

[0010] 图1是表示本公开的实施方式1中的说话者识别系统的结构的图。

[0011] 图2是用于说明本实施方式1中的说话者识别装置的说话者识别处理的动作的第1流程图。

[0012] 图3是用于说明本实施方式1中的说话者识别装置的说话者识别处理的动作的第2流程图。

[0013] 图4是表示本公开的实施方式2中的说话者识别系统的结构的图。

[0014] 图5是用于说明本实施方式2中的说话者识别装置的说话者识别处理的动作的第1

流程图。

[0015] 图6是用于说明本实施方式2中的说话者识别装置的说话者识别处理的动作的第2流程图。

[0016] 图7是表示本公开的实施方式3中的说话者识别系统的结构的图。

[0017] 图8是用于说明本实施方式3中的说话者识别装置的说话者识别处理的动作的第1流程图。

[0018] 图9是用于说明本实施方式3中的说话者识别装置的说话者识别处理的动作的第2流程图。

具体实施方式

[0019] (成为本公开的基础的见解)

[0020] 过去,已知有如下的说话者识别:获取识别对象说话者的输入声音数据,基于所获取的输入声音数据和预先登记的多个登记声音数据来对识别对象说话者是预先登记的多个说话者中的哪一个进行识别。在现有的说话者识别中,分别算出识别对象说话者的输入声音数据的特征量与多个登记说话者的登记声音数据的特征量的相似度得分。然后,将与算出的多个相似度得分当中的最高的相似度得分对应的登记声音数据的登记说话者识别为识别对象说话者。

[0021] 但在现有的说话者识别中,在识别对象说话者的输入声音数据中包含噪声的情况下,或者在输入声音数据中不包含识别对象说话者的声音的情况下,虽然也输出说话者识别结果,但利用了包含噪声的输入声音数据或不包含识别对象说话者的声音的输入声音数据的说话者识别的精度变低。

[0022] 与此相对,根据上述的专利文献1的噪声抑制声音辨识装置,检测输入声音的声音区间,抑制声音区间的噪声来进行声音辨识。

[0023] 但在上述现有的技术那样通过信号处理来抑制输入声音的噪声的情况下,有可能使说话者的个人的特征失真,其结果,说话者辨识的精度有可能降低。此外,用于抑制输入声音的噪声的信号处理计算量变多。

[0024] 为了解决以上的课题,公开下述的技术。

[0025] (1) 本公开的一方式所涉及的说话者识别方法是计算机中的说话者识别方法,所述说话者识别方法获取识别对象声音数据,获取预先登记的多个登记声音数据,算出所述识别对象声音数据与所述多个登记声音数据各自的相似度,选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者,基于所算出的所述多个相似度来判定所述识别对象声音数据是否适于说话者识别,在判定为所述识别对象声音数据适于所述说话者识别的情况下,基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者,输出识别结果。

[0026] 根据该结构,算出识别对象声音数据与多个登记声音数据各自的相似度,基于算出的多个相似度来判定识别对象声音数据是否适于说话者识别。然后,在判定为识别对象声音数据适于说话者识别的情况下,基于最高的相似度来判定是否将所选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0027] 算出多个相似度的处理的计算量少于抑制识别对象声音数据中所包含的噪声的

信号处理的计算量。此外,由于基于算出的多个相似度来判定识别对象声音数据是否适于说话者识别,因此,不针对识别对象声音数据进行对有可能使说话者的个人的特征失真的噪声进行抑制的信号处理。因此,不使计算量增加就能够提升对识别对象的说话者是预先登记的多个说话者中的哪一个进行识别的精度。

[0028] (2)在上述(1)所记载的说话者识别方法中,也可以是,在所述识别对象声音数据是否适于所述说话者识别的判定中,判定所算出的所述多个相似度当中的最高的相似度是否比第1阈值高,在判定为最高的所述相似度比所述第1阈值高的情况下,判定为所述识别对象声音数据适于所述说话者识别。

[0029] 根据该结构,通过将所算出的多个相似度当中的最高的相似度与第1阈值进行比较,能容易地判定识别对象声音数据是否适于说话者识别。

[0030] (3)在上述(1)所记载的说话者识别方法中,也可以是,在所述识别对象声音数据是否适于所述说话者识别的判定中,将所算出的所述多个相似度的方差值算出,判定所算出的所述方差值是否比第1阈值高,在判定为所述方差值比所述第1阈值高的情况下,判定为所述识别对象声音数据适于所述说话者识别。

[0031] 在识别对象声音数据不适于说话者识别的情况下,算出的多个相似度的方差值变低。因此,通过将算出的多个相似度的方差值与第1阈值进行比较,能容易地判定识别对象声音数据是否适于说话者识别。

[0032] (4)在上述(2)或(3)所记载的说话者识别方法中,也可以是,在是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者的判定中,判定所算出的所述多个相似度当中的最高的相似度是否比高于所述第1阈值的第2阈值高,在判定为最高的所述相似度比所述第2阈值高的情况下,将所选出的所述登记说话者识别为所述识别对象声音数据的所述识别对象说话者。

[0033] 根据该结构,通过将算出的多个相似度当中的最高的相似度与高于第1阈值的第2阈值进行比较,能容易地判别所选出的登记说话者是否是识别对象声音数据的识别对象说话者。

[0034] (5)在上述(1)所记载的说话者识别方法中,所述多个登记声音数据包含对识别对象的多个登记说话者所发声的声音预先进行登记而得到的多个第1登记声音数据、和对所述识别对象的所述多个登记说话者以外的多个其他登记说话者所发声的声音预先进行登记而得到的多个第2登记声音数据,在所述相似度的算出中,算出所述识别对象声音数据与所述多个第1登记声音数据各自的第1相似度,并且算出所述识别对象声音数据与所述多个第2登记声音数据各自的第2相似度,在所述登记说话者的选出中,选出与所算出的多个第1相似度当中的最高的第1相似度对应的第1登记声音数据的登记说话者,在所述识别对象声音数据是否适于所述说话者识别的判定中,判定所算出的所述多个第1相似度以及所述多个第2相似度当中的最高的第1相似度或第2相似度是否比第1阈值高,在判定为最高的所述第1相似度或所述第2相似度比所述第1阈值高的情况下,判定为所述识别对象声音数据适于所述说话者识别。

[0035] 在识别对象声音数据能进行说话者识别的情况下,通过增加多个登记声音数据的数量,识别对象声音数据与多个登记声音数据的任一者相似的可能性提高。因此,通过使用根据对识别对象的多个登记说话者所发声的声音预先进行登记所得到的多个第1登记

声音数据而算出的多个第1相似度,还使用根据对识别对象的多个登记说话者以外的多个其他登记说话者所发声的声音预先进行登记所得到的多个第2登记声音数据而算出的多个第2相似度,能可靠地判定识别对象声音数据是否适于说话者识别。

[0036] (6)在上述(5)所记载的说话者识别方法中,也可以是,所述多个第2登记声音数据不包含噪声,仅包含所述其他登记说话者所发声的所述声音。

[0037] 根据该结构,通过使用仅包括不包含噪声的干净的声音的多个第2登记声音数据,能稳定地算出识别对象声音数据与多个第2登记声音数据各自的第2相似度。

[0038] (7)在上述(5)或(6)所记载的说话者识别方法中,也可以是,在是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者的判定中,判定所算出的所述多个第1相似度当中的最高的第1相似度是否比高于所述第1阈值的第2阈值高,在判定为最高的所述第1相似度比所述第2阈值高的情况下,将所选出的所述登记说话者识别为所述识别对象声音数据的所述识别对象说话者。

[0039] 根据该结构,通过将算出的多个第1相似度当中的最高的第1相似度与高于第1阈值的第2阈值进行比较,能容易地识别所选出的登记说话者是否是识别对象声音数据的识别对象说话者。

[0040] (8)在上述(1)~(7)任一者所记载的说话者识别方法中,也可以是,进一步地在判定为所述识别对象声音数据不适于所述说话者识别的情况下,对催促所述识别对象说话者再输入所述识别对象声音数据的错误消息进行输出。

[0041] 根据该结构,在识别对象声音数据不适于说话者识别的情况下,能催促识别对象说话者再输入识别对象声音数据,能使用再输入的识别对象声音数据来进行说话者识别。

[0042] (9)在上述(1)~(7)任一者所记载的说话者识别方法中,也可以是,在所述识别对象声音数据的获取中,获取从所述识别对象说话者所发声的声音数据截取给定的区间而得到的所述识别对象声音数据,进一步地在判定为所述识别对象声音数据不适于所述说话者识别的情况下,获取从所述声音数据截取与所述给定的区间不同的区间而得到的其他识别对象声音数据。

[0043] 例如,在最初截取的区间的识别对象声音数据中不包含识别对象说话者的声音的情况下,判定为识别对象声音数据不适于说话者识别。在该情况下,获取从声音数据截取与最初的区间不同的区间而得到的其他识别对象声音数据。因此,在判定为识别对象声音数据不适于说话者识别的情况下,能使用其他识别对象声音数据来进行说话者识别。

[0044] 此外,本公开不仅能实现为执行以上那样的特征性的处理的说话者识别方法,还能实现为具备与说话者识别方法所执行的特征性的方法对应的特征性的结构的说话者识别装置等。此外,还能实现为使计算机执行这样的说话者识别方法中所包含的特征性的处理的计算机程序。因此,在以下的其他方式中,也能起到与上述的说话者识别方法同样的效果。

[0045] (10)本公开的另一方式所涉及的说话者识别装置具备:识别对象声音数据获取部,其获取识别对象声音数据;登记声音数据获取部,其获取预先登记的多个登记声音数据;算出部,其算出所述识别对象声音数据与所述多个登记声音数据各自的相似度;选出部,其选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者;相似度判定部,其基于所算出的所述多个相似度来判定所述识别对象声音数据是否适

