(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109754782 B
(45)授权公告日 2020.10.09

(21)申请号 201910081818.3

(22)申请日 2019.01.28

(65)同一申请的已公布的文献号
申请公布号 CN 109754782 A

(43)申请公布日 2019.05.14

(73)专利权人 武汉恩特拉信息技术有限公司
地址 430223 湖北省武汉市东湖新技术开发区光谷大道3号激光工程设计总部二期研发楼06幢06单元15层5号(Y413)

(72)发明人 不公告发明人

(74)专利代理机构 北京知元同创知识产权代理事务所(普通合伙) 11535
代理人 张田勇 张祖萍

(51)Int.Cl.
*G10L 15/02*(2006.01)
*G10L 15/10*(2006.01)

(56)对比文件
CN 105103221 A,2015.11.25
CN 101510424 A,2009.08.19
CN 1622195 A,2005.06.01
JP 2018036413 A,2018.03.08
JP 2016151709 A,2016.08.22
WO 2011025532 A1,2011.03.03
于泓.合成语音检测算法研究.《中国博士学位论文全文库》.2018,
张立.计算机合成语音与自然语音鉴别技术的研究.《中国硕士学位论文全文库》.2018,

审查员 刘红梅

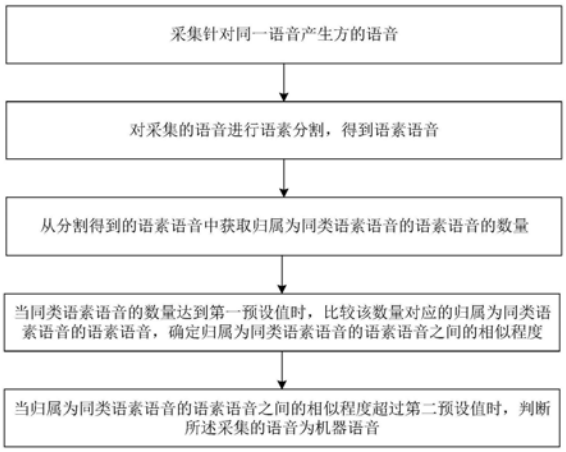
权利要求书2页 说明书6页 附图1页

(54)发明名称

一种辨别机器语音和自然语音的方法及装置

(57)摘要

本发明涉及一种辨别机器语音和自然语音的方法及装置,其中方法包括:采集针对同一语音产生方的语音;对采集的语音进行语素分割,得到语素语音;从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量;当同类语素语音的数量达到第一预设值时,比较该数量对应的归属为同类语素语音的语素语音,确定归属为同类语素语音的语素语音之间的相似程度;当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时,判断所述采集的语音为机器语音。本发明实施例提出的辨别机器语音和自然语音的方法及装置,能够快速准确地辨别出语音是机器语音还是自然语音。




CN 109754782 B

1.一种辨别机器语音和自然语音的方法,其特征在于,包括:

采集针对同一语音产生方的语音;

对采集的语音进行语素分割,得到语素语音;

从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量;

当同类语素语音的数量达到第一预设值时,比较该数量对应的归属为同类语素语音的语素语音,确定归属为同类语素语音的语素语音之间的相似程度;

当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时,判断所述采集的语音为机器语音;

其中,所述对采集的语音进行语素分割,是基于发音间隔进行语素分割,无需考虑语素的语义;

所述对采集的语音进行语素分割,得到语素语音,包括:

对采集的语音进行语音识别,并获取采集的语音的音节之间的发音间隔;

将识别的语音中发音间隔小于预设间隔值的一个或多个音节,作为一个语素语音。

2.根据权利要求1所述的方法,其特征在于,所述从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量,包括:

获取分割得到的语素语音的声音特征值;

将声音特征值接近的语素语音归属为同类语素语音;

对归属为同类语素语音的语素语音进行计数。

3.根据权利要求1所述的方法,其特征在于,所述比较该数量对应的归属为同类语素语音的语素语音,包括:

从采集的语音获取归属为同类语素语音的语素语音的发音长短和/或语调;

对归属为同类语素语音的语素语音的发音长短和/或语调进行比较。

4.根据权利要求1所述的方法,其特征在于,所述采集的语音包括对外部提问进行回答的语音。

5.一种辨别机器语音和自然语音的装置,其特征在于,包括:

