(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112802498 B
(45)授权公告日 2023.11.24

(21)申请号 202011590006.0
(22)申请日 2020.12.29
(65)同一申请的已公布的文献号
申请公布号 CN 112802498 A
(43)申请公布日 2021.05.14
(73)专利权人 深圳追一科技有限公司
地址 518051 广东省深圳市南山区粤海街道科技园社区科苑路8号讯美科技广场3号楼23A、23B
(72)发明人 袁丁 周维聪 蒋志宇 刘云峰
(74)专利代理机构 华进联合专利商标代理有限公司 44224
专利代理师 陈小娜
(51)Int.Cl.
G10L 25/87(2013.01)
G10L 25/24(2013.01)
G10L 25/30(2013.01)

(56)对比文件
CN 108877778 A,2018.11.23
CN 102543063 A,2012.07.04
CN 105118502 A,2015.12.02
CN 108010515 A,2018.05.08
CN 109473123 A,2019.03.15
CN 110335621 A,2019.10.15
US 2018068653 A1,2018.03.08
US 2020118571 A1,2020.04.16

审查员 郭忆

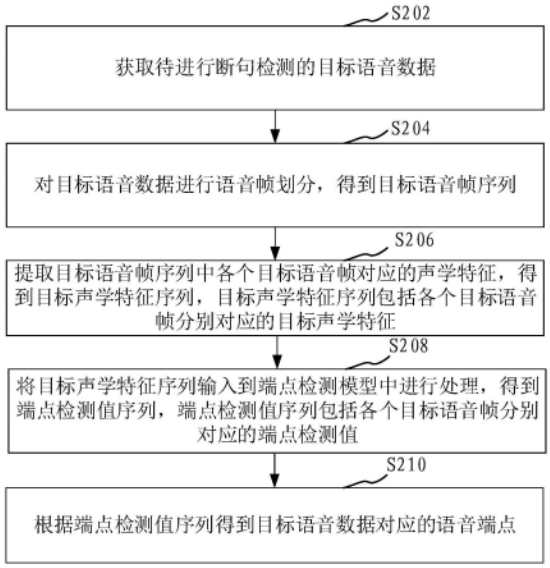
权利要求书3页 说明书14页 附图5页

(54)发明名称

语音检测方法、装置、计算机设备和存储介质

(57)摘要

本申请涉及一种语音检测方法、装置、计算机设备和存储介质。所述方法包括：获取待进行断句检测的目标语音数据；对所述目标语音数据进行语音帧划分，得到目标语音帧序列；提取所述目标语音帧序列中各个目标语音帧对应的声学特征，得到目标声学特征序列，所述目标声学特征序列包括各个所述目标语音帧分别对应的目标声学特征；将所述目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列，所述端点检测值序列包括各个所述目标语音帧分别对应的端点检测值；根据所述端点检测值序列得到所述目标语音数据对应的语音端点。采用本方法能够提高语音检测准确度。



CN 112802498 B

1.一种语音检测方法,其特征在于,所述方法包括:

获取待进行断句检测的目标语音数据;

对所述目标语音数据进行语音帧划分,得到目标语音帧序列;

提取所述目标语音帧序列中各个目标语音帧对应的声学特征,得到目标声学特征序列,所述目标声学特征序列包括各个所述目标语音帧分别对应的目标声学特征;

将所述目标声学特征序列输入到端点检测模型中,所述端点检测模型结合所述目标声学特征序列输出各个所述目标语音帧对应的端点检测概率;

当所述目标语音帧对应的端点检测概率大于预设概率时,获取第一预设值作为端点检测值;

当所述目标语音帧对应的端点检测概率小于预设概率时,获取第二预设值作为端点检测值;

将所述目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列,所述端点检测值序列包括各个所述目标语音帧分别对应的端点检测值;

获取所述端点检测值序列中,所述第二预设值的连续排列数量大于第二数量阈值的第二检测值区域;

将所述第二检测值区域所对应的检测区域语音点作为所述目标语音数据对应的语音起始点;

从所述语音起始点开始,获取所述端点检测值序列中,所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域;

将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音结束点,所述第一数量阈值大于所述第二数量阈值。

2.根据权利要求1所述的方法,其特征在于,所述方法还包括:

将语音结束点与语音起始点之间的语音数据作为噪声语音数据,去除所述目标语音数据中的噪声语音数据。

3.根据权利要求1所述的方法,其特征在于,所述端点检测模型的训练步骤包括:

获取训练语音数据;

获取所述训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧,作为正样本语音帧;

获取所述训练语音数据中同一说话者对应的暂时停顿的语音帧,作为负样本语音帧;

根据所述正样本语音帧以及所述负样本语音帧进行模型训练,得到所述端点检测模型。

4.根据权利要求1所述的方法,其特征在于,所述获取待进行断句检测的目标语音数据包括:

获取会话对端发送的当前语音,当所述当前语音所对应的语音帧到达预设数量时,将当前语音作为待进行断句检测的目标语音数据;

所述方法还包括:

