



(12) 发明专利申请

(10) 申请公布号 CN 104282303 A

(43) 申请公布日 2015. 01. 14

(21) 申请号 201310285509. 0

(22) 申请日 2013. 07. 09

(71) 申请人 威盛电子股份有限公司
地址 中国台湾新北市

(72) 发明人 张国峰

(74) 专利代理机构 北京市柳沈律师事务所
11105

代理人 钱大勇

(51) Int. Cl.

G10L 15/26 (2006. 01)

G10L 15/20 (2006. 01)

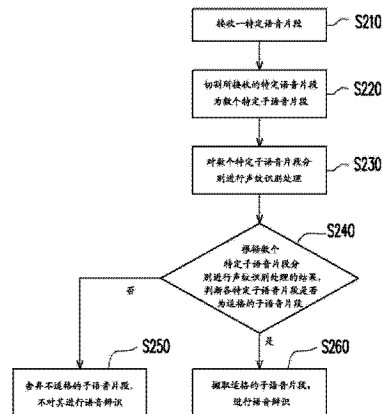
权利要求书2页 说明书4页 附图5页

(54) 发明名称

利用声纹识别进行语音辨识的方法及其电子装置

(57) 摘要

一种利用声纹识别进行语音辨识的方法及其电子装置,此方法包括下列步骤:接收一特定语音片段,切割所接收的特定语音片段为数个特定子语音片段,并对上述数个特定子语音片段分别进行一声纹识别处理。根据声纹识别处理的结果,判断各特定子语音片段是否为适格的子语音片段,撷取其中被判断为适格的子语音片段,进行语音辨识。



1. 一种利用声纹识别进行语音辨识的方法,包括:
接收一特定语音片段;
切割所接收的该特定语音片段为数个特定子语音片段;
对该数个特定子语音片段分别进行一声纹识别处理;
根据对该数个特定子语音片段分别进行该声纹识别处理的结果,判断各该特定子语音片段是否为适格的子语音片段;以及
撷取被判断为适格的子语音片段,进行语音辨识。
2. 如权利要求 1 所述的利用声纹识别进行语音辨识的方法,其中当判断各该特定子语音片段为不适格的子语音片段,则舍弃该些不适格的子语音片段,不对其进行语音辨识。
3. 如权利要求 1 所述的利用声纹识别进行语音辨识的方法,其中,根据对该数个特定子语音片段分别进行该声纹识别处理的结果,判断各该特定子语音片段是否为适格的子语音片段的步骤,还包括:
依照时间先后顺序,定义该数个特定子语音片段为第一子语音片段、第二子语音片段、第三子语音片段至第 N 子语音片段,其中, N 为切割该数个特定子语音片段的个数;
通过该声纹识别处理,辨识该第一子语音片段、该第二子语音片段及该第三子语音片段;
当该第一子语音片段、该第二子语音片段、该第三子语音片段通过该声纹识别处理,被识别为同一特定用户的声纹,则接着辨识该第四子语音片段至该第 N 子语音片段,是否亦为该同一特定用户的声纹;
当判断该第四子语音片段至该第 N 子语音片段其中的一个或多个子语音片段亦为该同一特定用户的声纹,则判断是该同一特定用户的声纹的子语音片段,为适格的子语音片段;以及
撷取被判断为适格的子语音片段,进行语音辨识。
4. 如权利要求 3 所述的利用声纹识别进行语音辨识的方法,其中,辨识该第四子语音片段至该第 N 子语音片段是否亦为该同一特定用户的声纹步骤,还包括:
当判断该第四子语音片段至该第 N 子语音片段其中的一个或多个子语音片段并非该同一特定用户的声纹,则判断不是该同一特定用户的声纹的子语音片段,为不适格的子语音片段;以及
舍弃该些不适格的子语音片段,不对其进行语音辨识。
5. 如权利要求 1 所述的利用声纹识别进行语音辨识的方法,其中,根据该数个特定子语音片段分别进行该声纹识别处理的结果,判断各该特定子语音片段是否为适格的子语音片段的步骤,包括:
在一数据库储存一预设声纹辨识数据;以及
当判断该数个特定子语音片段其中的一个或多个特定子语音片段符合该预设声纹辨识数据的声纹,则判断符合该预设声纹辨识数据的声纹的子语音片段为适格的子语音片段;以及
撷取被判断为适格的子语音片段,进行语音辨识。
6. 如权利要求 5 所述的利用声纹识别进行语音辨识的方法,其中,根据该数个特定子语音片段分别进行该声纹识别处理的结果,判断各该特定子语音片段是否为适格的子语音