采集模块,用于采集针对同一语音产生方的语音;

分割模块,用于对采集的语音进行语素分割,得到语素语音;

获取模块,用于从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量;

比较模块,用于当同类语素语音的数量达到第一预设值时,比较该数量对应的归属为同类语素语音的语素语音,确定归属为同类语素语音的语素语音之间的相似程度;

判断模块,用于当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时,判断所述采集的语音为机器语音;

其中,所述对采集的语音进行语素分割,是基于发音间隔进行语素分割,无需考虑语素的语义;

所述分割模块对采集的语音进行语素分割,得到语素语音,包括:

对采集的语音进行语音识别,并获取采集的语音的音节之间的发音间隔;

将识别的语音中发音间隔小于预设间隔值的一个或多个音节,作为一个语素语音。

6.根据权利要求5所述的装置,其特征在于,所述获取模块从分割得到的语素语音中获

取归属为同类语素语音的语素语音的数量,包括:

获取分割得到的语素语音的声音特征值;

将声音特征值接近的语素语音归属为同类语素语音;

对归属为同类语素语音的语素语音进行计数。

7.根据权利要求5所述的装置,其特征在于,所述比较模块比较该数量对应的归属为同类语素语音的语素语音,包括:

从采集的语音获取归属为同类语素语音的语素语音的发音长短和/或语调;

对归属为同类语素语音的语素语音的发音长短和/或语调进行比较。

8.根据权利要求5所述的装置,其特征在于,所述采集模块采集的语音包括对外部提问进行回答的语音。

# 一种辨别机器语音和自然语音的方法及装置

## 技术领域

[0001] 本发明属于语音识别技术领域，具体涉及一种辨别机器语音和自然语音的方法及装置。

## 背景技术

[0002] 语音识别技术已经是一个较为成熟的技术，已发展出众多的解决方案，满足不同的场景需求。例如在电话客服等场景中，准确率高的语音识别技术能够发挥一定的辅助作用，可以解决用户提出的简单的、常见的问题，但无法解决复杂的或者个性化的问题。因此，对于复杂的或者个性化的问题，用户还是希望客服是真正的人，以有效解决用户需要解决的问题。如何快速准确地分辨语音产生方的性质，成为本申请所要解决的技术问题。

## 发明内容

[0003] 为了解决上述快速准确地分辨语音产生方的性质的技术问题，本发明实施例提出了一种辨别机器语音和自然语音的方法及装置。

[0004] 在本发明的一方面，提出一种辨别机器语音和自然语音的方法，包括：

[0005] 采集针对同一语音产生方的语音；

[0006] 对采集的语音进行语素分割，得到语素语音；

[0007] 从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量；

[0008] 当同类语素语音的数量达到第一预设值时，比较该数量对应的归属为同类语素语音的语素语音，确定归属为同类语素语音的语素语音之间的相似程度；

[0009] 当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时，判断所述采集的语音为机器语音。

[0010] 在某些实施例中，所述对采集的语音进行语素分割，得到语素语音，包括：

[0011] 对采集的语音进行语音识别，并获取采集的语音的音节之间的发音间隔；

[0012] 将识别的语音中发音间隔小于预设间隔值的一个或多个音节，作为一个语素语音。

[0013] 在某些实施例中，所述从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量，包括：

[0014] 获取分割得到的语素语音的声音特征值；

[0015] 将声音特征值接近的语素语音归属为同类语素语音；

[0016] 对归属为同类语素语音的语素语音进行计数。

[0017] 在某些实施例中，所述比较该数量对应的归属为同类语素语音的语素语音，包括：

[0018] 从采集的语音获取归属为同类语素语音的语素语音的发音长短和/或语调；

[0019] 对归属为同类语素语音的语素语音的发音长短和/或语调进行比较。

[0020] 在某些实施例中，所述采集的语音包括对外部提问进行回答的语音。

[0021] 在本发明的另一方面，提出一种辨别机器语音和自然语音的装置，包括：

[0022] 采集模块，用于采集针对同一语音产生方的语音；

[0023] 分割模块，用于对采集的语音进行语素分割，得到语素语音；

[0024] 获取模块，用于从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量；

[0025] 比较模块，用于当同类语素语音的数量达到第一预设值时，比较该数量对应的归属为同类语素语音的语素语音，确定归属为同类语素语音的语素语音之间的相似程度；

[0026] 判断模块，用于当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时，判断所述采集的语音为机器语音。

