(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 106920558 B
(45)授权公告日 2021.04.13

(21)申请号 201510993729.8

(22)申请日 2015.12.25

(65)同一申请的已公布的文献号
申请公布号 CN 106920558 A

(43)申请公布日 2017.07.04

(73)专利权人 展讯通信(上海)有限公司
地址 201203 上海市浦东新区张江高科技园区祖冲之路2288弄展讯中心1号楼

(72)发明人 孙廷玮

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 郭学秀 吴敏

(51)Int.Cl.
G10L 17/04(2013.01)
G10L 17/16(2013.01)

(56)对比文件

Lindasalwa.“Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient and DTW techniques”.《Journal of Computing》.2010,第2卷(第3期),

Abhijeet Kumar.“Voice Command Recognition system based on MFCC and DTW”.《International Journal or engineering Science and Technology》.2010,

赵晓慧.“时间序列动态模糊聚类的研究”.《中国优秀硕士学位论文全文数据库 信息科技辑》.2014,

吴康妍.“一种结合端点检测可检错的DTW乐谱跟随算法”.《计算机应用与软件》.2015,

刘志镜.“加权DTW距离的自动步态识别”.《中国图像图形学报》.2010,

审查员 董小东

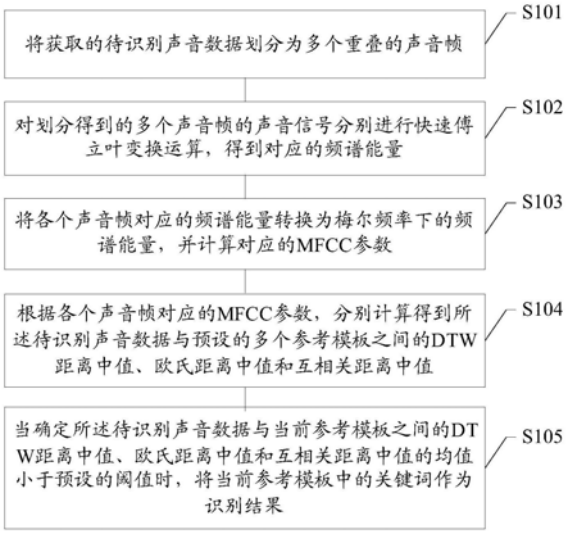
权利要求书2页 说明书6页 附图3页

(54)发明名称

关键词识别方法及装置

(57)摘要

关键词识别方法及装置,所述方法包括:将获取的待识别声音数据划分为多个重叠的声音帧;对划分得到的多个声音帧的声音信号分别进行快速傅立叶变换运算,得到对应的频谱能量;将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量,并计算对应的MFCC参数;根据各个声音帧对应的MFCC参数,分别计算得到所述待识别声音数据与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值;当确定所述待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时,将当前参考模板中的关键词作为识别结果。上述的方案,可以提高关键词识别的准确率,并节约计算资源。



CN 106920558 B

1.一种关键词识别方法,其特征在于,包括:

将获取的待识别声音数据划分为多个重叠的声音帧;

对划分得到的多个声音帧的声音信号分别进行快速傅立叶变换运算,得到对应的频谱能量;

将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量,并计算对应的MFCC参数;

根据各个声音帧对应的MFCC参数,分别计算得到所述待识别声音数据与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值;所述多个参考模板中分别包括对应的关键词的语音内容;在计算所述待识别的声音数据与所述预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值时,将所述待识别的声音数据与参考模板划分为I帧;用于DTW距离、欧式距离和互相关距离计算计算的每跳大小为0.1I帧,在计算得到当前待识别声音数据的I帧与参考模板的I帧的DTW距离、欧式距离和互相关距离之后,将I个DTW距离的中值作为所述待识别声音数据与对应的参考模板的DTW中值,将I个欧式距离的中值作为所述待识别声音数据与对应的参考模板的欧式距离中值,将I个互相关距离的中值作为所述待识别声音数据与对应的参考模板的互相关距离中值;

当确定所述待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时,将当前参考模板中的关键词作为所述待识别声音数据的关键词识别结果。

2.根据权利要求1所述的关键词识别方法,其特征在于,在所述待识别声音数据的频谱能量大于预设的能量阈值时,执行所述将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量,并计算对应的MFCC参数的操作。