片段的步骤,还包括:

当判断该数个特定子语音片段其中的一个或多个特定子语音片段不符合该预设声纹辨识数据的声纹,则判断不符合该预设声纹辨识数据的声纹的子语音片段为不适格的子语音片段;以及

舍弃该些不适格的子语音片段,不对其进行语音辨识。

7. 如权利要求 1 所述的利用声纹识别进行语音辨识的方法,其中,该声纹识别处理是利用梅尔倒频谱系数运算方法实施。

8. 一种电子装置,包括:

一收发模块,接收一特定语音片段;以及

一控制模块,耦接至该收发模块,切割所接收的该特定语音片段为数个特定子语音片段,分别对该数个特定子语音片段进行一声纹识别处理,根据该数个特定子语音片段分别进行该声纹识别处理的结果,判断各该特定子语音片段是否为适格的子语音片段,撷取被判断为适格的子语音片段,进行语音辨识。

9. 如权利要求 8 所述的电子装置,其中该控制模块依照时间先后顺序,定义该数个特定子语音片段为第一子语音片段、第二子语音片段、第三子语音片段至第 N 子语音片段,其中, N 为该数个特定子语音片段的个数,该控制模块并通过该声纹识别处理,辨识该第一子语音片段、该第二子语音片段及该第三子语音片段,当该第一子语音片段、该第二子语音片段、该第三子语音片段通过该声纹识别处理,被识别为同一特定用户的声纹,则该控制模块接着辨识该第四子语音片段至该第 N 子语音片段,是否亦为该同一特定用户的声纹,当该控制模块判断该第四子语音片段至该第 N 子语音片段其中的一个或多个子语音片段亦为该同一特定用户的声纹,则判断是该同一特定用户的声纹的子语音片段为适格的子语音片段,该控制模块撷取被判断为适格的子语音片段,进行语音辨识。

10. 如权利要求 9 所述的电子装置,其中该控制模块当判断该第四子语音片段至该第 N 子语音片段其中的一个或多个子语音片段并非该同一特定用户的声纹,则判断不是该同一特定用户的声纹的子语音片段为不适格的子语音片段,该控制模块舍弃该些不适格的子语音片段,不对其进行语音辨识。

11. 如权利要求 10 所述的电子装置,还包括:

一数据库,耦接至该控制模块,预先储存一预设声纹辨识数据,当该控制模块判断该数个特定子语音片段其中的一个或多个特定子语音片段符合该预设声纹辨识数据的声纹时,则判断符合该预设声纹辨识数据的声纹的子语音片段为适格的子语音片段,该控制模块接着撷取被判断为适格的子语音片段,进行语音辨识。

12. 如权利要求 11 所述的电子装置,还包括当该控制模块判断该特定数个子语音片段其中的一个或多个特定子语音片段不符合该预设声纹辨识数据的声纹,则判断不符合该预设声纹辨识数据的声纹的子语音片段为不适格的子语音片段,该控制模块舍弃该些不适格的子语音片段,不对其进行语音辨识。

13. 如权利要求 8 所述的电子装置,其中,该声纹识别处理是利用梅尔倒频谱系数运算方法实施。

利用声纹识别进行语音辨识的方法及其电子装置

技术领域

[0001] 本发明涉及一种语音辨识的方法及其电子装置,特别是涉及一种利用声纹识别进行语音辨识的方法及其电子装置。

背景技术

[0002] 以现有技术而言,大多数的电子装置,例如手机、笔记型计算机、平板电脑等,都具有语音辨识模块,使用者可藉由声控的方式,控制电子装置执行各种指令。但是,在处理使用者语音对话输入时,由于外界的干扰噪音,很容易使得语音辨识模块错误的将使用者语音输入之外的其他声音作识别,而导致识别错误率上升。因此,如何提升使用者语音输入的正确辨识率,乃是目前语音辨识领域需要研究的课题。