于说话者识别;和说话者判定部,其在判定为所述识别对象声音数据适于所述说话者识别的情况下,基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者;和输出部,其输出识别结果。

[0046] (11) 本公开的另一方式所涉及的说话者识别程序使计算机如下地发挥功能:获取识别对象声音数据,获取预先登记的多个登记声音数据,算出所述识别对象声音数据与所述多个登记声音数据各自的相似度,选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者,基于所算出的所述多个相似度来判定所述识别对象声音数据是否适于说话者识别,在判定为所述识别对象声音数据适于所述说话者识别的情况下,基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者,输出识别结果。

[0047] (12) 本公开的另一方式所涉及的非瞬时性计算机可读的记录介质记录有说话者识别程序,所述说话者识别程序使计算机如下地发挥功能:获取识别对象声音数据,获取预先登记的多个登记声音数据,算出所述识别对象声音数据与所述多个登记声音数据各自的相似度,选出与所算出的多个相似度当中的最高的相似度对应的登记声音数据的登记说话者,基于所算出的所述多个相似度来判定所述识别对象声音数据是否适于说话者识别,在判定为所述识别对象声音数据适于所述说话者识别的情况下,基于最高的所述相似度来判定是否将所选出的所述登记说话者识别为所述识别对象声音数据的识别对象说话者,输出识别结果。

[0048] 以下,参考附图来说明本公开的实施方式。另外,以下的实施方式是将本公开具体化的一例,并不限定本公开的技术的范围。

[0049] (实施方式1)

[0050] 图1是表示本公开的实施方式1中的说话者识别系统的结构的图。

[0051] 图1所示的说话者识别系统具备麦克风1以及说话者识别装置2。另外,说话者识别装置2既可以不具备麦克风1,也可以具备麦克风1。

[0052] 麦克风1对说话者所发声的声音进行拾音,变换成声音数据,并向说话者识别装置2输出。麦克风1在识别说话者时,将说话者所发声的识别对象声音数据向说话者识别装置2输出。此外,麦克风1可以在预先登记声音数据时,将说话者所发声的登记对象声音数据向说话者识别装置2输出。麦克风1可以固定于有识别对象说话者的空间内,也可以能移动。

[0053] 说话者识别装置2具备识别对象声音数据获取部21、第1特征量算出部22、登记声音数据存储部23、登记声音数据获取部24、第2特征量算出部25、相似度得分算出部26、说话者选出部27、相似度得分判定部28、说话者判定部29、识别结果输出部30以及错误处理部31。

[0054] 另外,识别对象声音数据获取部21、第1特征量算出部22、登记声音数据获取部24、第2特征量算出部25、相似度得分算出部26、说话者选出部27、相似度得分判定部28、说话者判定部29、识别结果输出部30以及错误处理部31通过处理器来实现。处理器例如由中央运算处理装置(CPU)等构成。

[0055] 登记声音数据存储部23通过存储器来实现。存储器例如由ROM(Read Only Memory,只读存储器)或EEPROM(Electrically Erasable Programmable Read Only Memory,电可擦除可编程只读存储器)等构成。

[0056] 另外,说话者识别装置2例如可以是计算机、智能手机、平板型计算机或服务器。

[0057] 识别对象声音数据获取部21获取从麦克风1输出的识别对象声音数据。

[0058] 另外,在说话者识别装置2是服务器的情况下,麦克风1可以被组装入识别对象说话者所使用的智能手机等终端。在该情况下,终端可以将识别对象声音数据发送到说话者识别装置2。登记声音数据获取部24例如可以是通信部,可以接收由终端发送的识别对象声音数据。

[0059] 第1特征量算出部22算出由识别对象声音数据获取部21获取的识别对象声音数据的特征量。特征量例如是i-vector。i-vector是通过对GMM(Gaussian Mixture Model,高斯混合模型)超矢量使用因子分析而根据声音数据算出的低维度矢量的特征量。另外,关于i-vector的算出方法,由于是现有技术,因此省略详细的说明。此外,特征量并不限定于i-vector,例如也可以是x-vector等其他特征量。

[0060] 登记声音数据存储部23预先存储将与说话者相关的信息建立对应的多个登记声音数据。与说话者相关的信息例如是用于识别说话者的说话者ID或说话者的姓名。

[0061] 另外,说话者识别装置2可以还具备:登记部,其将从麦克风1输出的登记对象声音数据作为登记声音数据而登记于登记声音数据存储部23;和输入接受部,其接受登记声音数据的与说话者相关的信息的输入。然后,登记部将登记声音数据与由输入接受部接受到的与说话者相关的信息建立对应地登记到登记声音数据存储部23。

[0062] 此外,识别对象声音数据以及登记声音数据的发声内容可以是任何内容。此外,识别对象声音数据以及登记声音数据可以是特定的单词或语句。

[0063] 登记声音数据获取部24获取预先登记于登记声音数据存储部23的多个登记声音数据。登记声音数据获取部24将预先登记的多个登记声音数据从登记声音数据存储部23读出。

[0064] 第2特征量算出部25算出由登记声音数据获取部24获取的多个登记声音数据的特征量。特征量例如是i-vector。

[0065] 相似度得分算出部26算出识别对象声音数据的特征量与多个登记声音数据的特征量各自的相似度得分。相似度得分是对识别对象声音数据的特征量和登记声音数据的特征量以何种程度相似进行数值化后的产物。相似度得分表示识别对象声音数据的特征量与登记声音数据的特征量的相似度。

[0066] 相似度得分算出部26使用概率性线性判别分析(Probabilistic Linear Discriminant Analysis:PLDA)来算出相似度得分。关于相似度得分,视作从概率性的模型来生成发声的特征量,用对数似然比来表示2个发声是否是从相同生成模型(相同说话者)生成的。相似度得分基于下述的式来算出。

[0067] 相似度得分= $\log$ (是相同说话者发声的似然/是不同说话者发声的似然)

[0068] 相似度得分算出部26从400维度的i-vector特征量中自动选择对说话者的识别有效的特征量,算出对数似然比,作为相似度得分。在识别对象声音数据的说话者和登记声音数据的说话者相同的情况下算出的相似度得分高于在识别对象声音数据的说话者和登记声音数据的说话者不同的情况下算出的相似度得分。此外,根据包含大于给定的音量的噪声的不适于说话者识别的识别对象声音数据而算出的相似度得分低于根据包含小于给定的音量的噪声的适于说话者识别的识别对象声音数据而算出的相似度得分。

[0069] 另外,由于利用了PLDA的相似度得分的算出是已知的,因此,省略详细的说明。此外,在本实施方式1中,相似度得分算出部26可以算出识别对象声音数据与多个登记声音数据各自的相似度得分。

[0070] 说话者选出部27算出与由相似度得分算出部26算出的多个相似度得分当中的最高的相似度得分对应的登记声音数据的登记说话者。

[0071] 相似度得分判定部28基于由相似度得分算出部26算出的多个相似度得分来判定识别对象声音数据是否适于说话者识别。在此,相似度得分判定部28判定由相似度得分算出部26算出的多个相似度得分当中的最高的相似度得分是否比第1阈值高。相似度得分判定部28在判定为最高的相似度得分比第1阈值高的情况下,判定为识别对象声音数据适于说话者识别。另一方面,相似度得分判定部28在判定为最高的相似度得分为第1阈值以下的情况下,判定为识别对象声音数据不适于说话者识别。

[0072] 在由相似度得分判定部28判定为识别对象声音数据适于说话者识别的情况下,说话者判定部29基于最高的相似度得分来判定是否将由说话者选出部27选出的登记说话者识别为识别对象声音数据的识别对象说话者。在此,说话者判定部29判定由相似度得分算出部26算出的多个相似度得分当中的最高的相似度得分是否比高于第1阈值的第2阈值高。说话者判定部29在判定为最高的相似度得分比第2阈值高的情况下,判定为将由说话者选出部27选出的登记说话者识别为识别对象声音数据的识别对象说话者。另一方面,说话者判定部29在判定为最高的相似度得分为第2阈值以下的情况下,判定为不将由说话者选出部27选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0073] 另外,在本实施方式1中,说话者判定部29也可以在由相似度得分判定部28判定为识别对象声音数据适于说话者识别的情况下,将由说话者选出部27选出的登记说话者识别为识别对象声音数据的识别对象说话者。在该情况下,说话者判定部29也可以不对由相似度得分算出部26算出的多个相似度得分当中的最高的相似度得分是否比第2阈值高进行判定,而将由说话者选出部27选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0074] 识别结果输出部30输出说话者判定部29的识别结果。在将所选出的登记说话者识别为识别对象声音数据的识别对象说话者的情况下,识别结果输出部30输出包含所选出的登记说话者的姓名或说话者ID的识别结果。此外,识别结果也可以包含相似度得分。此外,识别结果输出部30在不将所选出的登记说话者识别为识别对象声音数据的识别对象说话者的情况下,输出表示未将识别对象声音数据的识别对象说话者识别为预先登记的多个登记说话者中的任何一者的识别结果。

[0075] 识别结果输出部30例如是显示器或扬声器,在将所选出的登记说话者识别为识别对象声音数据的识别对象说话者的情况下,从显示器或扬声器输出表示识别对象声音数据的识别对象说话者是所选出的登记说话者的消息。另一方面,在未将所选出的登记说话者识别为识别对象声音数据的识别对象说话者的情况下,识别结果输出部30从显示器或扬声器输出表示识别对象声音数据的识别对象说话者不是预先登记的多个登记说话者中的任何一者的消息。

