(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109801646 A
(43)申请公布日 2019.05.24

(21)申请号 201910099804.4

(22)申请日 2019.01.31

(71)申请人 北京嘉楠捷思信息技术有限公司
地址 100094 北京市海淀区东北旺西路8号中关村软件园一期27号楼C座101号

(72)发明人 王彦 张楠赓

(74)专利代理机构 北京市中伦律师事务所 11410
代理人 杨黎峰 钟锦舜

(51)Int.Cl.
*G10L 25/87*(2013.01)
*G10L 25/03*(2013.01)

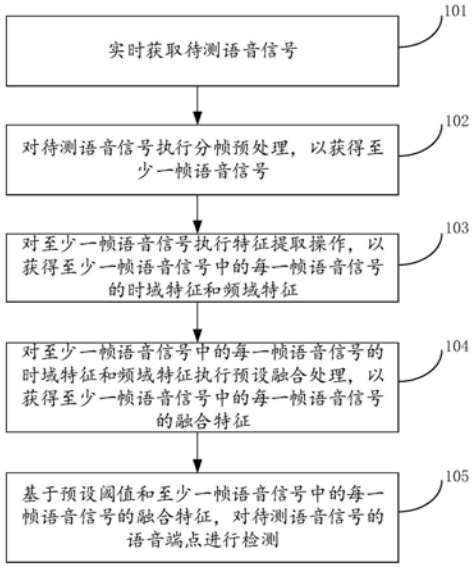
权利要求书3页 说明书14页 附图5页

(54)发明名称

一种基于融合特征的语音端点检测方法和装置

(57)摘要

本发明的实施方式提供了一种基于融合特征的语音端点检测方法和装置，方法包括：实时获取待测语音信号并执行分帧预处理，获得语音信号；对语音信号执行特征提取操作，获得每一帧语音信号的时域特征和频域特征；对每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得每一帧语音信号的融合特征；基于预设阈值和每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测。本发明还提供了对应的装置与计算机可读存储介质。通过上述检测方法基于较少的运算量，达到显著提高检测准确度的效果。




CN 109801646 A

1.一种基于融合特征的语音端点检测方法,其特征在于,所述方法包括:

实时获取待测语音信号;

对所述待测语音信号执行分帧预处理,以获得至少一帧语音信号;

对所述至少一帧语音信号执行特征提取操作,以获得所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征;

对所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理,以获得所述至少一帧语音信号中的每一帧语音信号的融合特征;

基于预设阈值和所述至少一帧语音信号中的每一帧语音信号的融合特征,对所述待测语音信号的语音端点进行检测。

2.根据权利要求1所述方法,其特征在于,所述时域特征至少包括能量特征和谱平坦度特征,所述频域特征至少包括频率特征。

3.根据权利要求1所述方法,其特征在于,基于预设阈值和所述至少一帧语音信号中的每一帧语音信号的融合特征,对所述待测语音信号的语音端点进行检测进一步包括:

对所述至少一帧语音信号中的每一帧语音信号逐帧进行判断;

若所述至少一帧语音信号中的任意一个语音信号的融合特征满足第一预设条件,则将所述任意一个语音信号作为第一语音信号划分到第一分类;

若所述至少一帧语音信号中的任意一个语音信号的融合特征未满足所述第一预设条件,则将所述任意一个语音信号作为第二语音信号划分到第二分类;

其中,所述第一预设条件基于所述预设阈值而形成。

4.根据权利要求3所述方法,其特征在于,在对所述至少一帧语音信号中的每一帧语音信号逐帧进行判断的过程中,所述方法进一步包括:

若所述至少一帧语音信号中,超过第一帧数的第一语音信号被连续划分到所述第一分类,则将所述超过第一帧数的第一语音信号的起点判定为第一语音端点;

若所述至少一帧语音信号中,在检测到所述第一语音端点之后,超过第二帧数的第二语音信号被连续划分到所述第二分类,则将所述超过第二帧数的第二语音信号的起点判定为第二语音端点。

5.根据权利要求3所述方法,其特征在于,若所述至少一帧语音信号中的任意一个语音信号的融合特征未满足所述第一预设条件,将所述任意一个语音信号作为第二语音信号划分到第二分类时,所述方法还包括:

根据所述第二语音信号的融合特征更新所述预设阈值;

根据所述预设阈值更新所述第一预设条件。

6.根据权利要求1所述的方法,其特征在于,所述方法还包括:

获取所述待测语音信号中的至少一个第三语音信号;

根据所述至少一个第三语音信号的能量特征均值与谱平坦度特征最小值,以获取所述待测语音信号的初始时域特征;

根据所述至少一个第三语音信号的频率特征最小值,以获取所述待测语音信号的初始频域特征;

根据所述待测语音信号的所述初始时域特征与所述初始频域特征,获取所述预设阈值。

7.根据权利要求2中所述的方法，其特征在于，所述至少一帧语音信号中的每帧待测语音信号的频率特征为所述至少一帧语音信号中的每帧待测语音信号的最高频率值。

8.根据权利要求1所述的方法，其特征在于，所述对所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理至少包括下列中的部分或全部：

根据预设的决策树模型，对所述至少一帧语音信号的时域特征和频域特征执行预设融合处理；

根据预设的权值参数，对所述至少一帧语音信号的时域特征和频域特征执行预设融合处理。

9.一种基于融合特征的语音端点检测装置，其特征在于，所述装置包括：

获取模块，用于实时获取待测语音信号；

分帧模块，用于对所述待测语音信号执行分帧预处理，以获得至少一帧语音信号；

提取模块，用于对所述至少一帧语音信号执行特征提取操作，以获得所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

融合模块，用于对所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得所述至少一帧语音信号中的每一帧语音信号的融合特征；

检测模块，用于基于预设阈值和所述至少一帧语音信号中的每一帧语音信号的融合特征，对所述待测语音信号的语音端点进行检测。

10.根据权利要求9所述装置，其特征在于，所述时域特征至少包括能量特征和谱平坦度特征，所述频域特征至少包括频率特征。

11.根据权利要求9所述装置，其特征在于，所述检测模块进一步包括：

对所述至少一帧语音信号中的每一帧语音信号逐帧进行判断；

若所述至少一帧语音信号中的任意一个语音信号的融合特征满足第一预设条件，则将所述任意一个语音信号作为第一语音信号划分到第一分类；

若所述至少一帧语音信号中的任意一个语音信号的融合特征未满足所述第一预设条件，则将所述任意一个语音信号作为第二语音信号划分到第二分类；

其中，所述第一预设条件基于所述预设阈值而形成。

12.根据权利要求11所述装置，其特征在于，在对所述至少一帧语音信号中的每一帧语音信号逐帧进行判断的过程中，所述检测模块进一步用于：

若所述至少一帧语音信号中，超过第一帧数的第一语音信号被连续划分到所述第一分类，则判定所述第一语音信号的起点为第一语音端点；

若所述至少一帧语音信号中，在检测到所述第一语音端点之后，超过第二帧数的第二语音信号被连续划分到所述第二分类，则判定所述第二语音信号的起点为第二语音端点。

13.根据权利要求11所述装置，其特征在于，所述检测模块进一步用于：

若所述至少一帧语音信号中的任意一个语音信号的融合特征未满足所述第一预设条件，将所述任意一个语音信号作为第二语音信号划分到第二分类时，根据所述第二语音信号的融合特征更新所述预设阈值；