3.根据权利要求1所述的关键词识别方法,其特征在于,所述预设的阈值与所述待识别声音数据的噪音水平相关联。

4.根据权利要求3所述的关键词识别方法,其特征在于,所述待识别声音数据的噪音水平包括低噪音水平、中等噪音水平和高噪音水平,其中:

当p≥p1时,确定所述待识别声音数据具有低噪音水平,p表示所述待识别声音数据对应的绝对幅值,p1为预设的第一阈值;

当p2≥p＞p1时,确定所述待识别声音数据具有中等噪音水平,p2为预设的第二阈值,且p1＞p2;

当p＜p2时,确定所述待识别声音数据具有高噪音水平。

5.根据权利要求4所述的关键词识别方法,其特征在于,p1等于0.8,p2等于0.45。

6.根据权利要求1所述的关键词识别方法,其特征在于,所述参考模板中包括瞬态噪声、静态噪声和特定人的丰富的语音内容的信息。

7.一种关键词识别装置,其特征在于,包括:

分帧处理单元,适于将获取的待识别的声音数据划分为多个重叠的声音帧;

频域转换单元,适于对划分得到的多个声音帧的声音信号分别进行快速傅立叶变换运算,得到对应的频谱能量;

第一计算单元,适于将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量,并计算对应的MFCC参数;

第二计算单元，适于根据各个声音帧对应的MFCC参数，分别计算得到所述待识别声音数据与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值；所述多个参考模板中分别包括对应的关键词的语音内容；在计算所述待识别的声音数据与所述预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值时，将所述待识别的声音数据与参考模板划分为I帧；用于DTW距离、欧式距离和互相关距离计算计算的每跳大小为0.1I帧，在计算得到当前待识别声音数据的I帧与参考模板的I帧的DTW距离、欧式距离和互相关距离之后，将I个DTW距离的中值作为所述待识别声音数据与对应的参考模板的DTW中值，将I个欧式距离的中值作为所述待识别声音数据与对应的参考模板的欧式距离中值，将I个互相关距离的中值作为所述待识别声音数据与对应的参考模板的互相关距离中值；

判断单元，适于判断当前声音帧与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值三者的均值是否小于预设的阈值；

关键词识别单元，适于当确定所述待识别声音数据与当前参考模板之间的DTW距离、欧氏距离中值和互相关距离中值的均值小于预设的阈值时，将当前参考模板中的关键词作为所述待识别声音数据的关键词识别结果。

8.根据权利要求7所述的关键词识别装置，其特征在于，还包括触发单元，所述触发单元适于在所述待识别声音数据的频谱能量大于预设的能量阈值时，触发所述第一计算单元执行所述将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数的操作。

9.根据权利要求7所述的关键词识别装置，其特征在于，所述预设的阈值与所述待识别声音数据的噪音水平相关联。

10.根据权利要求9所述的关键词识别装置，其特征在于，所述待识别声音数据的噪音水平包括低噪音水平、中等噪音水平和高噪音水平，其中：

当p≥p1时，确定所述待识别声音数据具有低噪音水平，p表示所述待识别声音数据对应的绝对幅值，p1为预设的第一阈值；

当p2≥p＞p1时，确定所述待识别声音数据具有中等噪音水平，p2为预设的第二阈值，且p1＞p2；

当p＜p2时，确定所述待识别声音数据具有高噪音水平。

11.根据权利要求10所述的关键词识别装置，其特征在于，p1等于0.8，p2等于0.45。

12.根据权利要求7所述的关键词识别装置，其特征在于，所述参考模板中包括瞬态噪声、静态噪声和特定人的丰富的语音内容的信息。

关键词识别方法及装置

技术领域

[0001] 本发明涉及语音识别技术领域,特别是涉及一种关键词识别方法及装置。

背景技术

[0002] 语音识别是机器通过识别和理解过程将人的语音转换为对应的文本或指令的技术。作为语音识别领域的一个重要分支,关键词(Isolated Word Recognition,IWR)识别在通信、消费电子、自助服务、办公自动化等领域得到了广泛的应用。

[0003] 现有技术中,一般采用隐马尔可夫模型(Hidden Markov Model,HMM)hidden Markov models(HMMs)及其对应的参数,或者关键词识别系统(KWS)进行关键词识别。