当检测到所述当前语音包括语音结束点时,获取所述当前语音的开始点到所述语音结束点的语音数据,作为待答复的语音数据;

基于所述待答复的语音数据的语义确定会话答复数据,并输出所述会话答复数据至所

述会话对端。

5.一种语音检测装置,其特征在于,所述装置包括:

目标语音数据获取模块,用于获取待进行断句检测的目标语音数据;

语音帧划分模块,用于对所述目标语音数据进行语音帧划分,得到目标语音帧序列;

声学特征提取模块,用于提取所述目标语音帧序列中各个目标语音帧对应的声学特征,得到目标声学特征序列,所述目标声学特征序列包括各个所述目标语音帧分别对应的目标声学特征;

端点检测概率得到单元,用于将所述目标声学特征序列输入到端点检测模型中,所述端点检测模型结合所述目标声学特征序列输出各个所述目标语音帧对应的端点检测概率;

第一预设值获取单元,用于当所述目标语音帧对应的端点检测概率大于预设概率时,获取第一预设值作为端点检测值;

第二预设值获取单元,用于当所述目标语音帧对应的端点检测概率小于预设概率时,获取第二预设值作为端点检测值;

端点检测值序列得到单元,用于将所述目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列,所述端点检测值序列包括各个所述目标语音帧分别对应的端点检测值;

语音端点得到模块,用于获取所述端点检测值序列中,所述第二预设值的连续排列数量大于第二数量阈值的第二检测值区域;将所述第二检测值区域所对应的检测区域语音点作为所述目标语音数据对应的语音起始点;

第一检测值区域获取单元,用于从所述语音起始点开始,获取所述端点检测值序列中,所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域;

语音端点得到单元,用于将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音结束点,所述第一数量阈值大于所述第二数量阈值。

6.根据权利要求5所述的装置,其特征在于,所述装置还包括:

去除模块,用于将语音结束点与语音起始点之间的语音数据作为噪声语音数据,去除所述目标语音数据中的噪声语音数据。

7.根据权利要求5所述的装置,其特征在于,所述装置还包括所述端点检测模型的训练模块;所述训练模块包括:

训练语音数据获取单元,用于获取训练语音数据;

正样本语音帧得到单元,用于获取所述训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧,作为正样本语音帧;

样本语音帧得到单元,用于获取所述训练语音数据中同一说话者对应的暂时停顿的语音帧,作为负样本语音帧;

训练单元,用于根据所述正样本语音帧以及所述负样本语音帧进行模型训练,得到所述端点检测模型。

8.根据权利要求5所述的装置,其特征在于,所述目标语音数据获取模块用于:获取会话对端发送的当前语音,当所述当前语音所对应的语音帧到达预设数量时,将当前语音作为待进行断句检测的目标语音数据;

装置还包括:

待答复的语音数据获取模块，用于当检测到所述当前语音包括语音结束点时，获取所述当前语音的开始点到所述语音结束点的语音数据，作为待答复的语音数据；

会话答复数据确定模块，用于基于所述待答复的语音数据的语义确定会话答复数据，并输出所述会话答复数据至所述会话对端。

9.一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求1至4中任一项所述的方法的步骤。

10.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求1至4中任一项所述的方法的步骤。

语音检测方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及语音处理技术领域，特别是涉及一种语音检测方法、装置、计算机设备和存储介质。

背景技术

[0002] 随着人机信息交互技术的不断发展，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,VAD)是语音识别中的关键技术之一，是语音分析、语音合成、语音编码、说话人识别中的一个重要环节。语音端点检测是指在连续声音信号中找出语音的断句点，通过语音端点检测可以判断用户说话是否出现真正的断句。语音端点检测的准确性，会直接影响到语音识别系统的性能。

[0003] 在传统的语音端点检测方式中，主要是将静音时长与阈值进行比对，通过判断静音时长是否超过阈值，以此判断是否出现断句。例如，阈值可以设置为5秒，当检测到用户未说话的静音时长超过5秒时，则认为检测到语音端点，即用户说话出现断句。然而，经常出现语音端点检测错误的情况，即语音端点检测准确度低。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种语音检测方法、装置、计算机设备和存储介质。

[0005] 一种语音检测方法，所述方法包括：获取待进行断句检测的目标语音数据；对所述目标语音数据进行语音帧划分，得到目标语音帧序列；提取所述目标语音帧序列中各个目标语音帧对应的声学特征，得到目标声学特征序列，所述目标声学特征序列包括各个所述目标语音帧分别对应的目标声学特征；将所述目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列，所述端点检测值序列包括各个所述目标语音帧分别对应的端点检测值；根据所述端点检测值序列得到所述目标语音数据对应的语音端点。

[0006] 在一些实施例中，所述将所述目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列包括：将所述目标声学特征序列输入到端点检测模型中，所述端点检测模型结合所述目标声学特征序列输出各个所述目标语音帧对应的端点检测概率；当所述目标语音帧对应的端点检测概率大于预设概率时，获取第一预设值作为端点检测值；将所述目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列。