根据所述预设阈值更新所述第一预设条件。

14.根据权利要9所述的装置，其特征在于，所述装置还包括阈值模块，具体用于：

获取所述待测语音信号中的至少一个第三语音信号；

根据所述至少一个第三语音信号的能量特征均值与谱平坦度特征最小值，以获取所述待测语音信号的初始时域特征；

根据所述至少一个第三语音信号的频率特征最小值，以获取所述待测语音信号的初始频域特征；

根据所述待测语音信号的所述初始时域特征与所述初始频域特征，获取所述预设阈值。

15.根据权利要求10中所述的装置，其特征在于，所述至少一帧语音信号中的每帧待测语音信号的频率特征为所述至少一帧语音信号中的每帧待测语音信号的最高频率值。

16.根据权利要求9所述的装置，其特征在于，融合模块进一步用于：

根据预设的决策树模型，对所述至少一帧语音信号的时域特征和频域特征执行预设融合处理；

根据预设的权值参数，对所述至少一帧语音信号的时域特征和频域特征执行预设融合处理。

17.一种基于融合特征的端点检测装置，其特征在于，包括：

一个或者多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或者多个处理器执行时，使得所述一个或多个处理器实现：

实时获取待测语音信号；

对所述待测语音信号执行分帧预处理，以获得至少一帧语音信号；

对所述至少一帧语音信号执行特征提取操作，以获得所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

对所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得所述至少一帧语音信号中的每一帧语音信号的融合特征；

基于预设阈值和所述至少一帧语音信号中的每一帧语音信号的融合特征，对所述待测语音信号的语音端点进行检测。

18.一种计算机可读存储介质，所述计算机可读存储介质存储有程序，当所述程序被处理器执行时，使得所述处理器执行如权利要求1-8中任一项所述的方法。

# 一种基于融合特征的语音端点检测方法和装置

## 技术领域

[0001] 本发明涉及语音识别领域，具体涉及一种基于融合特征的语音端点检测方法和装置。

## 背景技术

[0002] 本部分旨在为权利要求书中陈述的本发明的实施方式提供背景或上下文。此处的描述不因为包括在本部分中就承认是现有技术。

[0003] 近年来，随着人机信息交互技术的发展，语音识别技术显示出其重要性。在语音识别系统中，语音端点检测(Voice Activity Detection，简称VAD)是语音识别中的关键技术之一。语音端点检测是指在连续声音信号中找出语音部分的起始点和终止点。端点检测准确与否，会直接影响到语音识别系统的性能。如果端点切分出现错误，则会导致漏识别或者误识别等情况的发生，进而可导致语音识别结果不准确。

[0004] 传统的语音端点检测方法主要是获取时域或频域能量，并与给定的阈值进行比较，从而判断出语音的起始点和终止点。随着深度学习技术的快速发展，提出了多种基于模型的语音端点检测方法。

[0005] 然而，在实现本发明的过程中发明人发现上述语音端点检测算法至少存在以下问题：(1)传统的语音端点检测方法适用于平稳噪声，且高信噪比的环境，但在非平稳噪声、较低信噪比环境下，上述语音端点检测方法的检测效果不好，所检测的语音端点的准确率较低；(2)对于基于模型的语音端点检测方法，比如以GMM模型判定的VAD，由于数学模型的复杂性，实现端点检测的计算量很大，尤其是应用到嵌入式装置上时会造成较大的资源损耗。

## 发明内容

[0006] 为了解决上述实现端点检测的计算量很大，会造成较大的资源损耗的技术问题，本发明的实施例提出一种基于融合特征的语音端点检测方法和装置，可以在减少运算的同时保持检测准确度。

[0007] 在本发明实施方式的第一方面，提出一种基于融合特征的语音端点检测方法，其特征在于，方法包括：

[0008] 实时获取待测语音信号；

[0009] 对待测语音信号执行分帧预处理，以获得至少一帧语音信号；

[0010] 对至少一帧语音信号执行特征提取操作，以获得至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

[0011] 对至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得至少一帧语音信号中的每一帧语音信号的融合特征；

[0012] 基于预设阈值和至少一帧语音信号中的每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测。

[0013] 可选地，其中，时域特征至少包括能量特征和谱平坦度特征，频域特征至少包括频

率特征。

[0014] 可选地，其中，基于预设阈值和至少一帧语音信号中的每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测进一步包括：

[0015] 对至少一帧语音信号中的每一帧语音信号逐帧进行判断；

[0016] 若至少一帧语音信号中的任意一个语音信号的融合特征满足第一预设条件，则将所述任意一个语音信号作为第一语音信号划分到第一分类；

[0017] 若至少一帧语音信号中的任意一个语音信号的融合特征未满足第一预设条件，则将所述任意一个语音信号作为第二语音信号划分到第二分类；

[0018] 其中，第一预设条件基于预设阈值而形成。

[0019] 可选地，其中，在对至少一帧语音信号中的每一帧语音信号逐帧进行判断的过程中，方法进一步包括：

[0020] 若至少一帧语音信号中，超过第一帧数的第一语音信号被连续划分到第一分类，则判定第一语音信号的起点为第一语音端点；

[0021] 若至少一帧语音信号中，在检测到第一语音端点之后，超过第二帧数的第二语音信号被连续划分到第二分类，则判定第二语音信号的起点为第二语音端点。

[0022] 可选地，其中，若至少一帧语音信号中的任意一个语音信号的融合特征未满足第一预设条件，将所述任意一个语音信号作为第二语音信号划分到第二分类时，方法还包括：

[0023] 根据第二语音信号的融合特征更新预设阈值；

[0024] 根据预设阈值更新第一预设条件。

[0025] 可选地，其中，方法还包括：

[0026] 获取待测语音信号中的至少一个第三语音信号；

[0027] 根据至少一个第三语音信号的能量特征均值与谱平坦度特征最小值，以获取待测语音信号的初始时域特征；

[0028] 根据至少一个第三语音信号的频率特征最小值，以获取待测语音信号的初始频域特征；

[0029] 根据待测语音信号的初始时域特征与初始频域特征，获取预设阈值。

[0030] 可选地，其中，特征提取操作还包括快速傅里叶变换操作。

[0031] 可选地，其中，至少一帧语音信号中的每帧待测语音信号的频率特征为至少一帧语音信号中的每帧待测语音信号的最高频率值。

[0032] 可选地，其中，对至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理至少包括下列中的部分或全部：

[0033] 根据预设的决策树模型，对至少一帧语音信号的时域特征和频域特征执行预设融合处理；

[0034] 根据预设的权值参数，对至少一帧语音信号的时域特征和频域特征执行预设融合处理。

[0035] 本发明实施方式的第二方面，提出一种基于融合特征的语音端点检测装置，其特征在于，装置包括：

[0036] 获取模块，用于实时获取待测语音信号；

[0037] 分帧模块，用于对待测语音信号执行分帧预处理，以获得至少一帧语音信号；

[0038] 提取模块，用于对至少一帧语音信号执行特征提取操作，以获得至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

[0039] 融合模块，用于对至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得至少一帧语音信号中的每一帧语音信号的融合特征；

[0040] 检测模块，用于基于预设阈值和至少一帧语音信号中的每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测。

[0041] 可选地，其中，时域特征至少包括能量特征和谱平坦度特征，频域特征至少包括频率特征。

[0042] 可选地，其中，检测模块进一步包括：

[0043] 对至少一帧语音信号中的每一帧语音信号逐帧进行判断；

[0044] 若至少一帧语音信号中的任意一个语音信号的融合特征满足第一预设条件，则将所述任意一个语音信号作为第一语音信号划分到第一分类；

[0045] 若至少一帧语音信号中的任意一个语音信号的融合特征未满足第一预设条件，则将所述任意一个语音信号作为第二语音信号划分到第二分类；

[0046] 其中，第一预设条件基于预设阈值而形成。

[0047] 可选地，其中，在对至少一帧语音信号中的每一帧语音信号逐帧进行判断的过程中，检测模块进一步用于：

[0048] 若至少一帧语音信号中，超过第一帧数的第一语音信号被连续划分到第一分类，则判定第一语音信号的起点为第一语音端点；

[0049] 若至少一帧语音信号中，在检测到第一语音端点之后，超过第二帧数的第二语音信号被连续划分到第二分类，则判定第二语音信号的起点为第二语音端点。

[0050] 可选地，其中，检测模块进一步用于：

[0051] 在至少一帧语音信号中的任意一个语音信号的融合特征未满足第一预设条件，将所述任意一个语音信号作为第二语音信号划分到第二分类时，根据所述第二语音信号的融合特征更新预设阈值；