[0004] 但是,现有技术中关键词识别方法需要建立对应的模型,并需要对应的翻译操作训练模型参数,存在着计算量大且识别准确率低的问题。

发明内容

[0005] 本发明实施例解决的问题是提高关键词识别的准确率,并节约计算资源。

[0006] 为解决上述问题,本发明实施例提供了一种关键词识别方法,所述关键词识别方法包括:

[0007] 将获取的待识别声音数据划分为多个重叠的声音帧;

[0008] 对划分得到的多个声音帧的声音信号分别进行快速傅立叶变换运算,得到对应的频谱能量;

[0009] 将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量,并计算对应的MFCC参数;

[0010] 根据各个声音帧对应的MFCC参数,分别计算得到所述待识别声音数据与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值;

[0011] 当确定所述待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时,将当前参考模板中的关键词作为识别结果。

[0012] 可选地,在所述待识别声音数据的频谱能量大于预设的能量阈值时,执行所述将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量,并计算对应的MFCC参数的操作。

[0013] 可选地,所述预设的阈值与所述待识别声音数据的噪音水平相关联。

[0014] 可选地,所述待识别声音数据的噪音水平包括低噪音水平、中等噪音水平和高噪音水平,其中:

[0015] 当p≥p1时,确定所述待识别声音数据具有低噪音水平,p表示所述待识别声音数据对应的绝对幅值,p1为预设的第一阈值;

[0016] 当p2≥p＞p1时,确定所述待识别声音数据具有中等噪音水平,p2为预设的第二阈值,且p1＞p2;

[0017] 当p＜p2时,确定所述待识别声音数据具有高噪音水平。

[0018] 可选地，p1等于0.8，p2等于0.45。

[0019] 可选地，所述参考模板中包括瞬态噪声、静态噪声和特定人的丰富的语音内容的信息。

[0020] 本发明实施例还提供了一种关键词识别装置，所述装置包括：

[0021] 分帧处理单元，适于将获取的待识别的声音数据划分为多个重叠的声音帧；

[0022] 频域转换单元，适于对划分得到的多个声音帧的声音信号分别进行快速傅立叶变换运算，得到对应的频谱能量；

[0023] 第一计算单元，适于将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数；

[0024] 第二计算单元，适于根据各个声音帧对应的MFCC参数，分别计算得到所述待识别声音数据与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值；

[0025] 判断单元，适于判断当前声音帧与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值三者的均值是否小于预设的阈值；

[0026] 关键词识别单元，适于当确定所述待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时，将当前参考模板中的关键词作为识别结果。

[0027] 可选地，还包括触发单元，所述触发单元适于在所述待识别声音数据的频谱能量大于预设的能量阈值时，触发所述第一计算单元执行所述将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数的操作。

[0028] 可选地，所述预设的阈值与所述待识别声音数据的噪音水平相关联。

[0029] 可选地，所述待识别声音数据的噪音水平包括低噪音水平、中等噪音水平和高噪音水平，其中：

[0030] 当p≥p1时，确定所述待识别声音数据具有低噪音水平，p表示所述待识别声音数据对应的绝对幅值，p1为预设的第一阈值；

[0031] 当p2≥p＞p1时，确定所述待识别声音数据具有中等噪音水平，p2为预设的第二阈值，且p1＞p2；

[0032] 当p＜p2时，确定所述待识别声音数据具有高噪音水平。

[0033] 可选地，p1等于0.8，p2等于0.45。

[0034] 可选地，所述参考模板中包括瞬态噪声、静态噪声和特定人的丰富的语音内容的信息。

[0035] 与现有技术相比，本发明的技术方案具有以下的优点：

[0036] 上述的方案，通过基于对应MFCC参数计算得到的待识别声音数据与参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值与预设的阈值进行比较，来确定声音帧中是否包括关键词，而无需建立对应的数学识别模型，也不需要对关键词进行相应的翻译，因此，可以节关键词识别的的计算资源，并可以提高关键词识别的准确率。

[0037] 进一步地，当待识别声音数据的频谱能量大于预设的能量阈值时，才对对应的待识别声音数据进行关键词识别，反之，则不对待识别声音数据进行关键词识别，因此，可以进一步节约计算资源，并提高关键词识别的速度。