[0076] 另外,识别结果输出部30也可以将说话者判定部29的识别结果向说话者识别装置2以外的其他装置输出。在说话者识别装置2是服务器的情况下,识别结果输出部30例如也

可以包含通信部,也可以向识别对象说话者所使用的智能手机等终端发送识别结果。终端可以具备显示器或扬声器。终端的显示器或扬声器可以输出接收到的识别结果。

[0077] 在由相似度得分判定部28判定为识别对象声音数据不适于说话者识别的情况下,错误处理部31对催促识别对象说话者再输入识别对象声音数据的错误消息进行输出。错误处理部31例如输出“请靠近麦克风,或者在安静的场所进行发声。”这样的错误消息。

[0078] 错误处理部31例如是显示器或扬声器,在由相似度得分判定部28判定为识别对象声音数据不适于说话者识别的情况下,从显示器或扬声器输出催促识别对象说话者再输入识别对象声音数据的错误消息。

[0079] 另外,错误处理部31也可以将催促识别对象说话者再输入识别对象声音数据的错误消息向说话者识别装置2以外的其他装置输出。在说话者识别装置2是服务器的情况下,错误处理部31例如也可以包含通信部,也可以向识别对象说话者所使用的智能手机等终端发送错误消息。终端可以具备显示器或扬声器。终端的显示器或扬声器也可以输出接收到的错误消息。

[0080] 接着,说明本公开的实施方式1中的说话者识别装置2的说话者识别处理的动作。

[0081] 图2是用于说明本实施方式1中的说话者识别装置2的说话者识别处理的动作的第1流程图,图3是用于说明本实施方式1中的说话者识别装置2的说话者识别处理的动作的第2流程图。

[0082] 首先,在步骤S1,识别对象声音数据获取部21获取从麦克风1输出的识别对象声音数据。识别对象说话者朝向麦克风1发声。麦克风1收集由识别对象说话者发声的声音,输出识别对象声音数据。

[0083] 接下来,在步骤S2,第1特征量算出部22算出由识别对象声音数据获取部21获取的识别对象声音数据的特征量。

[0084] 接下来,在步骤S3,登记声音数据获取部24从登记声音数据存储部23获取登记声音数据。这时,登记声音数据获取部24从登记于登记声音数据存储部23的多个登记声音数据中获取1个登记声音数据。

[0085] 接下来,在步骤S4,第2特征量算出部25算出由登记声音数据获取部24获取的登记声音数据的特征量。

[0086] 接下来,在步骤S5,相似度得分算出部26算出识别对象声音数据的特征量与登记声音数据的特征量的相似度得分。

[0087] 接下来,在步骤S6,相似度得分算出部26判定是否算出了识别对象声音数据的特征量与存储于登记声音数据存储部23的全部登记声音数据的特征量的相似度得分。在此,在判定为未算出识别对象声音数据的特征量与全部登记声音数据的特征量的相似度得分的情况下(步骤S6中的“否”),处理回到步骤S3。然后,登记声音数据获取部24从存储于登记声音数据存储部23的多个登记声音数据中获取未算出相似度得分的登记声音数据。

[0088] 另一方面,在判定为算出了识别对象声音数据的特征量与全部登记声音数据的特征量的相似度得分的情况下(步骤S6中的“是”),在步骤S7,说话者选出部27选出与由相似度得分算出部26算出的多个相似度得分当中的最高的相似度得分对应的登记声音数据的登记说话者。

[0089] 接下来,在步骤S8,相似度得分判定部28判定最高的相似度得分是否比第1阈值

高。

[0090] 在此,在判定为最高的相似度得分为第1阈值以下的情况下(步骤S8中的“否”),在步骤S9,错误处理部31输出催促识别对象说话者再输入识别对象声音数据的错误消息。

[0091] 另一方面,在判定为最高的相似度得分比第1阈值高的情况下(步骤S8中的“是”),在步骤S10,说话者判定部29判定由相似度得分算出部26算出的多个相似度得分当中的最高的相似度得分是否比高于第1阈值的第2阈值高。

[0092] 在此,在判定为最高的相似度得分比第2阈值高的情况下(步骤S10中的“是”),在步骤S11,说话者判定部29将由说话者选出部27选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0093] 另一方面,在判定为最高的相似度得分为第2阈值以下的情况下(步骤S10中的“否”),在步骤S12,说话者判定部29判定为由说话者选出部27选出的登记说话者不是识别对象声音数据的识别对象说话者。

[0094] 接下来,在步骤S13,识别结果输出部30输出说话者判定部29的识别结果。在将所选出的登记说话者识别为识别对象声音数据的识别对象说话者的情况下,识别结果输出部30输出表示识别对象声音数据的识别对象说话者是所选出的登记说话者的消息。另一方面,在判定为所选出的登记说话者不是识别对象声音数据的识别对象说话者的情况下,识别结果输出部30输出表示识别对象声音数据的识别对象说话者不是预先登记的多个登记说话者中的任何一者的消息。

[0095] 如此地,算出识别对象声音数据与多个登记声音数据各自的相似度得分,基于算出的多个相似度得分来判定识别对象声音数据是否适于说话者识别。然后,在判定为识别对象声音数据适于说话者识别的情况下,基于最高的相似度得分来判定是否将所选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0096] 算出多个相似度得分的处理的计算量少于抑制识别对象声音数据中包含的噪声的信号处理的计算量。此外,由于基于算出的多个相似度得分来判定识别对象声音数据是否适于说话者识别,因此,不对识别对象声音数据进行对有可能使说话者的个人的特征失真的噪声进行抑制的信号处理。因此,不使计算量增加就能提升对识别对象的说话者是预先登记的多个说话者中的哪一个进行识别的精度。

[0097] 另外,在本实施方式1中,错误处理部31输出催促识别对象说话者再输入识别对象声音数据的错误消息,但本公开并不特别限于此。识别对象声音数据获取部21也可以获取从识别对象说话者所发声的声音数据中截取给定的区间而得到的识别对象声音数据。这时,在截取给定的区间而得到的识别对象声音数据中有可能不包含识别对象说话者的声音。在该情况下,由相似度得分判定部28判定为识别对象声音数据不适于说话者识别。因此,在由相似度得分判定部28判定为识别对象声音数据不适于说话者识别的情况下,错误处理部31也可以获取从声音数据截取与给定的区间不同的区间而得到的其他识别对象声音数据。然后,处理回到步骤S2,第1特征量算出部22也可以算出由错误处理部31获取的其他识别对象声音数据的特征量。之后,也可以进行步骤S3以后的处理。

[0098] 如此地,例如,在最初截取的区间的识别对象声音数据中不包含识别对象说话者的声音的情况下,判定为识别对象声音数据不适于说话者识别。在该情况下,获取从声音数据截取与最初的区间不同的区间而得到的其他识别对象声音数据。因此,在判定为识别对

象声音数据不适于说话者识别的情况下,能使用其他识别对象声音数据来进行说话者识别。

[0099] (实施方式2)

[0100] 在上述的实施方式1中,对算出的多个相似度得分当中的最高的相似度得分是否比第1阈值高进行判定,在判定为最高的相似度得分比第1阈值高的情况下,判定为识别对象声音数据适于说话者识别。与此相对,在实施方式2中,将算出的多个相似度得分的方差值算出,判定算出的方差值是否比第1阈值高,在判定为方差值比第1阈值高的情况下,判定为识别对象声音数据适于说话者识别。

[0101] 图4是表示本公开的实施方式2中的说话者识别系统的结构的图。

[0102] 图4所示的说话者识别系统具备麦克风1以及说话者识别装置2A。另外,说话者识别装置2A可以不具备麦克风1,也可以具备麦克风1。

[0103] 另外,在本实施方式2中,对与实施方式1相同的结构标注相同的附图标记,并省略说明。

[0104] 说话者识别装置2A具备识别对象声音数据获取部21、第1特征量算出部22、登记声音数据存储部23、登记声音数据获取部24、第2特征量算出部25、相似度得分算出部26、说话者选出部27、相似度得分判定部28A、说话者判定部29、识别结果输出部30以及错误处理部31。

[0105] 相似度得分判定部28A基于由相似度得分算出部26算出的多个相似度得分来判定识别对象声音数据是否适于说话者识别。在此,相似度得分判定部28A算出由相似度得分算出部26算出的多个相似度得分的方差值。相似度得分判定部28A判定算出的方差值是否比第1阈值高。相似度得分判定部28A在判定为方差值比第1阈值高的情况下,判定为识别对象声音数据适于说话者识别。另一方面,相似度得分判定部28A在判定为方差值为第1阈值以下的情况下,判定为识别对象声音数据不适于说话者识别。

[0106] 在识别对象声音数据中包含噪声并且该识别对象声音数据不适于说话者识别的情况下,识别对象声音数据与多个登记声音数据的相似度得分均成为低的值。因此,若多个相似度得分的方差值低,则能判定为识别对象声音数据不适于说话者识别。

[0107] 接着,说明本公开的实施方式2中的说话者识别装置2A的说话者识别处理的动作。

[0108] 图5是用于说明本实施方式2中的说话者识别装置2A的说话者识别处理的动作的第1流程图,图6是用于说明本实施方式2中的说话者识别装置2A的说话者识别处理的动作的第2流程图。

[0109] 另外,步骤S21~步骤S27的处理由于与图2的步骤S1~步骤S7的处理相同,因此省略说明。

[0110] 接下来,在步骤S28,相似度得分判定部28A算出由相似度得分算出部26算出的多个相似度得分的方差值。

[0111] 接下来,在步骤S29,相似度得分判定部28A判定算出的方差值是否比第1阈值高。

[0112] 在此,在判定为方差值为第1阈值以下的情况下(步骤S29中的“否”),在步骤S30,错误处理部31输出催促识别对象说话者再输入识别对象声音数据的错误消息。