[0052] 根据预设阈值更新第一预设条件。

[0053] 可选地，其中，装置还包括阈值模块，具体用于：

[0054] 获取待测语音信号中的至少一个第三语音信号；

[0055] 根据至少一个第三语音信号的能量特征均值与谱平坦度特征最小值，以获取待测语音信号的初始时域特征；

[0056] 根据至少一个第三语音信号的频率特征最小值，以获取待测语音信号的初始频域特征；

[0057] 根据待测语音信号的初始时域特征与初始频域特征，获取预设阈值。

[0058] 可选地，其中，特征模块进一步用于执行快速傅里叶变换操作。

[0059] 可选地，其中，至少一帧语音信号中的每帧待测语音信号的频率特征为至少一帧语音信号中的每帧待测语音信号的最高频率值。

[0060] 可选地，其中，融合模块进一步用于：

[0061] 根据预设的决策树模型，对至少一帧语音信号的时域特征和频域特征执行预设融合处理；

[0062] 根据预设的权值参数，对至少一帧语音信号的时域特征和频域特征执行预设融合处理。

[0063] 本发明实施方式的第三方面，提出一种基于融合特征的端点检测装置，其特征在于，包括：

[0064] 一个或者多个处理器；

[0065] 存储器，用于存储一个或多个程序；

[0066] 当一个或多个程序被一个或者多个处理器执行时，使得一个或多个处理器实现：

[0067] 实时获取待测语音信号；

[0068] 对待测语音信号执行分帧预处理，以获得至少一帧语音信号；

[0069] 对至少一帧语音信号执行特征提取操作，以获得至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

[0070] 对至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得至少一帧语音信号中的每一帧语音信号的融合特征；

[0071] 基于预设阈值和至少一帧语音信号中的每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测。

[0072] 本发明实施方式的第四方面，提出一种计算机可读存储介质，计算机可读存储介质存储有程序，当程序被处理器执行时，使得处理器执行如上的方法。

[0073] 有益效果：正是通过本发明实施方式所提供的提出一种基于融合特征的语音端点检测方法、装置以及存储介质，可以在减少运算的同时保持检测准确度，尤其在应用到嵌入式装置时，进一步显著提高了生产效率并降低了生产成本。

## 附图说明

[0074] 通过参考附图阅读下文的详细描述，本发明示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中，以示例性而非限制性的方式示出了本发明的若干实施方式，其中：

[0075] 图1示出了根据本发明实施例的一种基于融合特征的语音端点检测方法流程图；

[0076] 图2示出了根据本发明实施例的一种决策树示意图；

[0077] 图3示出了根据本发明实施例的另一种基于融合特征的语音端点检测方法流程图；

[0078] 图4示出了根据本发明实施例的又一种基于融合特征的语音端点检测方法流程图；

[0079] 图5示出了根据本发明实施例的一种基于融合特征的语音端点检测装置示意图；

[0080] 图6示出了根据本发明实施例的另一种基于融合特征的语音端点检测的装置示意图；

[0081] 图7示出了根据本发明实施方式的基于融合特征的语音端点检测的装置的计算机可读存储介质的示意图。

[0082] 在附图中，相同或对应的标号表示相同或对应的部分。

## 具体实施方式

[0083] 下面将参考若干示例性实施方式来描述本发明的原理和精神。应当理解，给出这些实施方式仅仅是为了使本领域技术人员能够更好地理解进而实现本发明，而并非以任何方式限制本发明的范围。相反，提供这些实施方式是为了使本公开更加透彻和完整，并且能够将本公开的范围完整地传达给本领域的技术人员。

[0084] 示例性方法

[0085] 本发明实施例提出一种基于融合特征的语音端点检测方法。

[0086] 图1是根据本发明实施例的基于融合特征的语音端点检测的方法的示意性流程图。如图1所示，包括但不限于如下步骤：

[0087] 101、实时获取待测语音信号；

[0088] 102、对待测语音信号执行分帧预处理，以获得至少一帧语音信号；

[0089] 103、对至少一帧语音信号执行特征提取操作，以获得至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

[0090] 可选地，可以在执行特征提取操作之前对至少一帧语音信号进行短时傅里叶变换，从而将时域信号转换为频域信号。

[0091] 104、对至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得至少一帧语音信号中的每一帧语音信号的融合特征；

[0092] 具体地，预设融合处理可以是将多个特征信息作为融合特征的一个分量信息进行组合。例如，将多个特征信息a、b进过预设融合处理后，获得一个多维数组形式的融合特征(a,b)，特征信息与多维数组中的某一元素相对应。可选地，预设融合处理也可以是直接将多个特征信息的数值通过运算直接转化为标量数值形式的融合特征。例如，将多个特征信息a、b进过预设融合处理后，获得一个标量数值形式的融合特征a+b。

[0093] 105、基于预设阈值和至少一帧语音信号中的每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测。

[0094] 具体地，可以将融合特征与该预设阈值进行比较后，并将该比较结果作为语音端点检测的判断基础；该预设阈值可以基于经验值获得、或基于获取的待测语音信号中的底噪信号获得。

[0095] 举例来说，可以依照上述步骤101和步骤102，在获取待检测的语音信号后，以20ms为一帧语音信号的长度，并基于此进行分帧，获得至少一帧语音信号。进一步依照上述步骤103，针对至少一帧语音信号中的每一帧语音信号x[n]，依次提取时域特征a和频域特征b，比如能量特征、信号过零率、SFM特征以及频率特征等。进一步依照上述步骤104，将多个特征信息a、b经过预设融合处理后，获得一个多维数组形式的融合特征(a,b)；进一步依照步骤105，根据融合特征(a,b)对至少一帧语音信号中的每一帧语音信号进行判别，比如x[n-2]、x[n-1]、x[n]不超过该预设阈值、x[n+1]、x[n+2]、x[n+3]、x[n+4]超过该预设阈值，进而可以推测x[n-2]、x[n-1]、x[n]为底噪、可以推测x[n+1]、x[n+2]、x[n+3]、x[n+4]为有效语音，进一步可以检测得知x[n+1]可以是一段有效语音的起始端点，也即找到至少一帧语音信号中的语音端点。

[0096] 本实施例中将每帧语音信号的多个时域特征与频域特征融合在一起，形成融合特征，并基于该融合特征对每帧语音信号进行分析，从而能够将至少一帧语音信号区分为有

效语音或底噪，进而找到至少一帧语音信号中的语音端点。相较于采用单一的过零率、短时能量等特征进行端点检测，本实施例通过采用多个时域特征与频域特征的融合特征，能够在资源耗费量较小的情况下获得更加准确的检测结果。

[0097] 进一步地，在一实施例中，时域特征至少包括能量特征和谱平坦度特征，频域特征至少包括频率特征。以下结合示例对上述时域特征与频域特征进行详细说明。

[0098] 具体地，采用频率特征作为频域特征进行语音端点检测是因为有效语音的频率相较于底噪的频率要更高。因此，频率特征可以有效区分语音和底噪。

[0099] 具体地，采用能量特征作为时域特征是因为有效语音和底噪的区别可以体现在它们的能量上，有效语音的能量比底噪能量大。具体地，上述能量特征既可以直接从每一帧语音信号的时域信号进行运算提取，也可以从音频帧的频域信号中运算提取，现有技术中提取能量特征的技术方案以及较为完善，在此不再赘述。