[0038] 进一步地，在录制对应的参考模板时，所述参考模板中包括瞬态噪声、静态噪声和

特定人的丰富的语音内容的信息，使得参考模板可以与对应的特定人的语音和语音所属环境进行较为准确地记录，因此，可以进一步提高关键词识别的准确性。

附图说明

[0039] 图1是本发明实施例中的一种关键词识别方法的流程图；

[0040] 图2是本发明实施例中的另一种关键词识别方法的流程图；

[0041] 图3是本发明实施例中的一种关键词识别装置的结构示意图。

具体实施方式

[0042] 为解决现有技术中存在的上述问题，本发明实施例采用的技术方案通过在确定待识别声音数据与参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值与预设的阈值进行比较，来确定声音帧中是否包括关键词，可以节关键词识别的计算资源，并可以提高关键词识别的准确率。

[0043] 为使本发明的上述目的、特征和优点能够更为明显易懂，下面结合附图对本发明的具体实施例做详细的说明。

[0044] 图1示出了本发明实施例中的一种关键词识别方法的流程图。如图1所示的关键词识别方法，可以包括如下步骤：

[0045] 步骤S101：将获取的待识别声音数据划分为多个重叠的声音帧。

[0046] 在具体实施中，各个声音帧之间的重叠部分的大小可以根据实际的需要进行设置。例如，当各个声音帧的长度为32ms时，相邻声音帧之间的重叠部分的大小可以为16ms。

[0047] 步骤S102：对划分得到的多个声音帧的声音信号分别进行快速傅立叶变换运算，得到对应的频谱能量。

[0048] 在具体实施中，划分得到的多个声音信号为时域的声音信号，通过快速傅立叶变换运算(FFT)，可以将时域的声音信号转换为频域的声音信号。

[0049] 步骤S103：将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数。

[0050] 在具体实施中，经过快速傅立叶变换运算得到声音信号的频谱能量(功率谱)，可以按照预设的对应关系，转换为梅尔频率下的频谱能量，并根据梅尔频率下的频谱能量，计算各个声音帧对应的梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient，MFCC)参数。

[0051] 步骤S104：根据各个声音帧对应的MFCC参数，分别计算得到所述待识别声音数据与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值。

[0052] 在具体实施中，预设的多个参考模板中分别包括对应的关键词的语音内容。其中，预设的参考模板的数量可以根据实际的需要进行设置，本发明在此不做限制。

[0053] 步骤S105：当确定所述待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时，将当前参考模板中的关键词作为识别结果。

[0054] 在具体实施中，通过对预设的多个参考模板进行遍历，分别计算当前待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值，并将当前待识

别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值与预设的阈值进行比较，当确定所述待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时，可以将当前参考模板中的关键词作为识别结果；反之，则确定当前待识别的声音数据中不包括当前参考模板中的关键词的语音信息。

[0055] 下面将结合图2对本发明实施例中的关键词识别方法做进一步详细的介绍。

[0056] 图2示出了本发明实施例中的另一种关键词识别方法的流程图。如图2所示的关键词识别方法，可以包括如下的步骤：

[0057] 步骤S201：将获取的声音数据进行重叠分帧，得到对应的多个声音帧。

[0058] 在具体实施中，首先可以对所采集的声音信号进行模数转换，得到对应的声音数据。接着，可以将对应的声音数据进行重叠分帧，得到多个声音帧。对采集的声音数据进行分帧，实质是对声音数据进行短时分析。短时分析是把声音信号分成具有固定周期的时间短段，每个时间短段是相对固定的持续声音片段。其中，相邻的两个声音帧之间部分重叠，重叠范围可以根据实际情况进行选择。

[0059] 步骤S202：对所得到的多个声音帧进行加窗处理。

[0060] 在具体实施中，可以选择汉明窗、汉宁窗、矩形窗等语音信号处理常用的窗函数，帧长选择为10～40ms，典型值为20ms。其中，对语音信号进行分帧处理破坏了声音信号的自然度，通过使用声音帧进行加窗和回移处理等，可以解决这个问题。

[0061] 步骤S203：将经过加窗处理后的声音帧进行快速傅立叶变换运算，得到各个声音帧对应的频谱能量的信息。

[0062] 在具体实施中，声音数据理论上来说是随时间变化的，是一个非稳态的过程，不可以直接进行频域的转换。但是，由于对声音数据进行分帧处理(短时分析)，每帧的声音数据可以认为是相对稳定的，因而可以对其应用频域转换。