[0027] 在某些实施例中，所述分割模块对采集的语音进行语素分割，得到语素语音，包括：

[0028] 对采集的语音进行语音识别，并获取采集的语音的音节之间的发音间隔；

[0029] 将识别的语音中发音间隔小于预设间隔值的一个或多个音节，作为一个语素语音。

[0030] 在某些实施例中，所述获取模块从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量，包括：

[0031] 获取分割得到的语素语音的声音特征值；

[0032] 将声音特征值接近的语素语音归属为同类语素语音；

[0033] 对归属为同类语素语音的语素语音进行计数。

[0034] 在某些实施例中，所述比较模块比较该数量对应的归属为同类语素语音的语素语音，包括：

[0035] 从采集的语音获取归属为同类语素语音的语素语音的发音长短和/或语调；

[0036] 对归属为同类语素语音的语素语音的发音长短和/或语调进行比较。

[0037] 在某些实施例中，所述采集模块采集的语音包括对外部提问进行回答的语音。

[0038] 本发明的有益效果：本发明实施例提出的辨别机器语音和自然语音的方法及装置，采用从语音中筛选出出现频率高的语素语音，比较这些语素语音之间的相似程度，能够快速准确地辨别出语音是机器语音还是自然语音。

## 附图说明

[0039] 图1是本发明实施例提出的辨别机器语音和自然语音的方法的流程图；

[0040] 图2是本发明实施例提出的辨别机器语音和自然语音的装置的结构示意图。

## 具体实施方式

[0041] 为使本发明的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本发明进一步详细说明。但本领域技术人员知晓，本发明并不局限于附图和以下实施例。

[0042] 电脑合成语音的基本原理是通过采样和再现的方式来实现语音合成。简单而言，无论什么语言都是由基本语素所构成，根据语言的语法环境和体系，决定基本语素的大小、组合方式和重复率。通过采集原始样本，针对所有基本语素的发音形成语素语音集合；然后基于语素语音集合，将一段需要转化为语音的语言文字所包括的所有基本语素整理出来，

按照一一对应的方式将该段语言文字替换成语素语音，按序播放，即形成电脑合成语音。早期的电脑合成语音带有明显的机械感，用户很容易分辨出电脑合成语音并非是真正的人发出的自然语言。随着信息存储量和即时处理能力等技术的提升，电脑合成语音的效果越来越接近真正的人说话的实际效果。另外在新闻播报、电话服务语音等应用领域，由于播音的规律性变强，导致真人播报的效果也越来越接近电脑合成语音的效果。

[0043] 本申请的申请人通过研究发现，可以通过反向分析电脑合成语音的原理来解决分别语音产生方的性质的技术问题。

[0044] 本发明实施例提出了一种辨别机器语音和自然语音的方法，如图1所示，包括：

[0045] 采集针对同一语音产生方的语音；

[0046] 对采集的语音进行语素分割，得到语素语音；

[0047] 从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量；

[0048] 当同类语素语音的数量达到第一预设值时，比较该数量对应的归属为同类语素语音的语素语音，确定归属为同类语素语音的语素语音之间的相似程度；

[0049] 当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时，判断所述采集的语音为机器语音。

[0050] 所述语音可以是汉语、英语、日语等各国或各地区语言构成的语音。所述机器语音主要指电脑合成语音，但不排除其他非真实的人的语音；所述自然语音指的是真实的人的语音。

[0051] 所述对采集的语音进行语素分割，是基于发音间隔进行语素分割，无需考虑语素的语义。

[0052] 所述第一预设值可以根据辨别的准确度等要求进行设定。一般来说，第一预设值越大，准确度会越高，考虑到可行性，可以将第一预设值设定在一个可操作的层面。在一个实施例中，所述第一预设值可以为5～10。

[0053] 在一个实施例中，所述第二预设值可以基于产生机器语音的语素样本，从包含多次出现同类语素语音的语音中，获取同类语素语音之间在发音长短、语调等方面的相似程度，由此获得第二预设值。

[0054] 在另一实施例中，所述第二预设值可以基于真实的人的语音，获取其中同类语素语音之间在发音长短、语调等方面的相似程度，由此获得第二预设值。为了提高第二预设值的准确度，所述真实的人的语音可以是新闻联播的播音员的语音。