[0113] 另一方面,在判定为方差值比第1阈值高的情况下(步骤S29中的“是”),在步骤S31,说话者判定部29判定由相似度得分算出部26算出的多个相似度得分当中的最高的相

似度得分是否比高于第1阈值的第2阈值高。

[0114] 另外,步骤S31~步骤S34的处理由于与图3的步骤S9~步骤S12的处理相同,因此省略说明。

[0115] 在识别对象声音数据不适于说话者识别的情况下,算出的多个相似度得分的方差值变低。因此,通过将算出的多个相似度得分的方差值与第1阈值进行比较,能容易地判定识别对象声音数据是否适于说话者识别。

[0116] 另外,在实施方式1以及实施方式2中,相似度得分算出部26算出识别对象声音数据的特征量与多个登记声音数据的特征量各自的相似度得分,但本公开并不特别限定于此。相似度得分算出部26也可以算出识别对象声音数据与多个登记声音数据各自的相似度得分。在该情况下,不需要识别对象声音数据的特征量和多个登记声音数据的特征量的算出。

[0117] (实施方式3)

[0118] 在上述的实施方式1中,算出识别对象声音数据与对识别对象的多个登记说话者所发声的声音预先进行登记而得到的多个第1登记声音数据各自的第1相似度得分,基于算出的多个第1相似度得分来判定识别对象声音数据是否适于说话者识别。与此相对,在实施方式3中,进一步算出识别对象声音数据与对识别对象的多个登记说话者以外的多个其他登记说话者所发声的声音预先进行登记而得到的多个第2登记声音数据各自的第2相似度得分,基于算出的多个第1相似度得分以及多个第2相似度得分来判定识别对象声音数据是否适于说话者识别。

[0119] 图7是表示本公开的实施方式3中的说话者识别系统的结构的图。

[0120] 图7所示的说话者识别系统具备麦克风1以及说话者识别装置2B。另外,说话者识别装置2B可以不具备麦克风1,也可以具备麦克风1。

[0121] 另外,在本实施方式3中,对与实施方式1相同结构标注相同的附图标记,并省略说明。

[0122] 说话者识别装置2B具备识别对象声音数据获取部21、第1特征量算出部22、第1登记声音数据存储部23B、第1登记声音数据获取部24B、第2特征量算出部25B、相似度得分算出部26B、说话者选出部27B、相似度得分判定部28B、说话者判定部29B、识别结果输出部30、错误处理部31、第2登记声音数据存储部32、第2登记声音数据获取部33以及第3特征量算出部34。

[0123] 另外,识别对象声音数据获取部21、第1特征量算出部22、第1登记声音数据获取部24B、第2特征量算出部25B、相似度得分算出部26B、说话者选出部27B、相似度得分判定部28B、说话者判定部29B、识别结果输出部30、错误处理部31、第2登记声音数据获取部33以及第3特征量算出部34通过处理器来实现。第1登记声音数据存储部23B以及第2登记声音数据存储部32通过存储器来实现。

[0124] 第1登记声音数据存储部23B预先存储将与说话者相关的信息建立对应的多个第1登记声音数据。多个第1登记声音数据表示识别对象的多个登记说话者所发声的声音。多个第1登记声音数据与实施方式1中的多个登记声音数据相同。

[0125] 第1登记声音数据获取部24B获取预先登记于第1登记声音数据存储部23B的多个第1登记声音数据。

[0126] 第2特征量算出部25B算出由第1登记声音数据获取部24B获取的多个第1登记声音数据的特征量。特征量例如是i-vector。

[0127] 第2登记声音数据存储部32预先存储多个第2登记声音数据。多个第2登记声音数据表示识别对象的多个登记说话者以外的多个其他登记说话者所发声的声音。多个第2登记声音数据不包含噪声,仅包含声音。

[0128] 第2登记声音数据获取部33获取预先登记于第2登记声音数据存储部32的多个第2登记声音数据。

[0129] 第3特征量算出部34算出由第2登记声音数据获取部33获取的多个第2登记声音数据的特征量。特征量例如是i-vector。

[0130] 相似度得分算出部26B算出识别对象声音数据的特征量与多个第1登记声音数据的特征量各自的第1相似度得分,并且算出识别对象声音数据的特征量与多个第2登记声音数据的特征量各自的第2相似度得分。

[0131] 说话者选出部27B选出与由相似度得分算出部26B算出的多个第1相似度得分当中的最高的第1相似度得分对应的第1登记声音数据的登记说话者。

[0132] 相似度得分判定部28B基于由相似度得分算出部26B算出的多个第1相似度得分以及多个第2相似度得分,来判定识别对象声音数据是否适于说话者识别。在此,相似度得分判定部28B判定由相似度得分算出部26B算出的多个第1相似度得分以及多个第2相似度得分当中的最高的第1相似度得分或第2相似度得分是否比第1阈值高。相似度得分判定部28B在判定为最高的第1相似度得分或第2相似度得分比第1阈值高的情况下,判定为识别对象声音数据适于说话者识别。另一方面,相似度得分判定部28B在判定为最高的第1相似度得分或第2相似度得分为第1阈值以下的情况下,判定为识别对象声音数据不适于说话者识别。

[0133] 在识别对象声音数据能进行说话者识别的情况下,该识别对象声音数据与许多登记声音数据当中任一个登记声音数据相似的可能性较高。因此,本实施方式3中的第2登记声音数据存储部32预先存储不包含噪声且包含识别对象的多个登记说话者以外的多个其他登记说话者所发声的干净的声音的多个第2登记声音数据。多个其他登记说话者的人数例如为100人,多个第2登记声音数据的数量例如为100个。若在多个第2登记声音数据中有与识别对象声音数据相似的第2登记声音数据,则能判定为识别对象声音数据能进行说话者识别。

[0134] 在由相似度得分判定部28B判定为识别对象声音数据适于说话者识别的情况下,说话者判定部29B基于最高的第1相似度得分来判定是否将由说话者选出部27B选出的登记说话者识别为识别对象声音数据的识别对象说话者。在此,说话者判定部29B判定由相似度得分算出部26B算出的多个第1相似度得分当中的最高的第1相似度得分是否比高于第1阈值的第2阈值高。说话者判定部29B在判定为最高的第1相似度得分比第2阈值高的情况下,判定为将由说话者选出部27B选出的登记说话者识别为识别对象声音数据的识别对象说话者。另一方面,说话者判定部29B在判定为最高的第1相似度得分为第2阈值以下的情况下,判定为不将由说话者选出部27B选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0135] 另外,在本实施方式3中,说话者判定部29B也可以在由相似度得分判定部28B判定

为识别对象声音数据适于说话者识别的情况下,将由说话者选出部27B选出的登记说话者识别为识别对象声音数据的识别对象说话者。在该情况下,说话者判定部29也可以不对由相似度得分算出部26B算出的多个第1相似度得分当中的最高的第1相似度得分是否比第2阈值高进行判定,而将由说话者选出部27B选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0136] 接着,说明本公开的实施方式3中的说话者识别装置2B的说话者识别处理的动作。

[0137] 图8是用于说明本实施方式3中的说话者识别装置2B的说话者识别处理的动作的第1流程图,图9是用于说明本实施方式3中的说话者识别装置2B的说话者识别处理的动作的第2流程图。

[0138] 另外,步骤S41以及步骤S42的处理由于与图2的步骤S1以及步骤S2的处理相同,因此省略说明。

[0139] 接下来,在步骤S43,第1登记声音数据获取部24B从第1登记声音数据存储部23B获取第1登记声音数据。这时,第1登记声音数据获取部24B从登记于第1登记声音数据存储部23B的多个第1登记声音数据中获取1个第1登记声音数据。

[0140] 接下来,在步骤S44,第2特征量算出部25B算出由第1登记声音数据获取部24B获取的第1登记声音数据的特征量。

[0141] 接下来,在步骤S45,相似度得分算出部26B算出识别对象声音数据的特征量与第1登记声音数据的特征量的第1相似度得分。

[0142] 接下来,在步骤S46,相似度得分算出部26B判定是否算出了识别对象声音数据的特征量与存储于第1登记声音数据存储部23B的全部第1登记声音数据的特征量的第1相似度得分。在此,在判定为未算出识别对象声音数据的特征量与全部第1登记声音数据的特征量的第1相似度得分的情况下(步骤S46中的“否”),处理回到步骤S43。然后,第1登记声音数据获取部24B从存储于第1登记声音数据存储部23B的多个第1登记声音数据中获取未算出第1相似度得分的第1登记声音数据。

[0143] 另一方面,在判定为算出了识别对象声音数据的特征量与全部第1登记声音数据的特征量的第1相似度得分的情况下(步骤S46中的“是”),在步骤S47,第2登记声音数据获取部33从第2登记声音数据存储部32获取第2登记声音数据。这时,第2登记声音数据获取部33从登记于第2登记声音数据存储部32的多个第2登记声音数据中获取1个第2登记声音数据。

[0144] 接下来,在步骤S48,第3特征量算出部34算出由第2登记声音数据获取部33获取的第2登记声音数据的特征量。

[0145] 接下来,在步骤S49,相似度得分算出部26B算出识别对象声音数据的特征量与第2登记声音数据的特征量的第2相似度得分。

[0146] 接下来,在步骤S50,相似度得分算出部26B判定是否算出了识别对象声音数据的特征量与存储于第2登记声音数据存储部32的全部第2登记声音数据的特征量的第2相似度得分。在此,在判定为未算出识别对象声音数据的特征量与全部第2登记声音数据的特征量的第2相似度得分的情况下(步骤S50中的“否”),处理回到步骤S47。然后,第2登记声音数据获取部33从存储于第2登记声音数据存储部32的多个第2登记声音数据中获取未算出第2相似度得分的第2登记声音数据。