[0063] 在具体实施中，可以采用短时傅立叶变换(Short-Time Fourier Transform/Short-Term Fourier Transform，STFT)对每帧的声音数据进行频域转换，以得到各个声音帧对应的频谱信息。其中，所得到的频谱中包括对应的声音信号的频率和能量的关系。

[0064] 步骤S204：将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数。

[0065] 在本发明一实施例中，当得到当前待识别声音数据的多个声音帧对应的频谱能量之后，可以首先判断当前待识别声音数据的频谱能量是否大于预设的能量阈值，当确定当前待识别声音数据的频谱能量大于所述能量阈值时，继续执行步骤S204，否则，确定当前待识别声音数据中不包括关键词的语音信息，因此，便可以停止对当前待识别声音数据的后续处理，以进一步节约计算资源。

[0066] 在具体实施中，可以按照预设的对应关系，将经过FFT运算得到的频谱能量转换成为梅尔(Mel)频率下的频谱能量，并计算每个声音帧对应的MFCC参数，作为每个声音帧的特征向量。

[0067] 步骤S205：根据各个声音帧对应的MFCC参数，计算得到当前声音帧与预设的多个参考模板中当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值。

[0068] 在本发明一实施例中，在计算当前待识别的声音数据与参考模板之间的DTW距离

时，当前待识别的声音数据和参考模板分别被划分成I帧。同时，本申请的发明人根据经验获知，在参考模板的录制过程中，说话者的发音会变得亢奋，且语速也比往常要慢。因此，将参考模板划分为I帧，用于DTW距离计算的每跳大小为0.1I帧，在计算得到当前待识别声音数据的I帧与参考模板的I帧的DTW距离之后，将I个DTW距离的中值作为当前待识别声音数据与对应的参考模板的DTW距离中值。类似地，我们可以得到当前待识别声音数据与对应的参考模板的欧式距离(ED)中值和互相关距离(CC)距离中值。

[0069] 步骤S206：判断待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值是否小于预设的阈值；当判断结果为是时，可以执行步骤S207，反之，则对预设的多个参考模板中的下一参考模板从步骤S205开始执行。

[0070] 在具体实施中，在计算得到当前待识别声音数据与参考模板之间的DTW距离中值、欧式距离中值和互相关距离中值之后，将三者的均值与预设的阈值进行比较。

[0071] 在本发明一实施例中，所述预设的阈值与当前待识别声音数据的噪音水平相关联，即不同的噪音水平，对应的预设的阈值将会不同。其中，当前待识别声音数据的绝对幅值概率大于当p≥p1时，确定所述待识别声音数据具有低噪音水平，p表示所述待识别声音数据对应的绝对幅值，p1为预设的第一阈值；p2≥p＞p1时，确定所述待识别声音数据具有中等噪音水平，p2为预设的第二阈值，且p1＞p2；当p＜p2时，确定所述待识别声音数据具有高噪音水平。在本发明一实施例中，p1为0.8，p2为0.45。

[0072] 步骤S207：将当前参考模板中的关键词作为识别结果并输出。

[0073] 在具体实施中，当确定预设的参考模板中的某个参考模板与当前待识别声音数据之间的DTW距离中值、欧式距离中值和互相关距离中值的均值小于预设的阈值时，可以确定当前待识别声音数据中包括参考模板中的关键词的语音信息。因此，可以将所述参考模板中的关键词作为当前待识别声音数据的关键词识别结果并输出。

[0074] 在具体实施中，当将上述的关键词识别方法应用于告警系统中时，在识别出对应的关键词时，告警系统可以执行告警操作。

[0075] 这里需要指出的是，在紧急情况或者其他的关键词应用中，单纯(如未经训练)的用户可以用于录制个性化关键词。为了确保良好的识别性能，参考模板变得非常重要。这可以通过简单的核查操作来确保参考模板的录制质量。

[0076] 因此，本申请的发明人提倡三种检测因素，即检测瞬态噪声源(如摔门声)，静态噪声源(如风扇或者交通噪声)，且丰富关键词的发音内容。上述三种因素需要同时满足，否则将需要重新录制关键词。其中，瞬态噪声的检测，可以使用连续25ms的声音帧，且每跳大小为5ms的声音信号的能量的绝对幅值的差异。其中，可以将每5个声音帧的绝对幅值进行平均。在静态噪声检测时，关键词的录制发生在安静环境中预设的5s时间窗内。与包括关键词的声音数据相比，在5s时间窗内，不包括关键词的参考模板的开头和结尾的信号能量具有较大的差异。在核查丰富的发音内容时，只有单一元音而没有如“啊”之类的辅音的关键词是被拒绝的，这种拒绝可以基于与关键词的发音内容相关的修正过零率做出。