[0100] 具体地，利用谱平坦度(SFM)特征作为时域特征进行语音端点检测是因为SFM特征是对功率谱的分布情况的度量，一般较高的谱平坦度说明语音谱在所有的频带上都有差不多的功率，频谱图相对比较平坦，例如白噪音。较低的谱平度则说明在频域上的功率谱的分布是不均匀的。因此，通常来说有效语音的谱平度很低。所以，谱平度能够有效地区分有效语音和底噪。

[0101] 进一步地，在本发明实施例中，提取SFM特征信息具体包括：首先从每一帧语音信号中获取几何平均值以及算数平均值，然后采用下列公式计算出每一帧语音信号的SFM特征信息：$SFM = |10\log\left(\frac{G_m}{A_m}\right)|$，其中，$G_m$为几何平均值，$A_m$为算数平均值。

[0102] 进一步地，在一实施例中，至少一帧语音信号中的每帧待测语音信号的频率特征为至少一帧语音信号中的每帧待测语音信号的最高频率值。

[0103] 本发明实施例中，提取上述频率特征具体可以包括：

[0104] 将每帧语音信号经过快速傅里叶变换转换为频域信号；

[0105] 采用下列公式选取每帧语音信号在频域内最高的频率值作为该帧语音信号的频率特征信息。

[0106] F[i]＝argmax(S[n])，

[0107] 可选地，上述特征信息还可以包括：每一帧语音信号在时域的信号过零率、每一帧语音信号在时域的信号信息熵，本实施例仅以上述特征为例进行描述，但不限于此。

[0108] 进一步地，在一实施例中，基于图1中所示出的语音端点检测方法，步骤104可以进一步包括：

[0109] 根据预设的决策树模型，对至少一帧语音信号的时域特征和频域特征执行预设融合处理；和/或

[0110] 根据预设的权值参数，对至少一帧语音信号的时域特征和频域特征执行预设融合处理。

[0111] 具体地，以下结合图2对上述根据预设的决策树模型，对至少一帧语音信号的时域特征和频域特征执行预设融合处理进行示例性描述：

[0112] 例如，如图2所示，将不同的特征信息设置为决策树中的不同节点，上述融合特征为多个特征信息的组合，其中不同的特征信息分量具有不同的优先级，若一个特征信息的

优先级越高，其位于越顶端的决策树节点。

[0113] A节点为SFM特征、B节点为频率特征、C节点为能量特征，从预设阈值中可以拆分提取出多个阈值分量，不同的特征信息对应于不同的阈值分量。

[0114] (1)若语音信号X的A特征未超过第一阈值分量，则直接判断该语音信号X为非语音帧；

[0115] (2)若语音信号X的A特征超过第一阈值分量，则进一步对B特征进行判断；

[0116] (3)若语音信号X的B特征未超过第二阈值分量的第一值，则直接判断语音信号X为非语音帧；

[0117] (4)若语音信号X的B特征在第二阈值分量的第一值与第二值之间，则进一步对C特征判断；

[0118] (5)若语音信号X的B特征超过第二阈值分量第二值，则直接判断语音信号X为语音帧；

[0119] (6)若语音信号X的C特征超过第三阈值分量，则直接判断语音信号X为语音帧；

[0120] (7)若语音信号X的C特征未超过第三阈值分量，则直接判断语音信号X为非语音帧。

[0121] 其中，如节点B—C所示，当对一个特征进行判断时，其所采用的阈值分量可以采用一个阈值，从而可以输出两种判断结果中的一种结果，也可以采用多个阈值，从而输出多种判断结果中的一种结果。

[0122] 可以理解的是，采用决策树模型进行特征融合操作，可以对多个特征的优先级进行区分，能控制每个特征的可容纳范围，提升检测准确度。本实施例可以根据需求设计出不同的决策树模型并进行判断，本实施例仅以上述列出的决策树模型为例进行描述，但不限于此。

[0123] 具体地，以下结合具体示例对上述根据预设的权值参数，对至少一帧语音信号的时域特征和频域特征执行预设融合处理进行示例性描述：

[0124] 其中，可以根据不同的权值参数与各个特征信息进行加权累加运算。相应地，预设阈值也可以拆分提取出与上述特征信息相对应的第一阈值分量、第二阈值分量与第三阈值分量。

[0125] 例如，预先定义一个初始值为0的count值，并进行如下运算：

[0126] 若语音信号X的SFM特征超过第一阈值分量，则count+2；

[0127] 若语音信号X的频率特征超过第二阈值分量，则count+1；

[0128] 若语音信号X的能量特征超过第三阈值分量，则count+1；

[0129] 最终，若count≥3，则判断语音信号X为有效语音，其中，上述多个特征信息对应不同的权值。可选地，上述不同特征信息也可以选择相同的加权参数。

[0130] 可以理解的是，采用权值参数进行特征融合操作所需求的计算量极少。本实施例可以根据需求设计出不同的权值方案，本实施例仅以上述示例为例进行描述，但不限于此。

[0131] 进一步地，在一实施例中，参见图3，示出了根据本发明实施例的又一种基于融合特征的语音端点检测方法流程图，结合图1，其中基于预设阈值和至少一帧语音信号中的每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测进一步包括：

[0132] 301、对至少一帧语音信号中的每一帧语音信号逐帧进行判断；

[0133] 302、判断语音信号a(至少一帧语音信号中的任意一个语音信号)的融合特征是否满足第一预设条件；

[0134] 303、若满足，则将该语音信号a作为第一语音信号划分到第一分类；

[0135] 304、若未满足，则将该语音信号a作为第二语音信号划分到第二分类。

[0136] 具体地，第一预设条件基于预设阈值而形成，第一预设条件可以是融合特征超过预设阈值或小于预设阈值。本实施例可以将根据融合特征的值将语音信号分为有效语音类别与底噪类别，其中第一分类也即有效语音类别、第二分类也即底噪类别。

[0137] 例如，预先定义一个初始值为0的count值作为语音信号的融合特征，并进行如下运算：若语音信号a的SFM特征超过第一阈值分量，则count+2；若语音信号a的频率特征超过第二阈值分量，则count+1；若语音信号a的能量特征超过第三阈值分量，则count+1。

[0138] 进一步地，获取第n帧、第n+1帧、第n+2帧、第n+3帧、…、第n+k帧语音信号，并对其逐帧地进行判断，利用上述方式具体计算出对应于每一帧语音信号的融合特征count，假设第一预设条件为count值≥3；其中，若第n+1帧语音信号、第n+2帧语音信号的count值≥3，则可以将第n+1帧语音信号、第n+2帧语音信号作为第一语音信号划分到第一分类，也即划分到有效语音类别中；若第n帧语音信号的count值＜3，则可以将第n帧语音信号作为第二语音信号划分到第二分类，也即划分到底噪类别。通过将每一帧语音信号进行分类，可以更加直观地观察出语音信号的特性。

[0139] 进一步地，在一实施例中，在对至少一帧语音信号中的每一帧语音信号逐帧进行判断的过程中，方法进一步包括：

[0140] 若至少一帧语音信号中，超过第一帧数的第一语音信号被连续划分到第一分类，则将超过第一帧数的第一语音信号的起点判定为第一语音端点；

[0141] 具体地，该第一分类为有效语音的分类，第一语音端点为有效语音的起始端点，第一帧数至少为一帧。

[0142] 例如：假设第一帧数为k，若第n-1帧语音信号被划分到第二分类，且第n帧语音信号、第n+1帧语音信号、…、第n+k帧语音信号均被连续划分到第一分类，可以看出，从第n帧到第n+k帧，被连续划分到第一分类的第一语音信号超过k帧，且该超过第一帧数k的第一语音信号的起点为第n帧语音信号，则可以将第n帧语音信号判定为第一语音端点，也即有效语音段的起始端点。