[0147] 另一方面,在判定为算出了识别对象声音数据的特征量与全部第2登记声音数据的特征量的第2相似度得分的情况下(步骤S50中的“是”),在步骤S51,说话者选出部27B选出与由相似度得分算出部26B算出的多个第1相似度得分当中的最高的第1相似度得分对应的第1登记声音数据的登记说话者。

[0148] 接下来,在步骤S52,相似度得分判定部28B判定最高的第1相似度得分或第2相似度得分是否比第1阈值高。

[0149] 在此,在判定为最高的第1相似度得分或第2相似度得分为第1阈值以下的情况下(步骤S52中的“否”),在步骤S53,错误处理部31输出催促识别对象说话者再输入识别对象声音数据的错误消息。

[0150] 另一方面,在判定为最高的第1相似度得分或第2相似度得分比第1阈值高的情况下(步骤S52中的“是”),在步骤S54,说话者判定部29B判定由相似度得分算出部26B算出的多个第1相似度得分当中的最高的第1相似度得分是否比高于第1阈值的第2阈值高。

[0151] 在此,在判定为最高的第1相似度得分比第2阈值高的情况下(步骤S54中的“是”),在步骤S55,说话者判定部29B将由说话者选出部27B选出的登记说话者识别为识别对象声音数据的识别对象说话者。

[0152] 另一方面,在判定为最高的第1相似度得分为第2阈值以下的情况下(步骤S54中的“否”),在步骤S56,说话者判定部29B判定为由说话者选出部27B选出的登记说话者不是识别对象声音数据的识别对象说话者。

[0153] 另外,步骤S57的处理与图3的步骤S12的处理相同,因此省略说明。

[0154] 在识别对象声音数据能进行说话者识别的情况下,通过增加多个登记声音数据的数量,从而识别对象声音数据与多个登记声音数据的任一者相似的可能性变高。因此,通过不仅使用根据对识别对象的多个登记说话者所发声的声音预先进行登记所得到的多个第1登记声音数据而算出的多个第1相似度得分,还使用根据对识别对象的多个登记说话者以外的多个其他登记说话者所发声的声音预先进行登记所得到的多个第2登记声音数据而算出的多个第2相似度得分,能可靠地判定识别对象声音数据是否适于说话者识别。

[0155] 另外,在本实施方式3中,相似度得分判定部28B基于由相似度得分算出部26B算出的多个第1相似度得分以及多个第2相似度得分来判定识别对象声音数据是否适于说话者识别,但本公开并不特别限定于此。相似度得分判定部28B也可以基于由相似度得分算出部26B算出的多个第2相似度得分来判定识别对象声音数据是否适于说话者识别。这时,相似度得分判定部28B也可以判定由相似度得分算出部26B算出的多个第2相似度得分当中的最高的第2相似度得分是否比第1阈值高。相似度得分判定部28B也可以在判定为最高的第2相似度得分比第1阈值高的情况下,判定为识别对象声音数据适于说话者识别。另一方面,相似度得分判定部28B也可以在最高的第2相似度得分为第1阈值以下的情况下,判定为识别对象声音数据不适于说话者识别。

[0156] 另外,在实施方式3中,相似度得分算出部26B算出识别对象声音数据的特征量与多个第1登记声音数据的特征量各自的第1相似度得分,并且算出识别对象声音数据的特征量与多个第2登记声音数据的特征量各自的第2相似度得分,但本公开并不特别限定于此。相似度得分算出部26B也可以算出识别对象声音数据与多个第1登记声音数据各自的第1相似度得分,并且算出识别对象声音数据与多个第2登记声音数据各自的第2相似度得分。在

该情况下,不需要识别对象声音数据的特征量、多个第1登记声音数据的特征量和多个第2登记声音数据的特征量的算出。

[0157] 另外,在上述各实施方式中,各构成要素可以由专用的硬件构成,或者通过执行适于各构成要素的软件程序来实现。各构成要素可以通过由CPU或处理器等程序执行部将记录于硬盘或半导体存储器等记录介质的软件程序读出并执行来实现。此外,可以通过将程序记录于记录介质而进行移送,或者将程序经由网络而移送,从而由独立的其他计算机系统实施程序。

[0158] 本公开的实施方式所涉及的装置的功能的一部分或全部典型地实现为作为集成电路的LSI (Large Scale Integration,大规模集成电路)。这些可以独立地单芯片化,也可以包含一部分或全部地单芯片化。此外,集成电路化并不限于LSI,也可以通过专用电路或通用处理器来实现。也可以利用能在LSI制造后进行编程的FPGA (Field Programmable Gate Array,现场可编程门阵列)、能对LSI内部的电路单元的连接、设定进行重构的可重构处理器。

[0159] 此外,也可以通过CPU等的处理器执行程序来实现本公开的实施方式所涉及的装置的功能的一部分或全部。

[0160] 此外,上述中使用的数字全都是为了具体说明本公开而例示出的,本公开并不受所例示的数字的限制。

[0161] 此外,对上述流程图中所示的各步骤进行执行的顺序是为了具体说明本公开而例示出的,在能得到同样的效果的范围内,也可以是上述以外的顺序。此外,上述步骤的一部分可以与其他步骤同时(并行)执行。

[0162] 产业上的可用性

[0163] 本公开所涉及的技术由于不使计算量增加就能够提升对识别对象的说话者是预先登记的多个说话者中的哪一个进行识别的精度,因此,作为识别说话者的技术是有用的。