发明内容

[0003] 本发明提供一种利用声纹识别进行语音辨识的方法,此方法包括下列步骤:接收一特定语音片段,切割所接收的特定语音片段为数个特定子语音片段,对此数个特定子语音片段分别进行一声纹识别处理,根据对此数个特定子语音片段分别进行声纹识别处理的结果,判断各特定子语音片段是否为适格的子语音片段。接着,撷取被判断为适格的子语音片段,进行语音辨识。

[0004] 在本发明的一实施例中,其中,根据对上述数个特定子语音片段分别进行前述声纹识别处理的结果,判断各特定子语音片段是否为适格的子语音片段的步骤,还包括:依照时间先后顺序,定义上述数个特定子语音片段为一第一子语音片段、第二子语音片段、第三子语音片段至第N子语音片段,其中,N为上述数个特定子语音片段的个数。接着,通过前述声纹识别处理,分别辨识第一子语音片段、第二子语音片段及第三子语音片段。当第一子语音片段、第二子语音片段、第三子语音片段通过声纹识别处理,被识别为同一特定用户的声纹,则接着辨识第四子语音片段至第N子语音片段,是否亦为此同一特定用户的声纹。当判断第四子语音片段至第N子语音片段其中的一个或多个子语音片段亦为此同一特定用户的声纹,则判断是此同一特定用户的声纹的子语音片段为适格的子语音片段,接着对这些适格的子语音片段进行语音辨识。

[0005] 在本发明的一实施例中,其中,辨识第四子语音片段至第N子语音片段是否亦为此同一特定用户的声纹的步骤,还包括:当判断第四子语音片段至第N子语音片段其中的一个或多个子语音片段并非此同一特定用户的声纹,则判断不是此同一特定用户的声纹的子语音片段为不适格的子语音片段。接着,舍弃这些不适格的子语音片段,不对其进行语音辨识。

[0006] 在本发明的一实施例中,其中,根据分别对上述数个特定子语音片段分别进行声纹识别处理的结果,判断各特定子语音片段是否为适格的子语音片段的步骤,包括:在一数据库提前储存一预设声纹辨识数据,当判断上述数个子语音片段其中的一个或多个子语音片段为符合预设声纹辨识数据的声纹时,则确定符合预设声纹辨识数据的声纹的子语音片

段为适格的子语音片段,接着,对这些适格的子语音片段进行语音辨识。

[0007] 在本发明的一实施例中,其中,根据分别对上述数个特定子语音片段分别进行声纹识别处理的结果,判断各特定子语音片段是否为适格的子语音片段的步骤,还包括:当判断上述数个语音片段其中的一个或多个子语音片段为不符合预设声纹辨识数据的声纹时,则确定不符合预设声纹辨识数据的声纹的子语音片段为不适格的子语音片段。接着,舍弃这些不适格的子语音片段,不对其进行语音辨识。

[0008] 在本发明的一实施例中,其中,上述的声纹识别处理是利用梅尔倒频谱系数运算方法(Mel-Frequency Cepstral Coefficients, MFCCs)实施。

[0009] 本发明提供一种利用声纹识别进行语音辨识的装置,包括:一收发模块、一控制模块。其中,收发模块接收一特定语音片段,而控制模块耦接至收发模块,控制模块切割所接收的特定语音片段为数个特定子语音片段,并对此数个特定子语音片段分别进行一声纹识别处理,根据此数个特定子语音片段分别进行声纹识别处理的结果,判断各特定子语音片段是否为适格的子语音片段。接着,撷取被判断为适格的子语音片段,进行语音辨识。

[0010] 基于上述,本发明提供一种利用声纹识别进行语音辨识的方法及其电子装置。在进行语音辨识之前,先通过声纹识别处理,判断经收发模块接收到的语音片段,切割之后的其中一个或多个子语音片段是否为适格的子语音片段,接着再撷取适格的子语音片段,对其进行语音辨识。依此使得语音交互的人机介面,不再因为其他噪音的干扰,而导致大量的错误辨识。