[0055] 为了提高第二预设值的准确性，在又一实施例中，所述第二预设值可以综合考虑前述两个实施例，选取中间值作为第二预设值；在某些实施例中，所述第二预设值还可以是针对多个同类语素语音各自获取的相似程度，均值化后的值作为第二预设值。

[0056] 当归属为同类语素语音的语素语音之间的相似程度未超过第二预设值时，判断所述采集的语音为自然语音。

[0057] 本发明实施例还提出了一种辨别机器语音和自然语音的装置，如图2所示，包括：

[0058] 采集模块，用于采集针对同一语音产生方的语音；

[0059] 分割模块，用于对采集的语音进行语素分割，得到语素语音；

[0060] 获取模块，用于从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量；

[0061] 比较模块，用于当同类语素语音的数量达到第一预设值时，比较该数量对应的归属为同类语素语音的语素语音，确定归属为同类语素语音的语素语音之间的相似程度；

[0062] 判断模块，用于当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时，判断所述采集的语音为机器语音。

[0063] 本发明实施例提出的辨别机器语音和自然语音的装置与辨别机器语音和自然语音的方法相同的内容不再赘述。

[0064] 本发明实施例提出的辨别机器语音和自然语音的技术方案，申请人考虑到电脑合成语音中同类语素语音均来自事先采集的语素样本，彼此之间的相似程度会远高于来自真实的人的语音中同类语素语音之间的相似程度，因此，采用从语音中筛选出出现频率高的语素语音，比较这些语素语音之间的相似程度，能够快速准确地辨别出语音是机器语音还是自然语音。

[0065] 下面对本发明实施例提出的辨别机器语音和自然语音的技术方案，作进一步的示例性描述。

[0066] 实施例1：

[0067] 在本实施例中，所述对采集的语音进行语素分割，得到语素语音，包括：

[0068] 对采集的语音进行语音识别，并获取采集的语音的音节之间的发音间隔；

[0069] 将识别的语音中发音间隔小于预设间隔值的一个或多个音节，作为一个语素语音。

[0070] 所述音节按照语言的特点进行定义，例如对于汉语，一个汉字读音一般对应一个音节，音节通常包括声母、韵母和声调，对于某些字，音节可能不包括声母和/或声调；对于英语，一个单词读音可以对应一个或多个音节。

[0071] 所述预设间隔值的大小只要保证能够区分出各个音节即可。

[0072] 在一个实施例中，所述预设间隔值可以根据语音产生方的语速进行调整，针对语速较快的语音产生方的预设间隔值小于语速较慢的语音产生方的预设间隔值。

[0073] 本发明实施例提出的辨别机器语音和自然语音的技术方案，基于发音间隔对采集的语音进行语素分割，而不是基于语义分析进行语素分析，可以减少运算量。

[0074] 实施例2：

[0075] 在本实施例中，所述从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量，包括：

[0076] 获取分割得到的语素语音的声音特征值；

[0077] 将声音特征值接近的语素语音归属为同类语素语音；

[0078] 对归属为同类语素语音的语素语音进行计数。

[0079] 在一个实施例中，所述声音特征值包括基于音节获得的特征值。

[0080] 在一个实施例中，同类语素语音可以归入同一个语素语音集合，通过对集合中的语素语音进行计数，可以获取归属为同类语素语音的语素语音的数量。

[0081] 在一个实施例中，如果同一个语素语音集合中的数量超过第三预设值，可以将该集合中的语素语音进一步按照语义进行二次分类，并以二次分类后的语素语音集合作为所述同类语素语音。

[0082] 本发明实施例提出的辨别机器语音和自然语音的技术方案，通过语素语音的声音

特征值来判断同类语素语音，同样无需考虑语素的语义，有利于增加归属为同类语素语音的语素语音的数量，缩短辨别时间。进一步地，本发明实施例提出的辨别机器语音和自然语音的技术方案，还通过在声音特征值的基础上融合语义进行二次分类，可以提高辨别准确度，减少运算量。