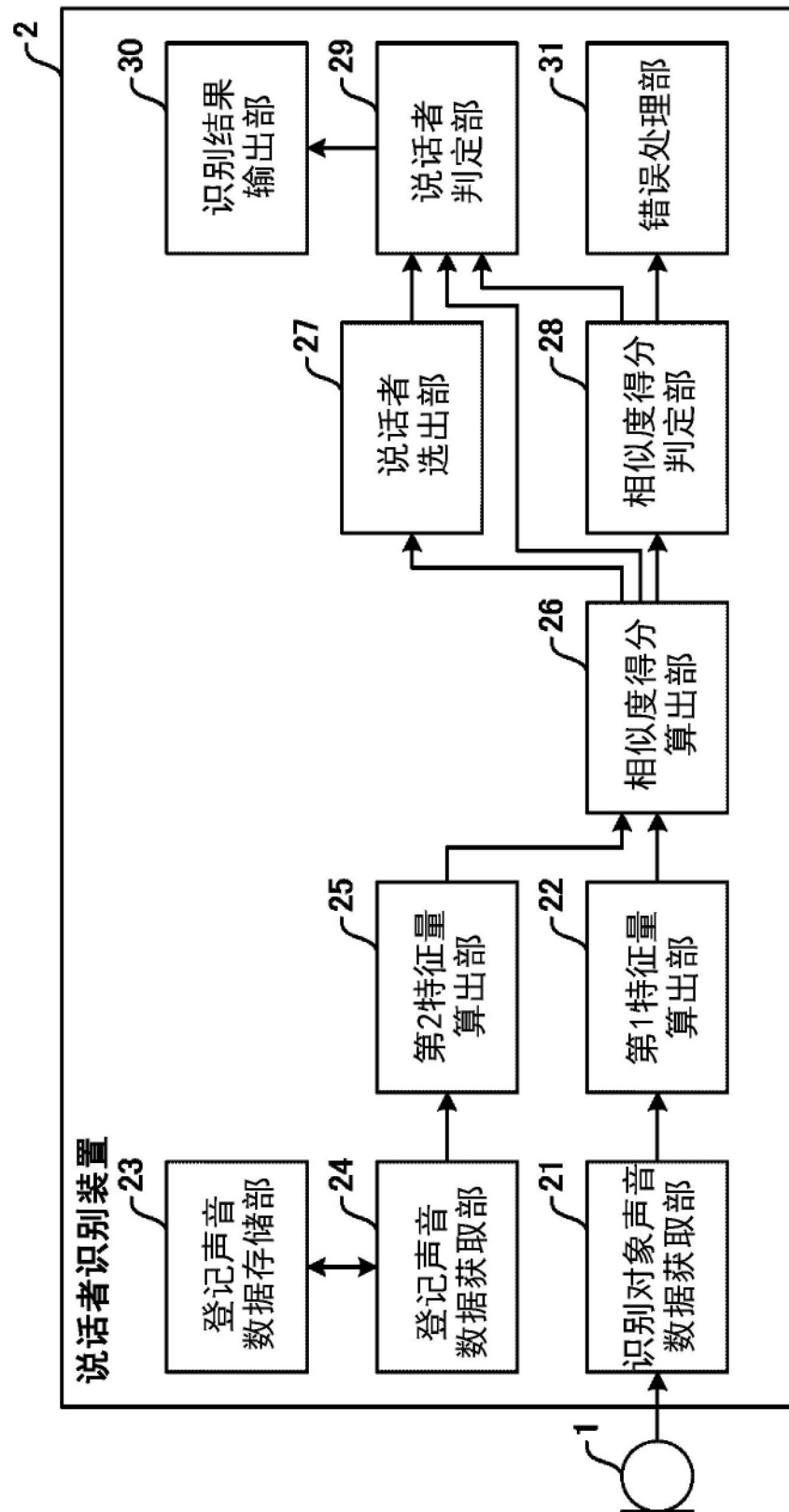


图1

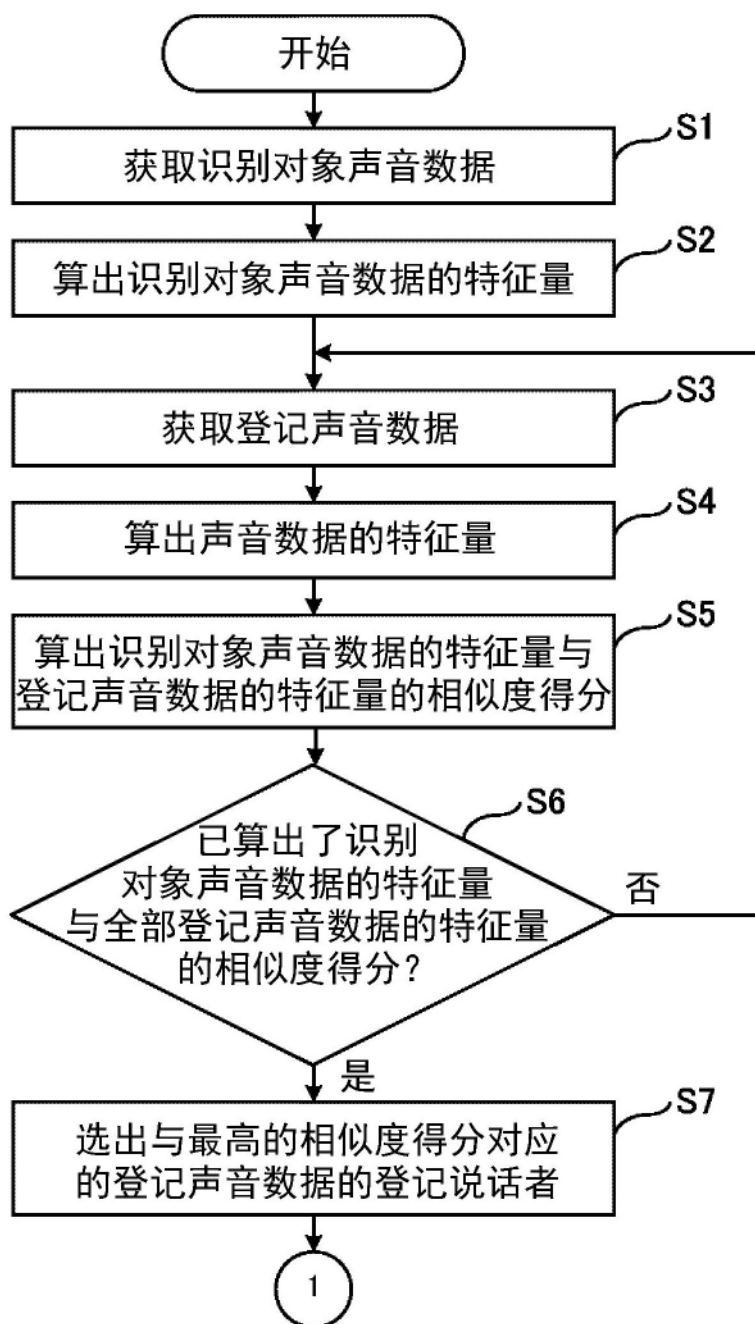


图2

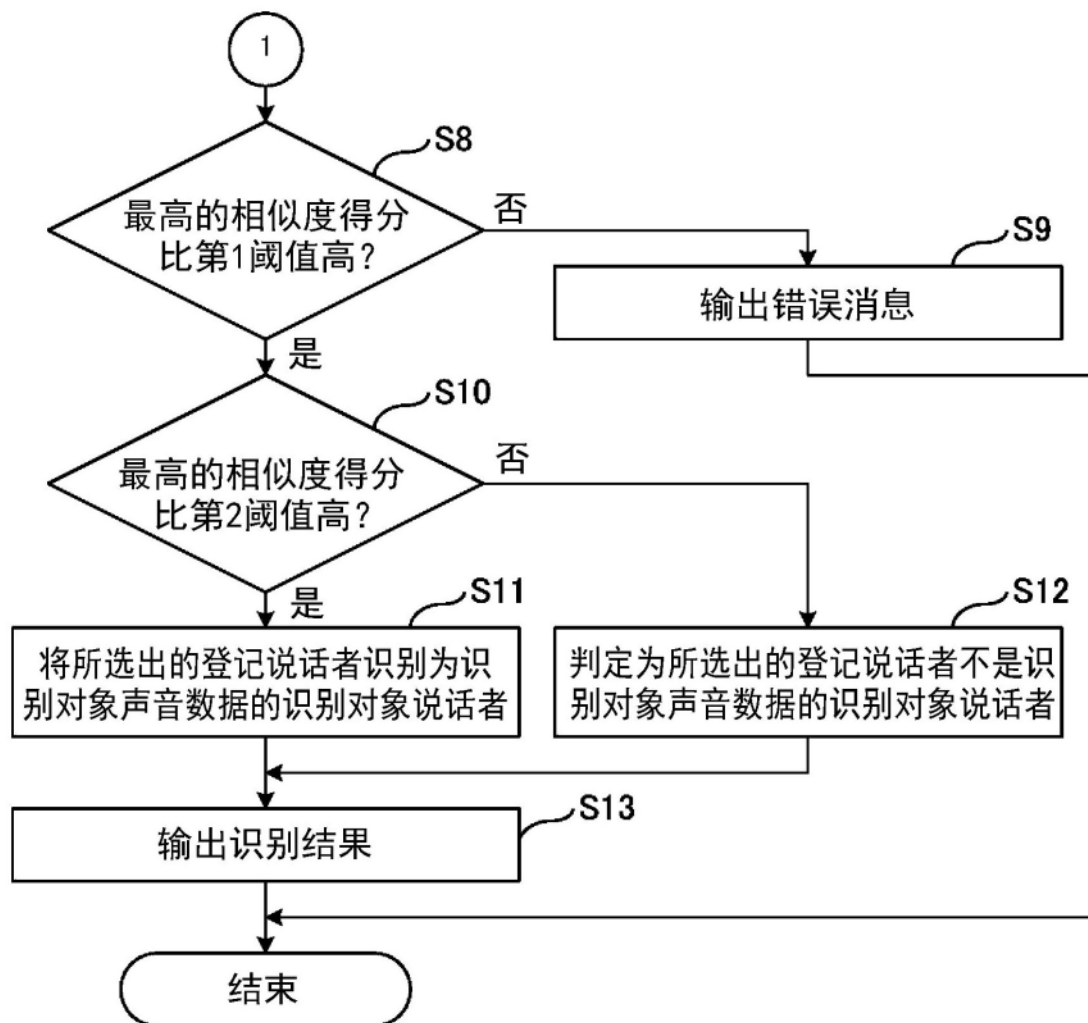


图3