[0011] 为使本发明的上述特征和优点能更明显易懂,下文特举实施例,并结合附图详细说明如下。

附图说明

[0012] 图 1 是依照本发明实施例的一种电子装置的示意图。

[0013] 图 2 是依照本发明实施例的一种利用声纹识别进行语音辨识的方法的流程图。

[0014] 图 3 是依照本发明图 1 实施例的另一种利用声纹识别进行语音辨识的方法的流程图。

[0015] 图 4A 是依照本发明另一实施例的一种电子装置的示意图。

[0016] 图 4B 是依照本发明图 4A 实施例的一种利用声纹识别进行语音辨识的方法的流程图。

[0017] 附图符号说明

[0018] 100 :电子装置

[0019] 110 :收发模块

[0020] 120 :控制模块

[0021] 130 :数据库

[0022] S210 ~ S260、S310 ~ S344、S410 ~ S424 :利用声纹识别进行语音辨识的方法流程步骤

具体实施方式

[0023] 图 1 是依照本发明实施例的一种电子装置的示意图,电子装置 100 实施本发明所

提出的利用声纹识别进行语音辨识的方法。电子装置 100 包括收发模块 110、控制模块 120，两者相互耦接。其中，收发模块 110 可以是麦克风、音讯检测器等收音元件及喇叭、扩大机等播音元件的组合。控制模块 120 可以是硬件和 / 或硬件所实现的功能模块，其中，硬件可包括中央处理器、芯片组、微处理器等具有运算功能的硬件设备或上述硬件设备的组合，而硬件则可以是操作系统、驱动程序等等，控制模块 120 为提供本发明利用声纹识别进行语音辨识整体过程的主控元件。

[0024] 图 2 是依照本发明实施例的一种利用声纹识别进行语音辨识的方法的流程图，请同时参照图 1 及图 2。首先，收发模块 110 接收一特定语音片段（步骤 210），控制模块 120 切割收发模块 110 所接收的此特定语音片段为数个特定子语音片段（步骤 S220），接着，对此数个特定子语音片段分别进行一声纹识别处理（步骤 S230）。其中，声纹识别处理是利用梅尔倒频谱系数运算方法（Mel-Frequency Cepstral Coefficients, MFCCs）实施，在语音辨识（Speech Identification）和声纹辨识（Voiceprint Recognition）方面，最常用到的语音特征就是梅尔倒频谱系数，此参数考虑到听觉对于不同频率的感受程度，因此特别适合用在语音或声纹辨识。梅尔倒频谱系数的计算首先用快速傅立叶转换（Fast Fourier Transform, FFT）将时域信号转化成频域，之后对其对数能量谱用依照梅尔（Mel）刻度分布的三角滤波器组进行卷积，最后对各个滤波器的输出构成的向量进行离散余弦变换，进而得到结果。

[0025] 在步骤 S230 之后，控制模块 120 根据上述数个特定子语音片段分别进行该声纹识别处理的结果，判断各特定子语音片段是否为适格的子语音片段（步骤 S240）。当步骤 S240 的判断为否，则该控制模块 120 舍弃该些不适格的子语音片段，不对其进行语音辨识（步骤 S250）。当步骤 S240 判断为是，控制模块 120 撷取被判断为适格的子语音片段，进行语音辨识（步骤 S260）。

[0026] 图 3 是依照本发明另一实施例的一种利用声纹识别进行语音辨识的方法的流程图，请同时参照图 1 及图 3。图 3 在步骤 S210 ~ S230 与图 2 所揭示的流程相同，不再赘述。而在图 2 所揭示的步骤 S240，根据进行声纹识别处理的结果，控制模块 120 判断各特定子语音片段是否为适格的子语音片段的步骤，还在图 3 的实施例中，细分为步骤 S310 ~ S344，将于以下论述。图 3 的实施例，在步骤 S230 对上述数个特定子语音片段分别进行声纹识别处理之后，依照时间先后顺序，定义该数个特定子语音片段为第一子语音片段、第二子语音片段、第三子语音片段至第 N 子语音片段，其中，N 为上述数个特定子语音片段的个数（步骤 S310）。