[0077] 下面将对本发明实施例中的关键词识别方法对应的装置做进一步详细的介绍。

[0078] 请参见图3，本发明实施例中的关键词识别装置300，可以包括分帧处理单元301、频域转换单元302、第一计算单元303、第二计算单元304、判断单元305和关键词识别单元306，其中：

[0079] 所述分帧处理单元301，适于将获取的待识别的声音数据划分为多个重叠的声音帧；

[0080] 所述频域转换单元302，适于对划分得到的多个声音帧进行遍历，并将遍历到的当前声音帧的声音信号进行快速傅立叶变换运算，得到对应的频谱能量；

[0081] 所述第一计算单元303，适于将所得到的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数；

[0082] 在具体实施中，在所述关键词识别装置300还可以设置一个触发单元(图中未示出)，该触发单元适于在遍历到的当前声音帧的频谱能量大于预设的能量阈值时，触发所述第一计算单元303执行所述将所得到频谱能量转换为Mel频率下的频谱能量，并计算对应的MFCC参数的操作；

[0083] 所述第二计算单元304，适于根据当前声音帧对应的MFCC参数，分别计算得到当前声音帧与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值；

[0084] 所述判断单元305，适于判断当前声音帧与参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值是否小于预设的阈值；

[0085] 在具体实施中，所述预设的阈值与当前声音帧的噪音水平相关联，其中，当$p \geq p1$时，确定当前声音帧具有低噪音水平，p表示当前声音帧对应的绝对幅值，p1为预设的第一阈值；当$p2 \geq p > p1$时，确定当前声音帧具有中等噪音水平，p2为预设的第二阈值，且$p1 > p2$；当$p < p2$时，确定当前声音帧具有高噪音水平。其中，在本发明一实施例中，p1等于0.8，p2等于0.45。

[0086] 在具体实施中，所述参考模板中包括瞬态噪声、静态噪声和特定人的丰富的语音内容的信息。

[0087] 所述关键词识别单元306，适于当确定当前声音帧与参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时，将当前参考模板中的关键词作为识别结果并输出。

[0088] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成，该程序可以存储于计算机可读存储介质中，存储介质可以包括：ROM、RAM、磁盘或光盘等。

[0089] 以上对本发明实施例的方法及系统做了详细的介绍，本发明并不限于此。任何本领域技术人员，在不脱离本发明的精神和范围内，均可作各种更动与修改，因此本发明的保护范围应当以权利要求所限定的范围为准。