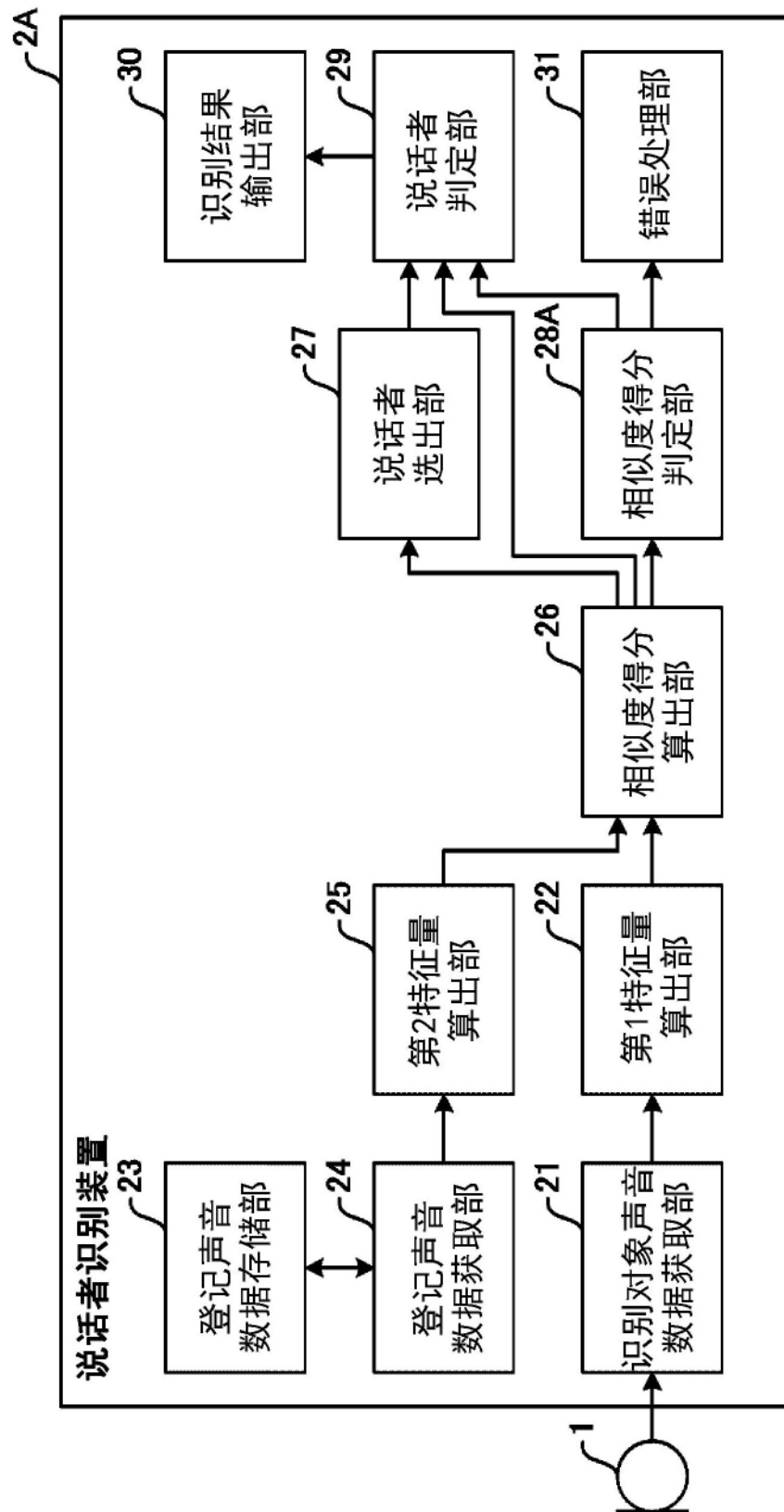


图4

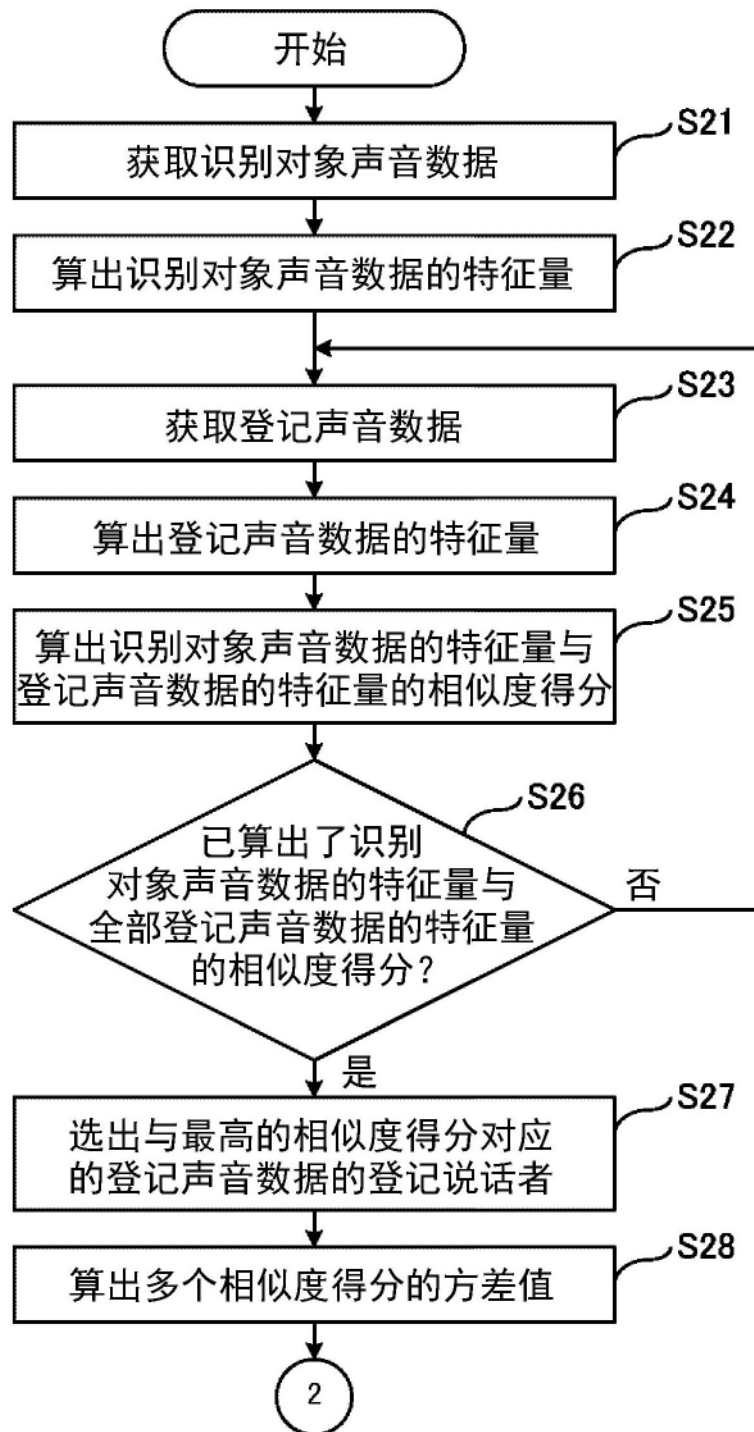


图5

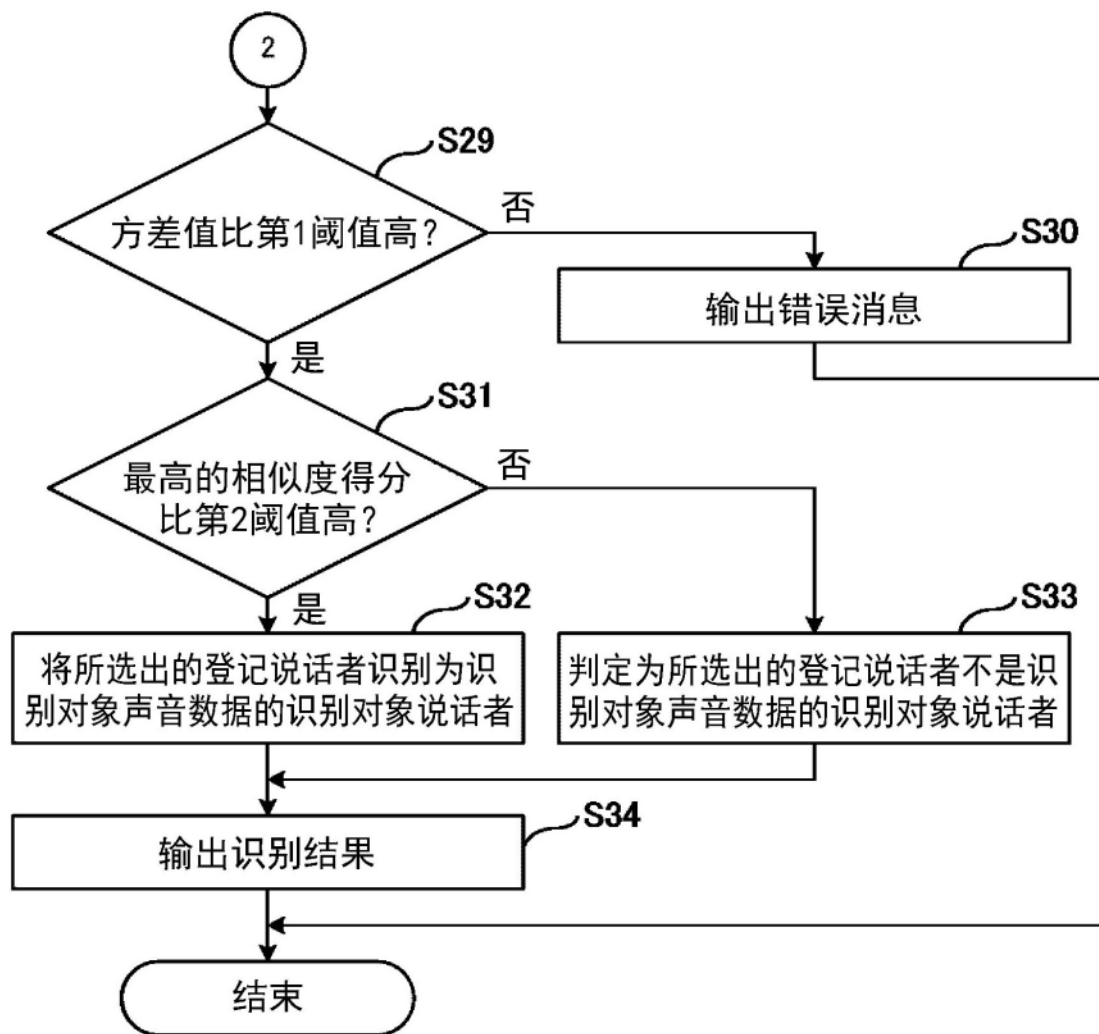


图6