[0143] 若至少一帧语音信号中，在检测到第一语音端点之后，超过第二帧数的第二语音信号被连续划分到第二分类，则将超过第二帧数的第二语音信号的起点判定为第二语音端点。

[0144] 例如：假设第二帧数为b，若第m-1帧语音信号被划分到第一分类，且第m帧语音信号、第m+1帧语音信号、…、第m+b帧语音信号均被连续划分到第二分类，可以看出，从第m帧到第m+b帧，被连续划分到第二分类的第二语音信号超过b帧，且该超过第二帧数b的第二语音信号的起点为第m帧语音信号，则可以将第m帧语音信号判定为第二语音端点，也即有效语音段的末尾端点。

[0145] 换言之，上述方法也可以理解为，若被连续划分到第一分类(有效语音类别)的语音信号的帧数小于第一帧数，则可以将其忽略，不将其识别为有效语音段的起点，也即语音起始端点。若被连续划分到第二分类(底噪类别)的语音信号的帧数小于第二帧数，则可以

将其忽略，不将其识别为有效语音段的终点，也即语音末尾端点。例如，可以忽略小于连续140ms的非语音帧，以及忽略小于连续100ms的语音帧。利用以上技术方案可以避免了语音端点的误识别，有利于提升端点检测准确度。

[0146] 进一步地，在一实施例中，若至少一帧语音信号中的任意一个语音信号b的融合特征未满足第一预设条件，进而将任意一个语音信号b作为第二语音信号划分到第二分类时，方法还可以包括：根据该语音信号b的融合特征更新预设阈值；根据预设阈值更新第一预设条件。

[0147] 具体地，由于实时获取的语音信号中，底噪可能随着时间变化发生一定改变，而当第二语音信号的融合特征未满足第一预设条件，可以判断第二语音信号为底噪信号，也就是可以根据作为底噪信号的第二语音信号去更新预设阈值。因此，本实施例中可以采取根据第二语音信号的融合特征更新预设阈值，这样有利于提高检测的准确性。

[0148] 进一步地，在一实施例中，参见图4，示出了根据本发明实施例的又一种基于融合特征的语音端点检测方法流程图，结合图1，方法还包括：

[0149] 401、获取待测语音信号中的至少一个第三语音信号；

[0150] 402、根据至少一个第三语音信号的能量特征均值与谱平坦度特征最小值，以获取待测语音信号的初始时域特征；

[0151] 403、根据至少一个第三语音信号的频率特征最小值，以获取待测语音信号的初始频域特征；

[0152] 404、根据待测语音信号的初始时域特征与初始频域特征，获取预设阈值。

[0153] 以下结合具体示例对本实施例进行具体描述。

[0154] 例如，提取待测语音信号的前50帧，并将其默认为非语音帧，并根据该待测语音信号的前50帧语音信号获取预设阈值，具体包括：

[0155] (1)根据下列公式计算出前50帧语音信号的平均能量$E_m$：

[0156] $$Em = \frac{1}{50}\sum_{1}^{50}\sum_{1}^{N} x[n]^2$$

[0157] 其中，x[n]指代前50帧语音信号中的每一帧语音信号的语流，N指代窗口长度。

[0158] (2)根据下列公式计算前50帧语音信号中最小的SFM值：

[0159] 首先，采用以下公式计算出每帧语音信号的SFM值：

[0160] $$\mathrm{SFM} = |10\,log\ (\frac{G_m}{A_m})\ |,$$

[0161] 其中，$G_m$为几何平均值，$A_m$为算数平均值。

[0162] 然后，选取前50帧语音信号中最小的SFM值：

[0163] $SFM_{min}$＝min(SFM)

[0164] (3)根据下列公式计算前50帧语音信号中最小的频率值：

[0165] 首先，将时域信号转化为频域信息：

[0166] S[n]＝FFT(x[n])

[0167] 其次，选择频谱幅度最大的频率值作为该帧主要频率：

[0168] F[i]＝argmax(S[n])

[0169] 最后，选取前50帧语音信号中最小的主要频率作为初始频域特征：

[0170] $F_{min}=min(F[i])$

[0171] 具体地，上述方式是通过假定的方式预设前50帧语音信号为非语音帧，然而在实际应用场景中，上述假定的非语音帧区间与实际的非语音帧区间并不完全相符，可能存在语音信号进入检测系统的时间点早于50帧的情况，仅一步容易导致初始阈值产生较大的偏差。

[0172] 在此情况下，由于SFM特征信息与频率特征信息相较于能量特征信息更为敏感。因此，本实施例通过选取前50帧信号的SFM值与频率值的最小值作为初始阈值的参数信息，能够有效提高语音端点检测的准确率。

[0173] 综上，本发明实施例提供的基于融合特征的语音端点检测方法通过将多个特征信息进行融合，再根据融合信息进行语音端点检测这一技术方案，在显著提高了检测准确度的同时保持了较小的运算量，显著提高了检测效率、并降低了检测成本。

[0174] 示例性装置

[0175] 本发明实施例提出一种基于融合特征的语音端点检测的装置。

[0176] 图5是根据本发明实施例的基于融合特征的语音端点检测的装置的示意性结构图。如图5所示，包括但不限于：

[0177] 获取模块501，用于实时获取待测语音信号；

[0178] 分帧模块502，用于对所述待测语音信号执行分帧预处理，以获得至少一帧语音信号；

[0179] 提取模块503，用于对所述至少一帧语音信号执行特征提取操作，以获得所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

[0180] 可选地，可以在执行特征提取操作之前对所述至少一帧语音信号进行短时傅里叶变换，从而将时域信号转换为频域信号。

[0181] 融合模块504，用于对所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得所述至少一帧语音信号中的每一帧语音信号的融合特征；

[0182] 具体地，预设融合处理可以是将多个特征信息作为融合特征的一个分量信息进行组合。例如，将多个特征信息a、b进过预设融合处理后，获得一个多维数组形式的融合特征(a,b)，特征信息与多维数组中的某一元素相对应。可选地，预设融合处理也可以是直接将多个特征信息的数值通过运算直接转化为标量数值形式的融合特征。例如，将多个特征信息a、b进过预设融合处理后，获得一个标量数值形式的融合特征a+b。

[0183] 检测模块505，用于基于预设阈值和所述至少一帧语音信号中的每一帧语音信号的融合特征，对所述待测语音信号的语音端点进行检测。

[0184] 具体地，可以将融合特征与该预设阈值进行比较后，并将该比较结果作为语音端点检测的判断基础；该预设阈值可以基于经验值获得、或基于获取的待测语音信号中的底噪信号获得。

[0185] 举例来说，可以利用上述获取模块501和分帧模块502，，在获取待检测的语音信号后，以20ms为一帧语音信号的长度，并基于此进行分帧，获得至少一帧语音信号。进一步利用上述提取模块503，针对至少一帧语音信号中的每一帧语音信号x[n]，依次提取时域特征a和频域特征b，比如能量特征、信号过零率、SFM特征以及频率特征等。进一步利用上述融合

模块504，，将多个特征信息a、b经过预设融合处理后，获得一个多维数组形式的融合特征(a,b)；进一步利用检测模块505，根据融合特征(a,b)对至少一帧语音信号中的每一帧语音信号进行判别，比如x[n-2]、x[n-1]、x[n]不超过该预设阈值、x[n+1]、x[n+2]、x[n+3]、x[n+4]超过该预设阈值，进而可以推测x[n-2]、x[n-1]、x[n]为底噪、可以推测x[n+1]、x[n+2]、x[n+3]、x[n+4]为有效语音，进一步可以检测得知x[n+1]可以是一段有效语音的起始端点，也即找到至少一帧语音信号中的语音端点。