[0027] 接着，通过前述声纹识别处理，辨识第一子语音片段、第二子语音片段及第三子语音片段（步骤 S320），判断此第一子语音片段、此第二子语音片段、此第三子语音片段通过前述声纹识别处理，是否被识别为同一特定用户的声纹（步骤 S330），也就是说，判断此第一子语音片段、此第二子语音片段、此第三子语音片段是否由同一位使用者的发出。当步骤 S330 的判断为否，则认定其为不适格的子语音片段（步骤 S342）。而当步骤 S330 的判断为是，则至步骤 S340，判断第四子语音片段至第 N 子语音片段其中的一个或多个子语音片段是否仍为此同一特定用户的声纹（步骤 S340）。当步骤 S340 的判断为否，则认定不是此同一特定用户的声纹的子语音片段为不适格的子语音片段（步骤 S342），当步骤 S340 的判断为是，则认定是此同一特定用户的声纹的子语音片段为适格的子语音片段（步骤 S344）。在步

骤 S342 之后,舍弃上述被判断为不适格的子语音片段,不对其进行语音辨识(步骤 S250)。而在步骤 S344 之后,撷取被判断为适格的子语音片段,进行语音辨识(步骤 S260)。

[0028] 图 4A 是依照本发明另一实施例的一种电子装置的示意图,图 4B 是依照本发明图 4A 实施例的一种利用声纹识别进行语音辨识的方法的流程图,请同时参照图 4A 及图 4B。图 4A 的功能方块图与图 1 实施例相较,其不同的处在于图 4A 具有数据库 130 耦接至控制模块 120,数据库 130 预先储存一预设声纹辨识数据。图 4B 在步骤 S210 ~ S230 与图 2 所揭示的流程相同,不再赘述。而在图 2 所揭示的步骤 S240,控制模块 120 根据进行声纹识别处理的结果,判断各特定子语音片段是否为适格的子语音片段的步骤,在图 4B 的实施例中,更细分为步骤 S410 ~ S424,将于以下论述。图 4B 的实施例,在数据库 130 预先储存一预设声纹辨识数据(步骤 S410)。接着,控制模块 120 判断上述数个特定子语音片段中的一个或多个子语音片段是否符合此预设声纹辨识数据的声纹(步骤 S420)。当步骤 S420 的判断为否,则判断不符合预设声纹辨识数据的声纹的子语音片段为不适格的子语音片段(步骤 S422)。当步骤 S420 的判断为是,则判断符合预设声纹辨识数据的声纹的子语音片段为适格的子语音片段(步骤 S424)。在步骤 S422 之后,舍弃上述被判断为不适格的子语音片段,不对其进行语音辨识(步骤 S250),而在步骤 S424 之后,撷取被判断为适格的子语音片段,进行语音辨识(步骤 S260)。

[0029] 综上所述,本发明提供一种利用声纹识别进行语音辨识的方法及其电子装置,在进行语音辨识之前,先通过声纹识别处理,判断经收发模块接收到的语音片段,切割之后的其中一个或多个子语音片段是否为适格的子语音片段,接着再撷取适格的子语音片段,对其进行语音辨识。藉由本发明所提供的判断流程,使得语音交互的人机介面,撷取真正属于使用者的语音对话片段,不再因为其他语音或噪音的干扰,而导致错误辨识,大大提升了语音辨识的正确率。

[0030] 虽然本发明已以实施例揭示如上,然其并非用以限定本发明,本领域技术人员,在不脱离本发明的精神和范围的前提下,可作些许的更动与润饰,故本发明的保护范围是以本发明的权利要求为准。