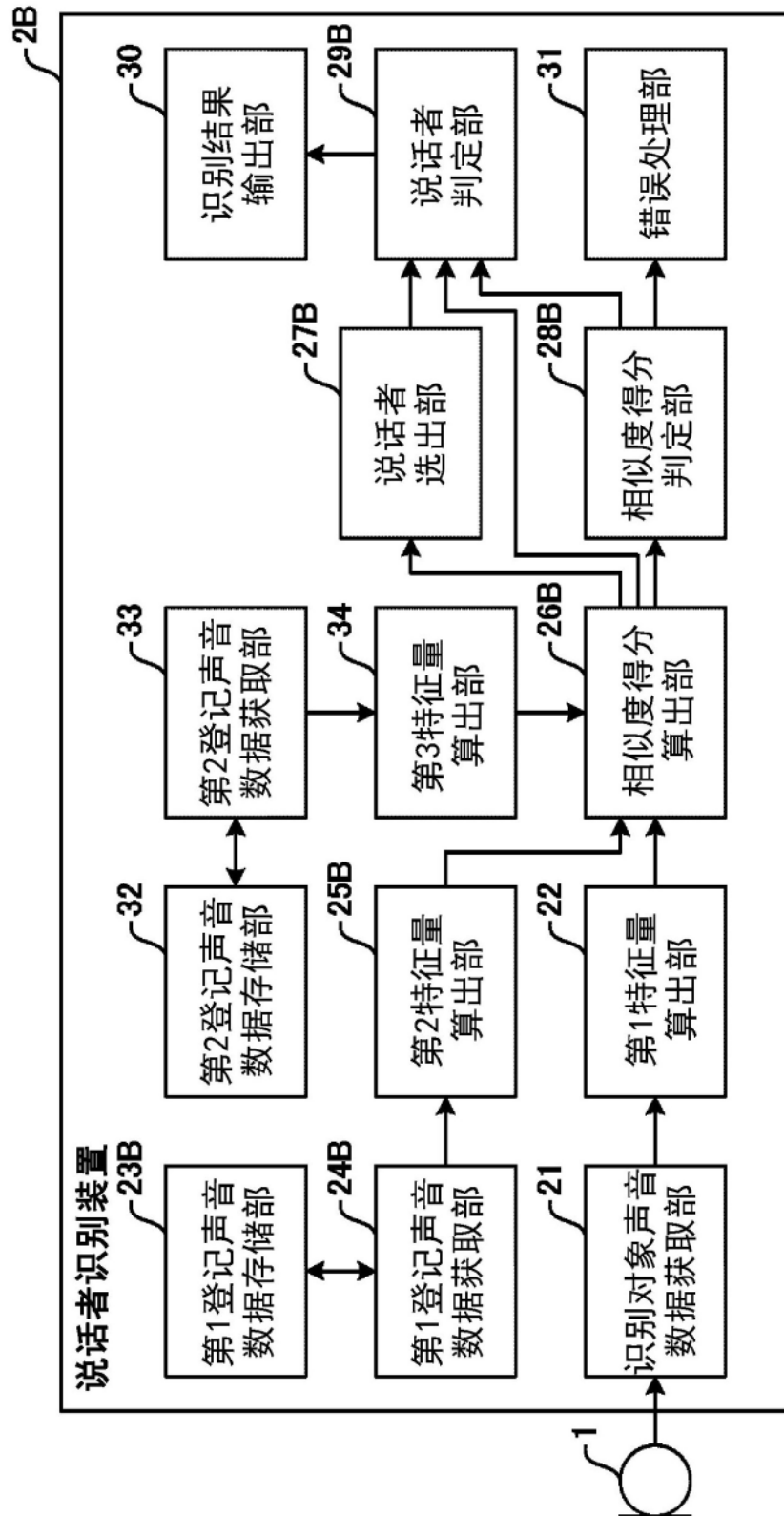


图7

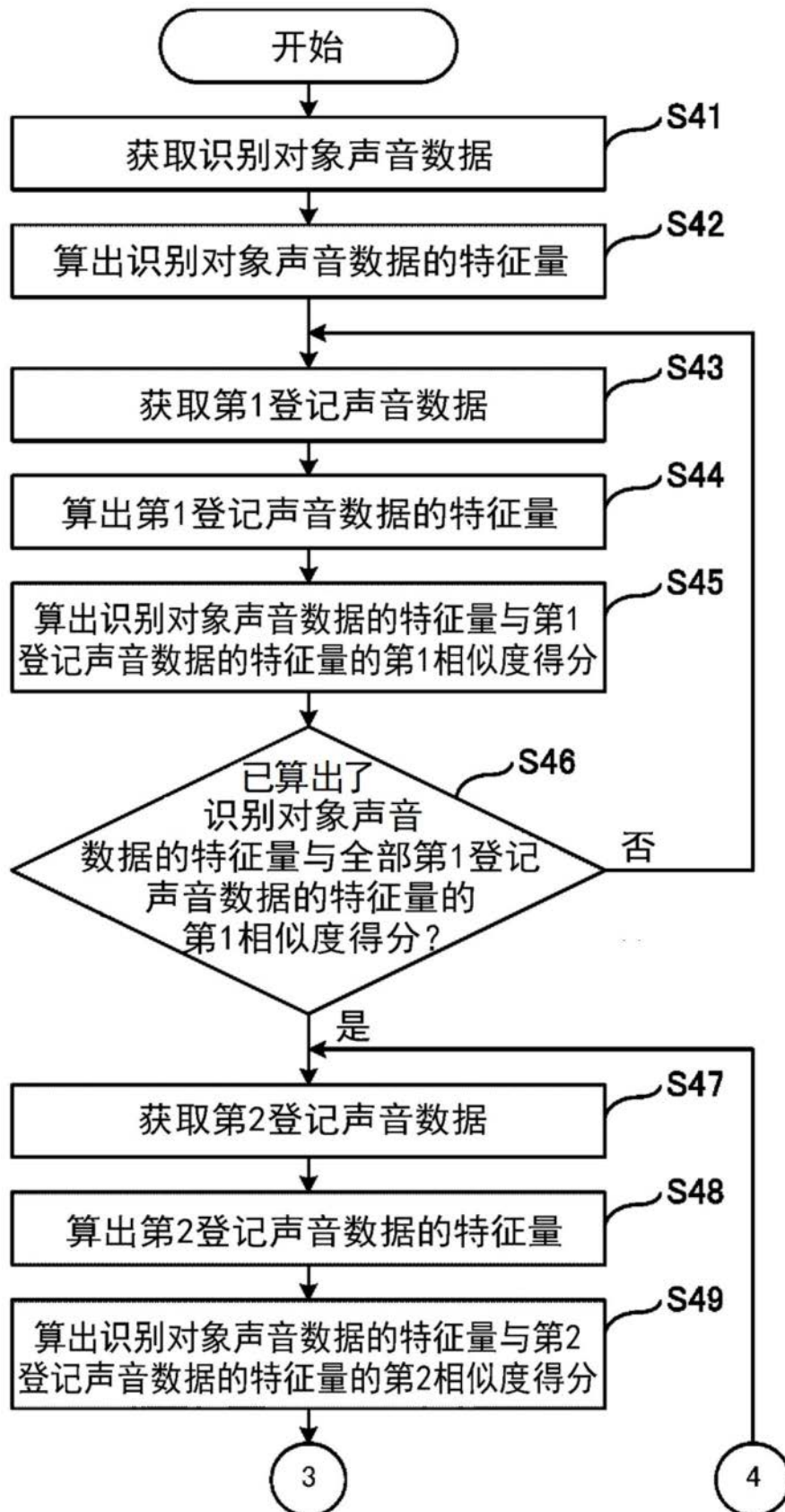


图8

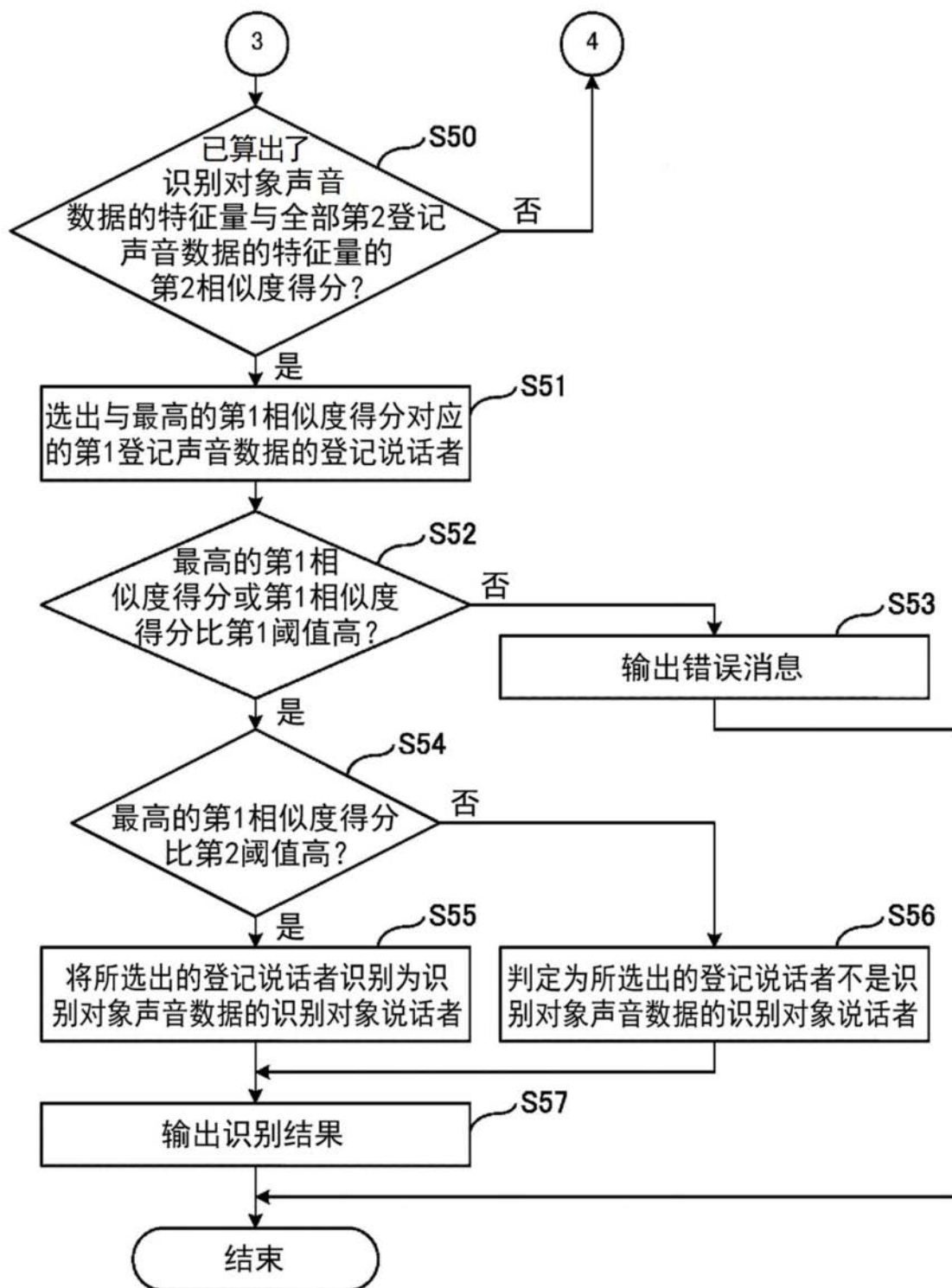


图9