[0186] 本实施例中将每帧语音信号的多个时域特征与频域特征融合在一起，形成融合特征，并基于该融合特征对每帧语音信号进行分析，从而能够将至少一帧语音信号区分为有效语音或底噪，进而找到至少一帧语音信号中的语音端点。相较于采用单一的过零率、短时能量等特征进行端点检测，本实施例通过采用多个时域特征与频域特征的融合特征，能够在资源耗费量较小的情况下获得更加准确的检测结果。

[0187] 进一步地，在一实施例中，时域特征至少包括能量特征和谱平坦度特征，频域特征至少包括频率特征。

[0188] 进一步地，在一实施例中，检测模块进一步包括：

[0189] 对至少一帧语音信号中的每一帧语音信号逐帧进行判断；

[0190] 若至少一帧语音信号中的任意一个语音信号的融合特征满足第一预设条件，则将任意一个语音信号作为第一语音信号划分到第一分类；

[0191] 若至少一帧语音信号中的任意一个语音信号的融合特征未满足第一预设条件，则将任意一个语音信号作为第二语音信号划分到第二分类；

[0192] 其中，第一预设条件基于预设阈值而形成。

[0193] 进一步地，在一实施例中，在对至少一帧语音信号中的每一帧语音信号逐帧进行判断的过程中，检测模块进一步用于：

[0194] 若至少一帧语音信号中，超过第一帧数的第一语音信号被连续划分到第一分类，则判定第一语音信号的起点为第一语音端点；

[0195] 若至少一帧语音信号中，在检测到第一语音端点之后，超过第二帧数的第二语音信号被连续划分到第二分类，则判定第二语音信号的起点为第二语音端点。

[0196] 进一步地，在一实施例中，检测模块进一步用于：

[0197] 若至少一帧语音信号中的任意一个语音信号的融合特征未满足第一预设条件，进而将任意一个语音信号作为第二语音信号划分到第二分类时，根据第二语音信号的融合特征更新预设阈值；

[0198] 根据预设阈值更新第一预设条件。

[0199] 进一步地，在一实施例中，装置还包括阈值模块，具体用于：

[0200] 获取待测语音信号中的至少一个第三语音信号；

[0201] 根据至少一个第三语音信号的能量特征均值与谱平坦度特征最小值，以获取待测语音信号的初始时域特征；

[0202] 根据至少一个第三语音信号的频率特征最小值，以获取待测语音信号的初始频域特征；

[0203] 根据待测语音信号的初始时域特征与初始频域特征，获取预设阈值。

[0204] 进一步地，在一实施例中，特征模块进一步用于执行快速傅里叶变换操作。

[0205] 进一步地，在一实施例中，至少一帧语音信号中的每帧待测语音信号的频率特征为至少一帧语音信号中的每帧待测语音信号的最高频率值。

[0206] 进一步地，在一实施例中，融合模块进一步用于：

[0207] 根据预设的决策树模型，对至少一帧语音信号的时域特征和频域特征执行预设融合处理；

[0208] 根据预设的权值参数，对至少一帧语音信号的时域特征和频域特征执行预设融合处理。

[0209] 综上，本发明实施例提供的基于融合特征的语音端点检测方法通过将多个特征信息进行融合，再根据融合信息进行语音端点检测这一技术方案，在显著提高了检测准确度的同时保持了较小的运算量，显著提高了检测效率、并降低了检测成本。

[0210] 示例性装置

[0211] 在介绍了本发明示例性实施方式的方法和装置之后，接下来，介绍根据本发明的另一方面的基于融合特征的语音端点检测装置。

[0212] 所属技术领域的技术人员能够理解，本发明的各个方面可以实现为装置、方法或计算机可读存储介质。因此，本发明的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式(包括固件、微代码等)，或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“装置”。

[0213] 在一些可能的实施方式中，本发明的基于融合特征的语音端点检测装置可以至少包括一个或多个处理器、以及至少一个存储器。其中，所述存储器存储有程序，当所述程序被所述处理器执行时，使得所述处理器执行如图1所示的步骤：

[0214] 101、实时获取待测语音信号；

[0215] 102、对所述待测语音信号执行分帧预处理，以获得至少一帧语音信号；

[0216] 103、对所述至少一帧语音信号执行特征提取操作，以获得所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征；

[0217] 104、对所述至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得所述至少一帧语音信号中的每一帧语音信号的融合特征；

[0218] 105、基于预设阈值和所述至少一帧语音信号中的每一帧语音信号的融合特征，对所述待测语音信号的语音端点进行检测。

[0219] 此外，尽管附图中未示出，但本发明的所述程序被所述处理器执行时，还使得所述处理器执行上述示例性方法中描述的其他操作或步骤。

[0220] 下面参照图6来描述根据本发明的这种实施方式的基于融合特征的语音端点检测装置1。图6显示的装置1仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

[0221] 如图6所示，装置1可以以通用计算装置的形式表现，包括但不限于：至少一个处理器10、至少一个存储器20、连接不同装置组件的总线60。

[0222] 总线60包括数据总线、地址总线和控制总线。

[0223] 存储器20可以包括易失性存储器，例如随机存取存储器(RAM)21和/或高速缓存存储器22，还可以进一步包括只读存储器(ROM)23。

[0224] 存储器20还可以包括程序模块24，这样的程序模块24包括但不限于：操作装置、一

个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。

[0225] 装置1还可以与一个或多个外部装置2(例如键盘、指向装置、蓝牙装置等)通信，也可与一个或者多个其他装置进行通信。这种通信可以通过输入/输出(I/O)接口40进行，并在显示单元30上进行显示。并且，装置1还可以通过网络适配器50与一个或者多个网络(例如局域网(LAN)，广域网(WAN)和/或公共网络，例如因特网)通信。如图所示，网络适配器50通过总线60与装置1中的其它模块通信。应当明白，尽管图中未示出，但可以结合装置1使用其它硬件和/或软件模块，包括但不限于：微代码、装置驱动器、冗余处理单元、外部磁盘驱动阵列、RAID装置、磁带驱动器以及数据备份存储装置等。

[0226] 示例性计算机可读存储介质

[0227] 在一些可能的实施方式中，本发明的各个方面还可以实现为一种计算机可读存储介质的形式，其包括程序代码，当所述程序代码在被处理器执行时，所述程序代码用于使所述处理器执行上面描述的方法。

[0228] 上面描述的方法包括了上面的附图中示出和未示出的多个操作和步骤，这里将不再赘述。

[0229] 所述计算机可读存储介质可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的装置、装置或器件，或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括：具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0230] 如图7所示，描述了根据本发明的实施方式的计算机可读存储介质70，其可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码，并可以在终端装置，例如个人电脑上运行。然而，本发明的计算机可读存储介质不限于此，在本文件中，可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行装置、装置或者器件使用或者与其结合使用。

[0231] 可以以一种或多种程序设计语言的任意组合来编写用于执行本发明操作的程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如Java、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算装置上执行、部分地在用户装置上执行部分在远程计算装置上执行、或者完全在远程计算装置或服务器上执行。在涉及远程计算装置的情形中，远程计算装置可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算装置，或者，可以连接到外部计算装置(例如利用因特网服务提供商来通过因特网连接)。

[0232] 此外，尽管在附图中以特定顺序描述了本发明方法的操作，但是，这并非要求或者暗示必须按照该特定顺序来执行这些操作，或是必须执行全部所示的操作才能实现期望的结果。附加地或备选地，可以省略某些步骤，将多个步骤合并为一个步骤执行，和/或将一个步骤分解为多个步骤执行。