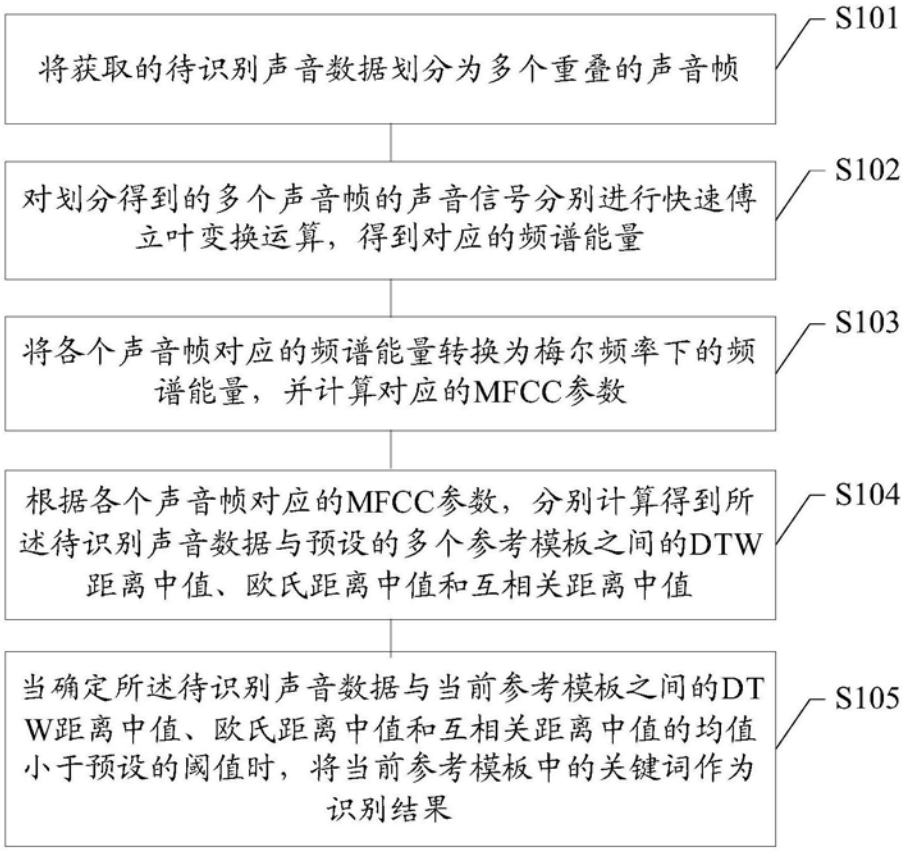

S101
将获取的待识别声音数据划分为多个重叠的声音帧
S102
对划分得到的多个声音帧的声音信号分别进行快速傅立叶变换运算，得到对应的频谱能量
S103
将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数
S104
根据各个声音帧对应的MFCC参数，分别计算得到所述待识别声音数据与预设的多个参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值
S105
当确定所述待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值时，将当前参考模板中的关键词作为识别结果