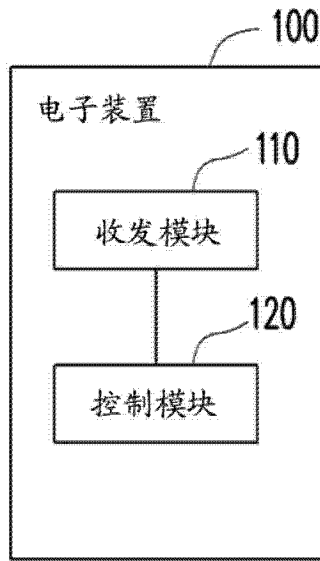


图 1

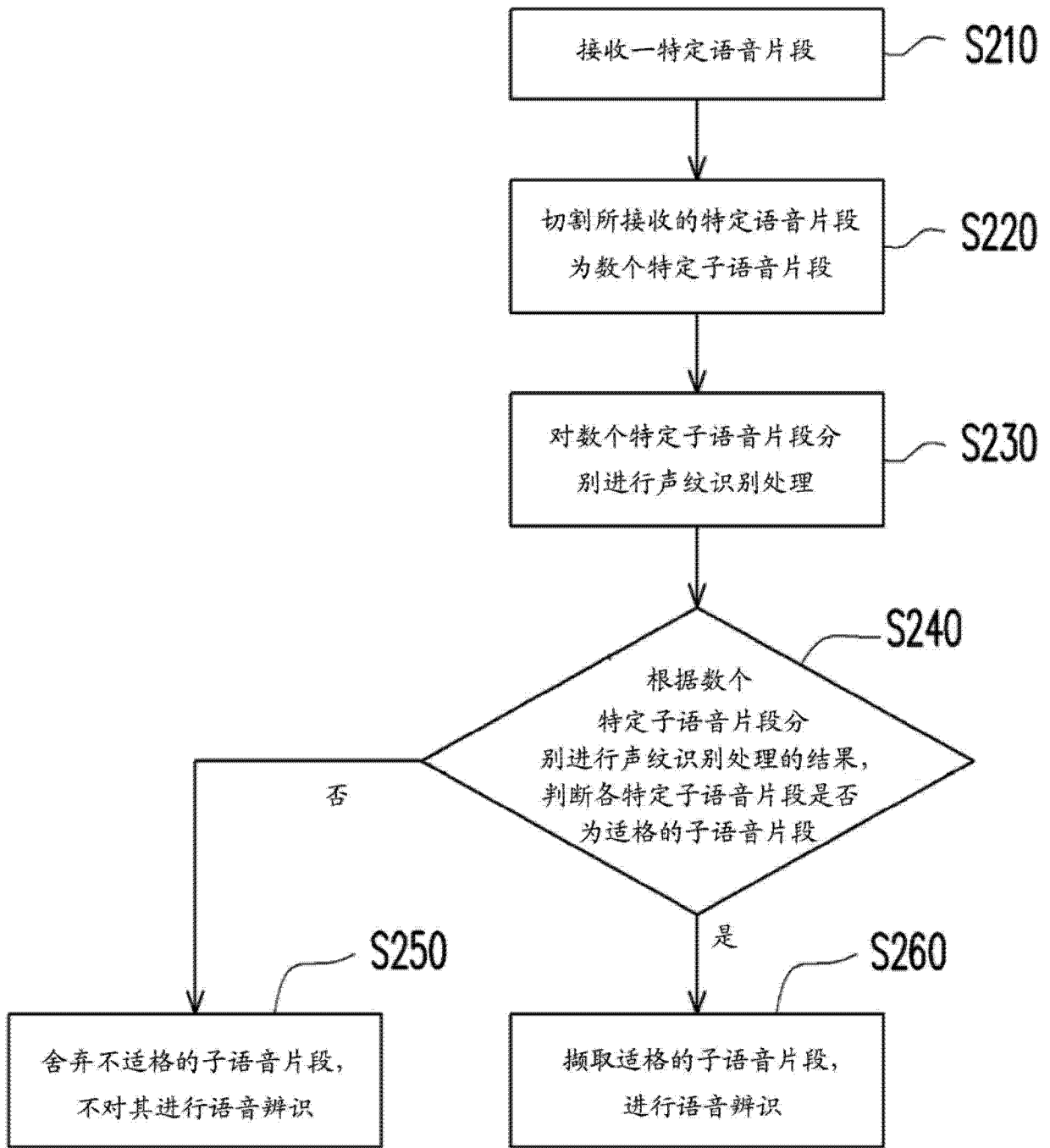