[0233] 虽然已经参考若干具体实施方式描述了本发明的精神和原理，但是应该理解，本发明并不限于所公开的具体实施方式，对各方面的划分也不意味着这些方面中的特征不能

组合以进行受益,这种划分仅是为了表述的方便。本发明旨在涵盖所附权利要求的精神和范围内所包括的各种修改和等同布置。

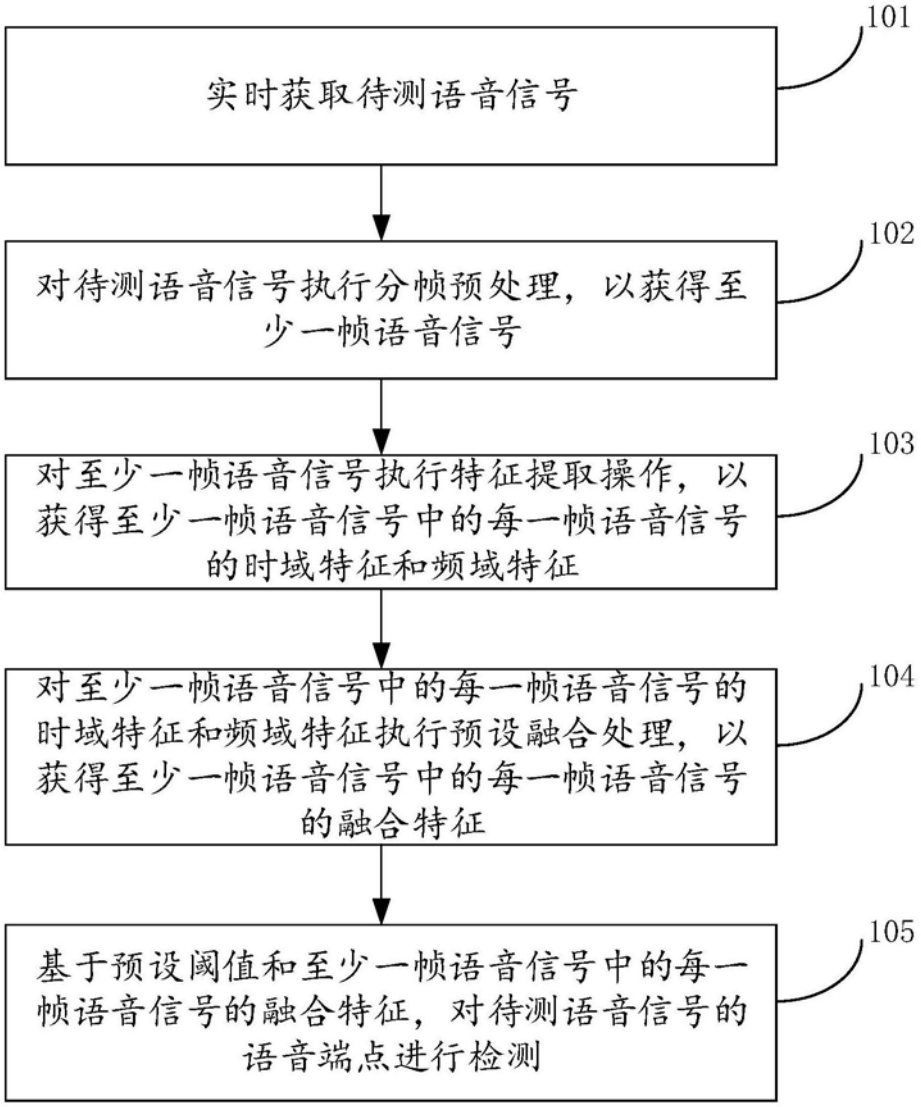

101
实时获取待测语音信号
102
对待测语音信号执行分帧预处理，以获得至少一帧语音信号
103
对至少一帧语音信号执行特征提取操作，以获得至少一帧语音信号中的每一帧语音信号的时域特征和频域特征
104
对至少一帧语音信号中的每一帧语音信号的时域特征和频域特征执行预设融合处理，以获得至少一帧语音信号中的每一帧语音信号的融合特征
105
基于预设阈值和至少一帧语音信号中的每一帧语音信号的融合特征，对待测语音信号的语音端点进行检测

图1

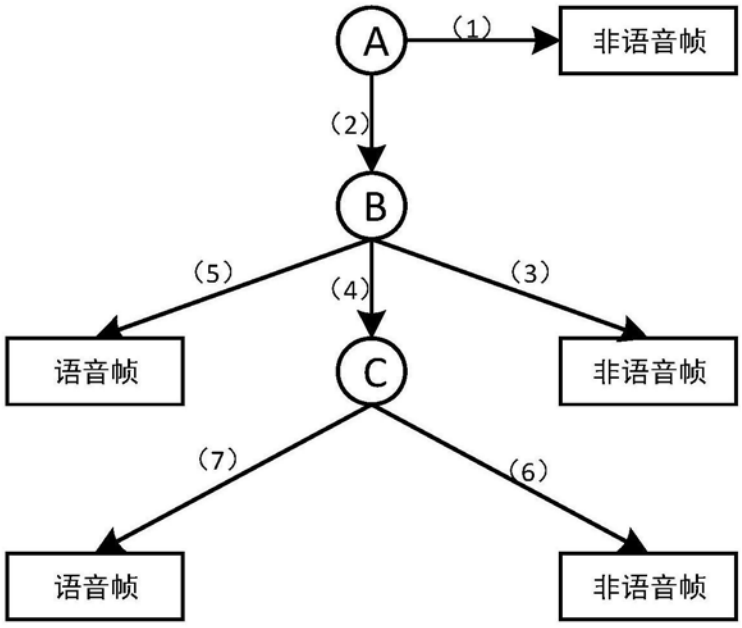

A
(1)
非语音帧
(2)
B
(5)
(4)
(3)
语音帧
C
非语音帧
(7)
(6)
语音帧
非语音帧

图2

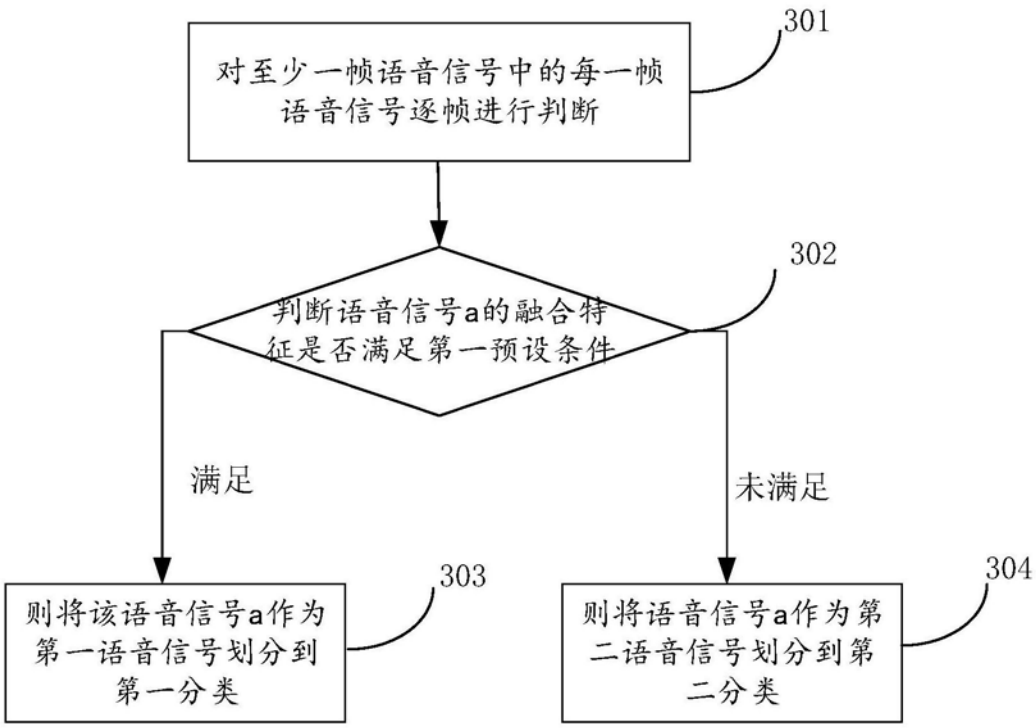

对至少一帧语音信号中的每一帧语音信号逐帧进行判断
301
判断语音信号a的融合特征是否满足第一预设条件
302
满足
未满足
则将该语音信号a作为第一语音信号划分到第一分类
303
则将语音信号a作为第二语音信号划分到第二分类
304