图1

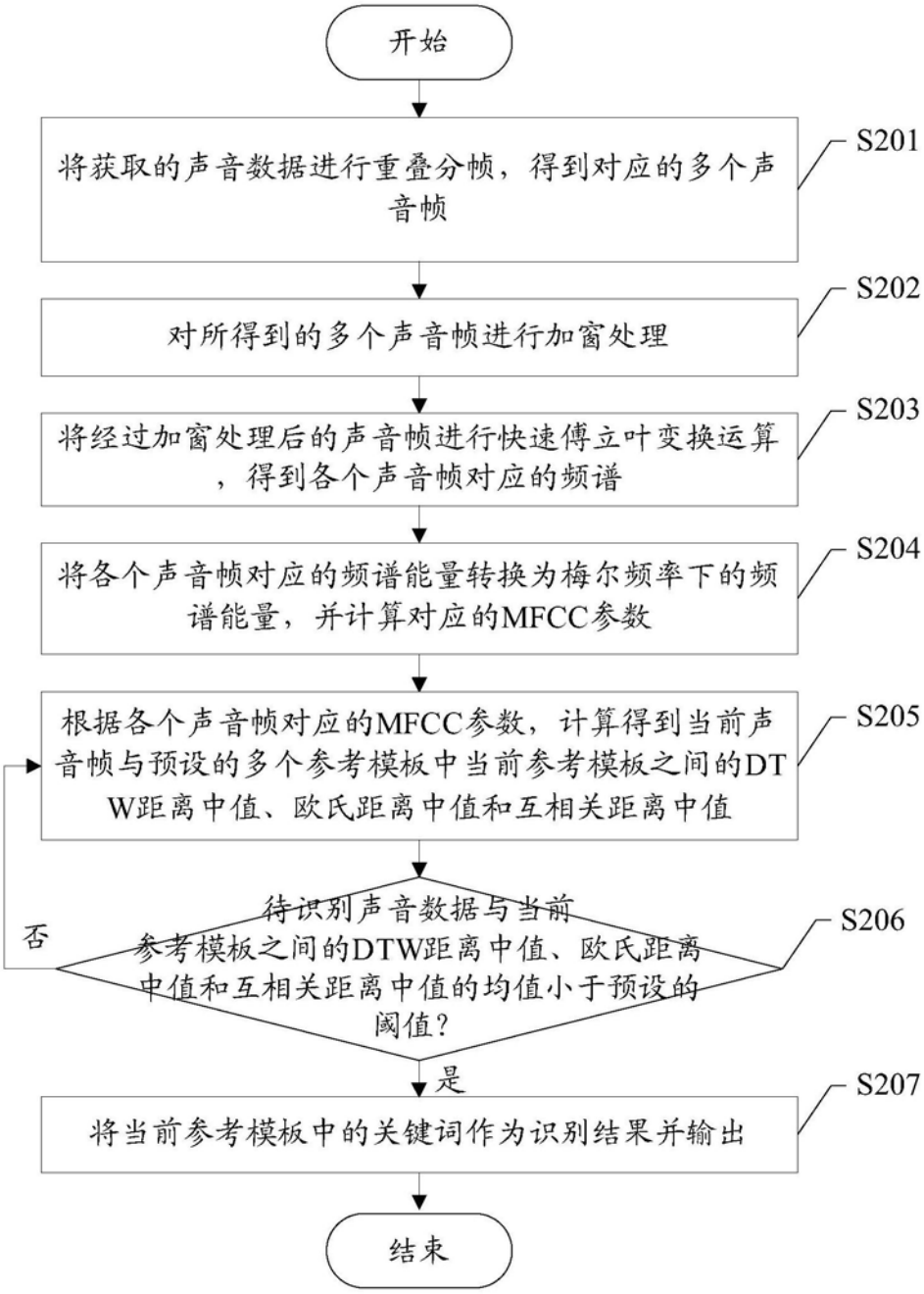

开始
将获取的声音数据进行重叠分帧，得到对应的多个声音帧
S201
对所得到的多个声音帧进行加窗处理
S202
将经过加窗处理后的声音帧进行快速傅立叶变换运算，得到各个声音帧对应的频谱
S203
将各个声音帧对应的频谱能量转换为梅尔频率下的频谱能量，并计算对应的MFCC参数
S204
根据各个声音帧对应的MFCC参数，计算得到当前声音帧与预设的多个参考模板中当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值
S205
待识别声音数据与当前参考模板之间的DTW距离中值、欧氏距离中值和互相关距离中值的均值小于预设的阈值？
S206
否
是
将当前参考模板中的关键词作为识别结果并输出
S207
结束

图2

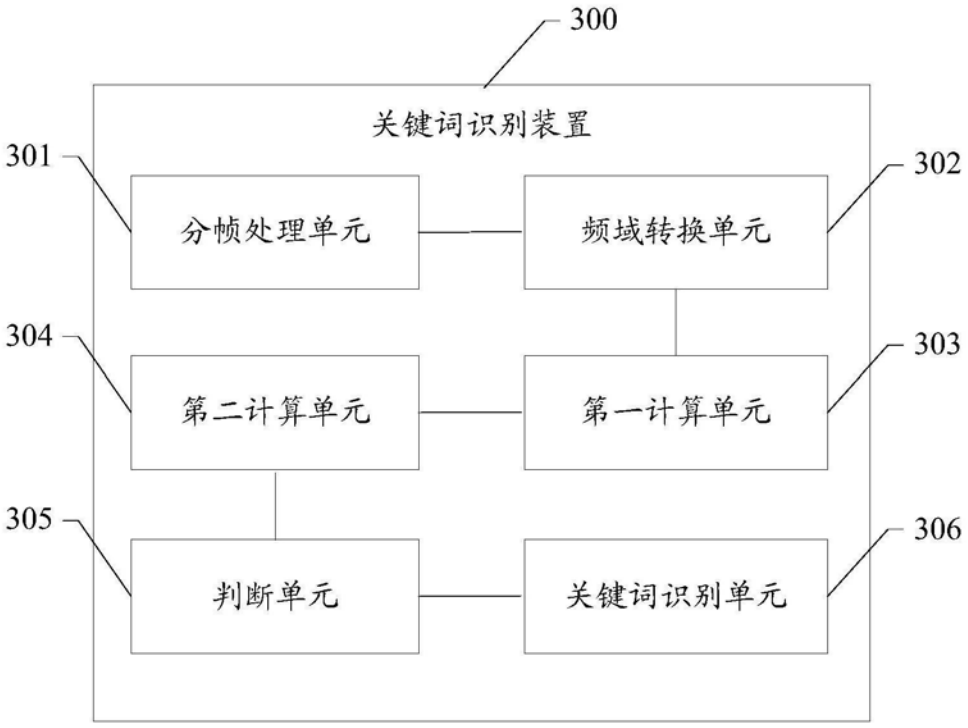

300
关键词识别装置
301
分帧处理单元
302
频域转换单元
304
第二计算单元
303
第一计算单元
305
判断单元
306
关键词识别单元

图3