图 2

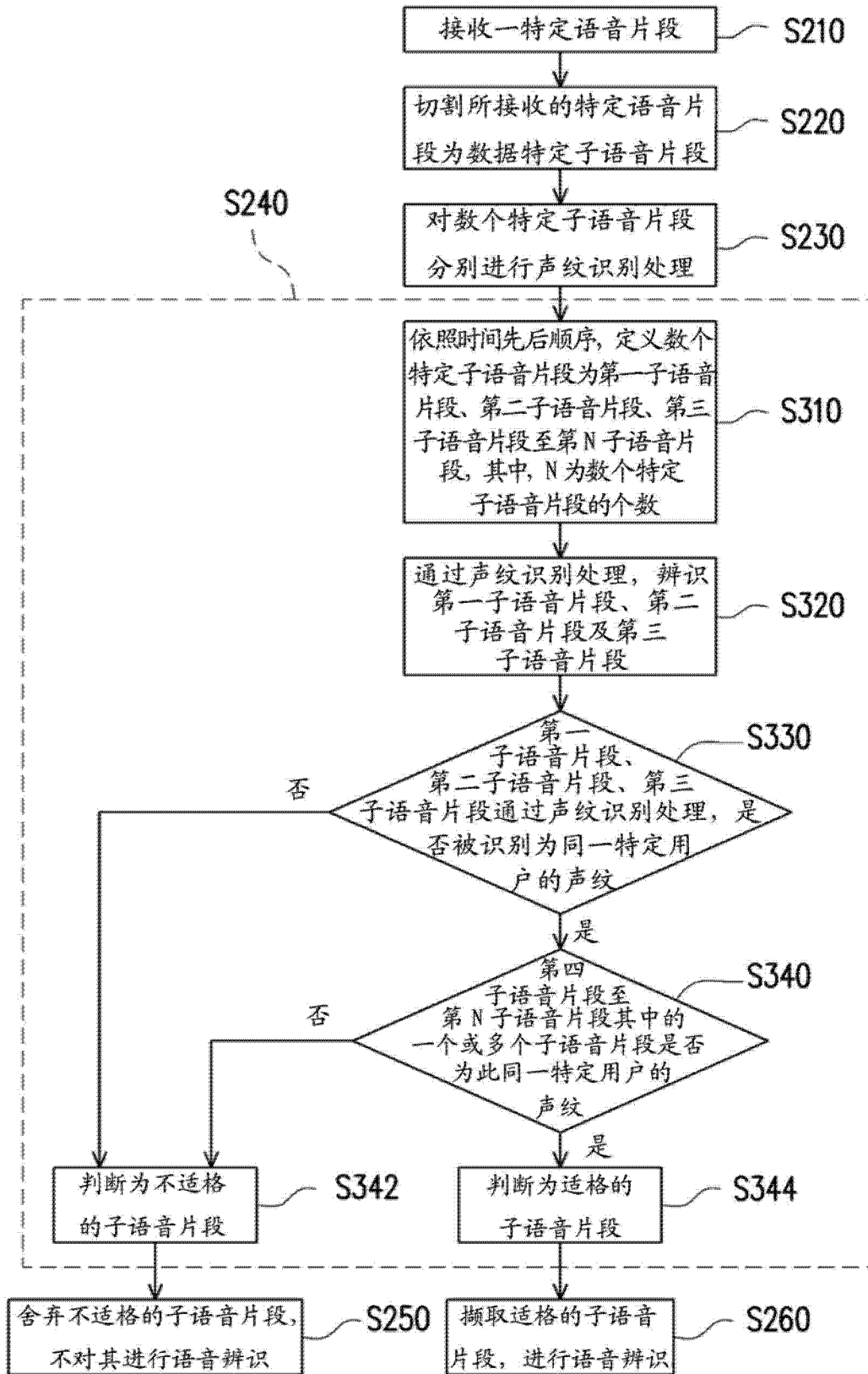


图 3