[0083] 实施例3：

[0084] 在本实施例中，所述比较该数量对应的归属为同类语素语音的语素语音，包括：

[0085] 从采集的语音获取归属为同类语素语音的语素语音的发音长短和/或语调；

[0086] 对归属为同类语素语音的语素语音的发音长短和/或语调进行比较。

[0087] 在一个实施例中，还包括从采集的语音中获取归属为同类语素语音的语素语音的发音间隔。发音间隔能够反映语音产生方特别是真实的人的说话习惯。

[0088] 本发明实施例提出的辨别机器语音和自然语音的技术方案，通过获取语素语音的发音长短和/或语调，甚至语素语音的发音间隔，有利于捕捉不同语音产生方(特别是真实的人)的说话习惯，提高辨别准确性和辨别速度。

[0089] 实施例4：

[0090] 在本实施例中，所述采集的语音包含对外部提问进行回答的语音。

[0091] 申请人发现，同一语素语音在不同环境下可能具有不同的发音效果。例如一段对话中出现了电话号码和时间时，电话号码中的“3”的发音和时间“三分钟”中的“三”字的发音，很可能存在比较大的差异。为了避免样本采集工作量的加大以及运算量的加大，并且避免降低辨别准确率和辨别速度，在本实施例中增加了用户通过外部提问，主动引导语音产生方进行回答，促使语音产生方提供有针对性的语音，从而有助于提高辨别准确性和辨别速度。

[0092] 所述对外部提问进行回答的语音的获得，包括：

[0093] 针对语音产生方生成问题，并将生成的问题通知用户；

[0094] 接收语音产生方对用户提出的该问题的回答的语音，其中，所述回答中包含有预设答案。

[0095] 在一个实施例中，可以基于已经采集到的语音产生方的语音，实时生成一个或多个问题，作为所述生成的问题。

[0096] 在另一实施例中，可以从事先生成的问题中获取一个或多个问题，作为所述生成的问题。

[0097] 所述问题会引导语音产生方给出包含预设答案的回答，例如，所述问题是询问语音产生方一个电话号码(该电话号码为预设答案)，或者，所述问题是询问语音产生方一个特定的地址或者专用名词(特定地址或专用名词为预设答案)，再或者，所述问题是询问询问语音产生方一件物品的名称(物品的名称为预设答案)，等等。

[0098] 在一个实施例中，所述对采集的语音进行语素分割，得到语素语音，包括对采集的对外部提问进行回答的语音进行语素分割，得到所述语素语音。

[0099] 在另一个实施例中，所述对采集的语音进行语素分割，得到语素语音，包括对包含对外部提问进行回答的语音在内的语音进行语素分割，得到语素语音。

[0100] 本发明实施例提出的辨别机器语音和自然语音的技术方案，通过主动引导用户对语音产生方进行提问，并将预设答案包含在语音产生方的回答中，借助预设答案反映出不

同环境下同一语素语音的不同发音效果，从而提高辨别准确性和辨别速度。更进一步地，可以只针对包含对外部提问进行回答的语音进行辨别，能够更加快速地获得辨别结果，辨别准确性更高。

[0101] 本发明实施例还提出一种计算机可读存储介质，存储有执行前述方法的计算机程序。

[0102] 本发明实施例还提出一种计算机设备，包括处理器和操作上与所述处理器连接的上述计算机可读存储介质，所述处理器运行执行计算机可读介质中的计算机程序。

[0103] 本领域技术人员可以理解，在流程图中表示或在此以其他方式描述的逻辑和/或步骤，例如，可以被认为是用于实现逻辑功能的可执行指令的定序列表，可以具体实现在任何计算机可读介质中，以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用，或结合这些指令执行系统、装置或设备而使用。就本说明书而言，“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。

[0104] 计算机可读介质的更具体的示例(非穷尽性列表)包括以下：具有一个或多个布线的电连接部(电子装置)，便携式计算机盘盒(磁装置)，随机存取存储器(RAM)，只读存储器(ROM)，可擦除可编辑只读存储器(EPROM或闪速存储器)，光纤装置，以及便携式光盘只读存储器(CDROM)。另外，计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质，因为可以例如通过对纸或其他介质进行光学扫描，接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序，然后将其存储在计算机存储器中。

[0105] 应当理解，本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或它们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列(PGA)，现场可编程门阵列(FPGA)等。

[0106] 在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

[0107] 以上，对本发明的实施方式进行了说明。但是，本发明不限定于上述实施方式。凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