图3

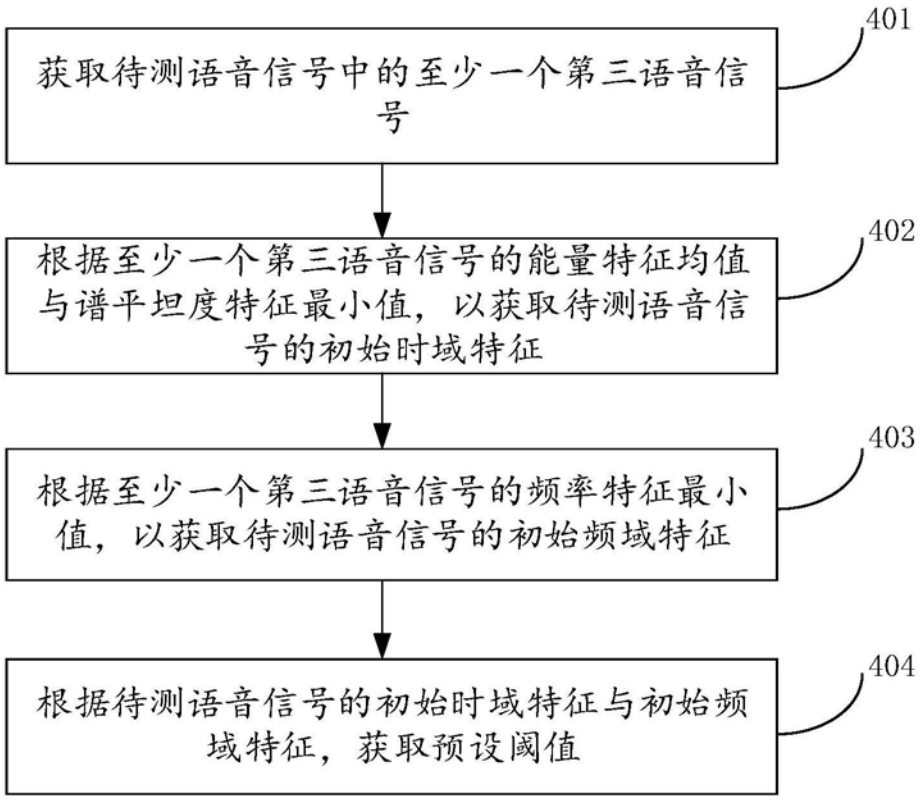

401
获取待测语音信号中的至少一个第三语音信号
402
根据至少一个第三语音信号的能量特征均值与谱平坦度特征最小值，以获取待测语音信号的初始时域特征
403
根据至少一个第三语音信号的频率特征最小值，以获取待测语音信号的初始频域特征
404
根据待测语音信号的初始时域特征与初始频域特征，获取预设阈值

图4


图5

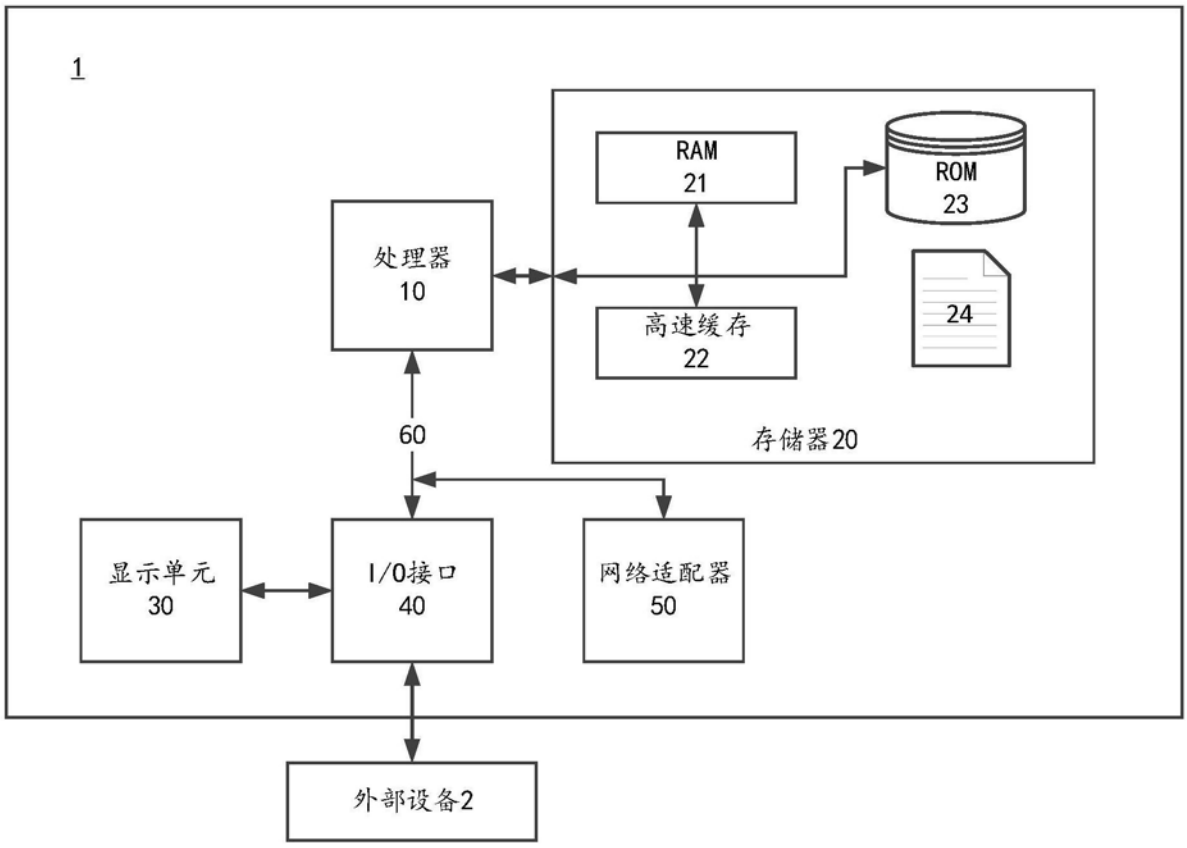

1
RAM
21
ROM
23
处理器
10
高速缓存
22
24
60
存储器20
显示单元
30
I/O接口
40
网络适配器
50
外部设备2

图6

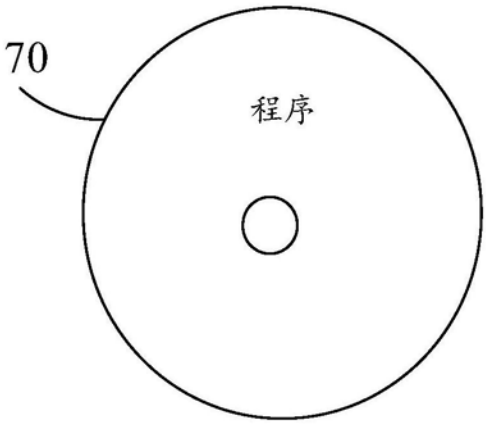

70
程序

图7