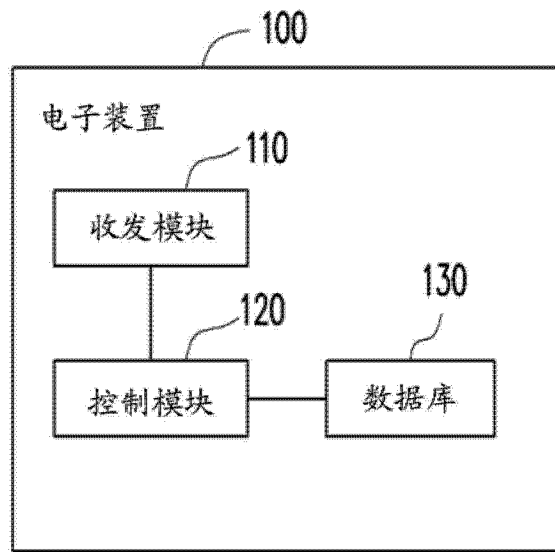


图 4A

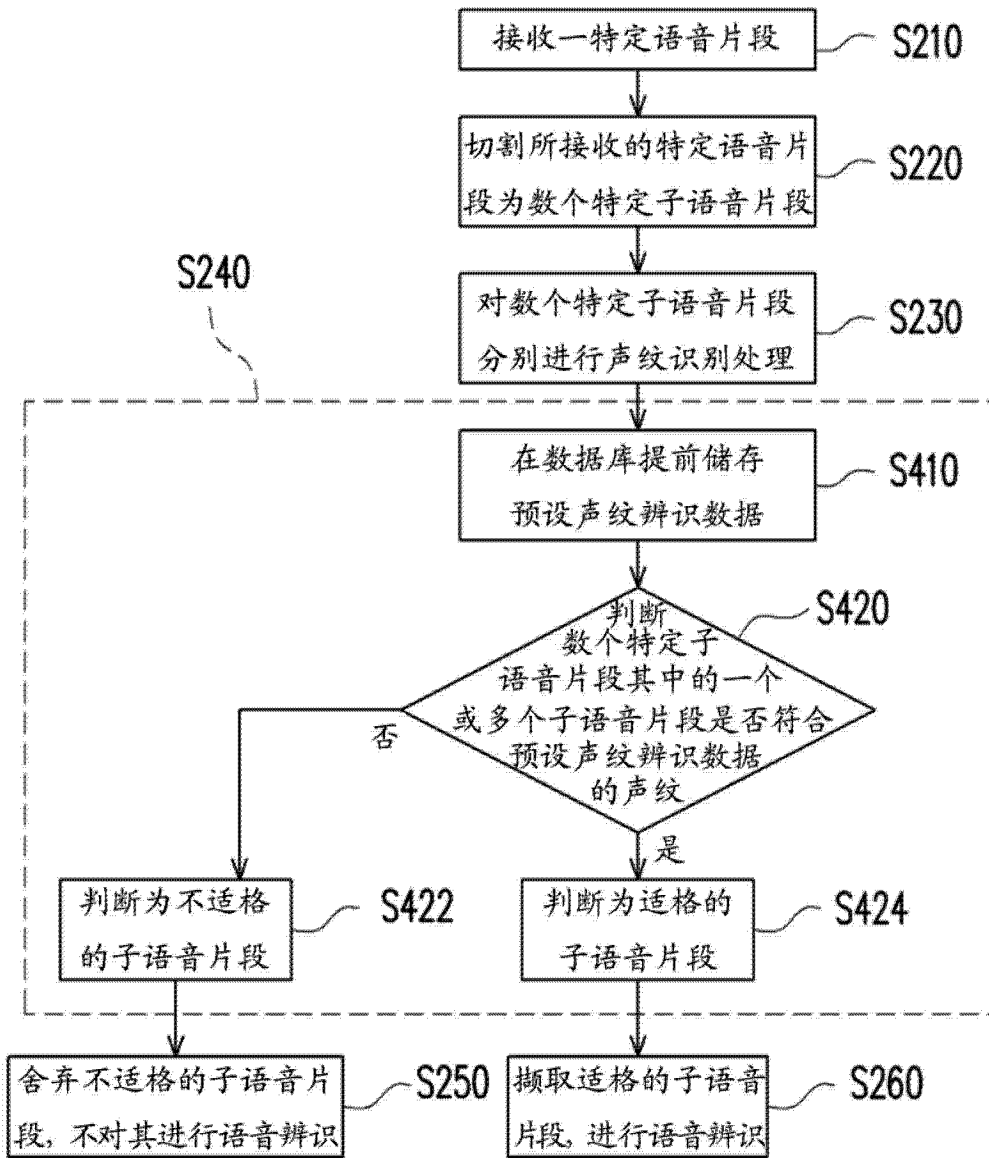


图 4B