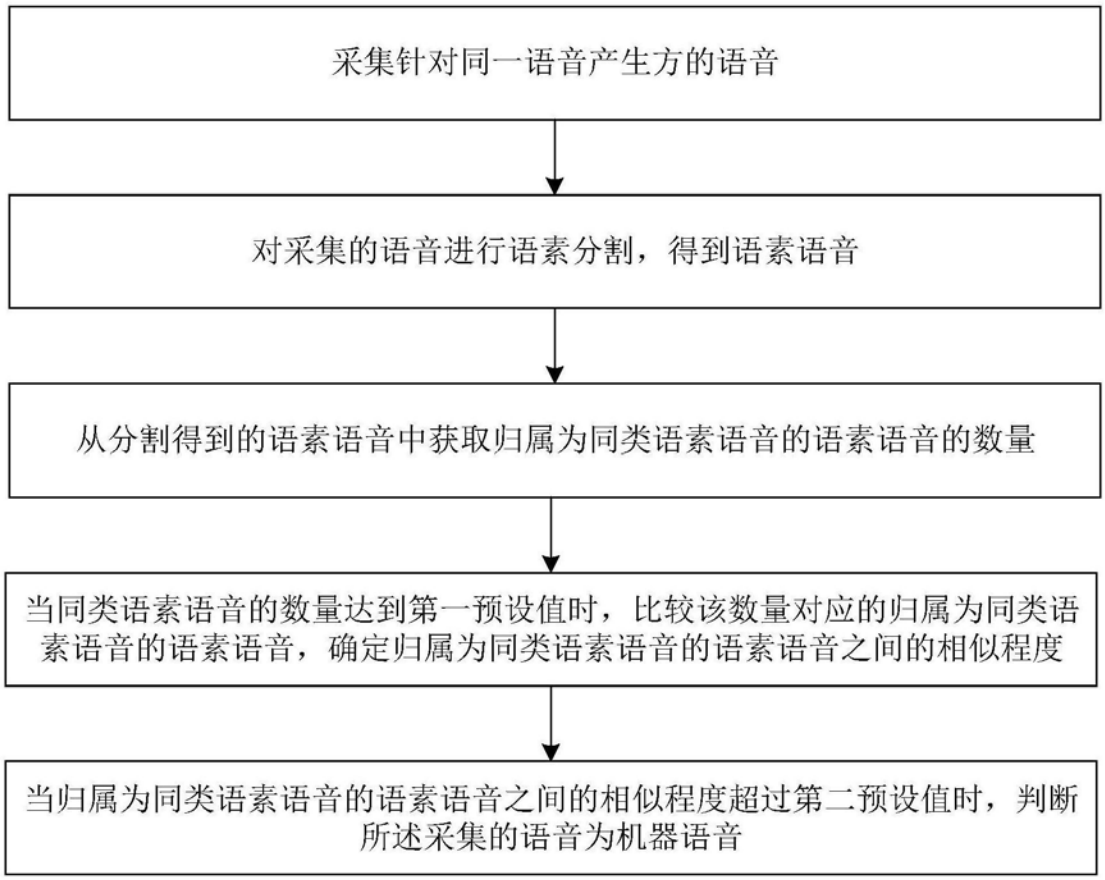

采集针对同一语音产生方的语音
对采集的语音进行语素分割，得到语素语音
从分割得到的语素语音中获取归属为同类语素语音的语素语音的数量
当同类语素语音的数量达到第一预设值时，比较该数量对应的归属为同类语素语音的语素语音，确定归属为同类语素语音的语素语音之间的相似程度
当归属为同类语素语音的语素语音之间的相似程度超过第二预设值时，判断所述采集的语音为机器语音

图1

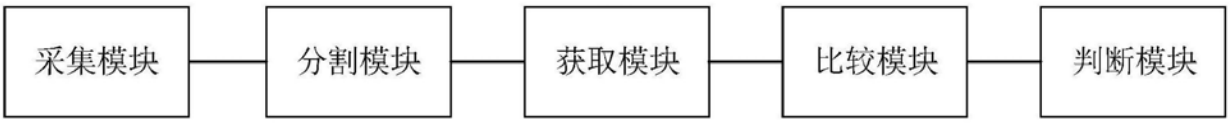

采集模块
分割模块
获取模块
比较模块
判断模块

图2