[0007] 在一些实施例中，所述根据所述端点检测值序列得到所述目标语音数据对应的语音端点包括：获取所述端点检测值序列中，所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域；将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点。

[0008] 在一些实施例中，所述将所述目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列还包括：当所述目标语音帧对应的端点检测概率小于预设概率时，获取第二预设值作为端点检测值；所述第一检测值区域所对应的前向语音点为语音结束

点，所述根据所述端点检测值序列得到所述目标语音数据对应的语音端点还包括：获取所述端点检测值序列中，所述第二预设值的连续排列数量大于第二数量阈值的第二检测值区域，所述第一数量阈值大于所述第二数量阈值；将所述第二检测值区域所对应的检测区域语音点作为所述目标语音数据对应的语音起始点；从所述语音起始点开始，进入获取所述端点检测值序列中，所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域；将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点的步骤。

[0009] 在一些实施例中，所述方法还包括：将语音结束点与语音起始点之间的语音数据作为噪声语音数据，去除所述目标语音数据中的噪声语音数据。

[0010] 在一些实施例中，所述端点检测模型的训练步骤包括：获取训练语音数据；获取所述训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧，作为正样本语音帧；获取所述训练语音数据中同一说话者对应的暂时停顿的语音帧，作为负样本语音帧；根据所述正样本语音帧以及所述负样本语音帧进行模型训练，得到所述端点检测模型。

[0011] 在一些实施例中，所述获取待进行断句检测的目标语音数据包括：获取会话对端发送的当前语音，当所述当前语音所对应的语音帧到达预设数量时，将当前语音作为待进行断句检测的目标语音数据；所述方法还包括：当检测到所述当前语音包括语音结束点时，获取所述当前语音的开始点到所述语音结束点的语音数据，作为待答复的语音数据；基于所述待答复的语音数据的语义确定会话答复数据，并输出所述会话答复数据至所述会话对端。

[0012] 一种语音检测装置，所述装置包括：目标语音数据获取模块，用于获取待进行断句检测的目标语音数据；语音帧划分模块，用于对所述目标语音数据进行语音帧划分，得到目标语音帧序列；声学特征提取模块，用于提取所述目标语音帧序列中各个目标语音帧对应的声学特征，得到目标声学特征序列，所述目标声学特征序列包括各个所述目标语音帧分别对应的目标声学特征；端点检测值序列得到模块，用于将所述目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列，所述端点检测值序列包括各个所述目标语音帧分别对应的端点检测值；语音端点得到模块，用于根据所述端点检测值序列得到所述目标语音数据对应的语音端点。

[0013] 在一些实施例中，所述端点检测值序列得到模块包括：端点检测概率单元，用于将所述目标声学特征序列输入到端点检测模型中，所述端点检测模型结合所述目标声学特征序列输出各个所述目标语音帧对应的端点检测概率；第一预设值获取单元，用于当所述目标语音帧对应的端点检测概率大于预设概率时，获取第一预设值作为端点检测值；端点检测值序列得到单元，用于将所述目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列。

[0014] 在一些实施例中，所述语音端点得到模块包括：第一检测值区域获取单元，用于获取所述端点检测值序列中，所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域；语音端点得到单元，用于将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点。

[0015] 在一些实施例中，所述端点检测值序列得到模块还包括：第二预设值获取单元，用于当所述目标语音帧对应的端点检测概率小于预设概率时，获取第二预设值作为端点检测

值;所述第一检测值区域所对应的前向语音点为语音结束点,所述语音端点得到模块还用于:获取所述端点检测值序列中,所述第二预设值的连续排列数量大于第二数量阈值的第二检测值区域,所述第一数量阈值大于所述第二数量阈值;将所述第二检测值区域所对应的检测区域语音点作为所述目标语音数据对应的语音起始点;从所述语音起始点开始,进入获取所述端点检测值序列中,所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域;将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点的步骤。

[0016] 在一些实施例中,所述装置还包括:去除模块,用于将语音结束点与语音起始点之间的语音数据作为噪声语音数据,去除所述目标语音数据中的噪声语音数据。

[0017] 在一些实施例中,所述端点检测模型的训练模块包括:训练语音数据获取单元,用于获取训练语音数据;正样本语音帧得到单元,用于获取所述训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧,作为正样本语音帧;负样本语音帧得到单元,用于获取所述训练语音数据中同一说话者对应的暂时停顿的语音帧,作为负样本语音帧;训练单元,用于根据所述正样本语音帧以及所述负样本语音帧进行模型训练,得到所述端点检测模型。

[0018] 在一些实施例中,所述目标语音数据获取模块用于:获取会话对端发送的当前语音,当所述当前语音所对应的语音帧到达预设数量时,将当前语音作为待进行断句检测的目标语音数据;所述装置还包括:待答复的语音数据获取模块,用于当检测到所述当前语音包括语音结束点时,获取所述当前语音的开始点到所述语音结束点的语音数据,作为待答复的语音数据;会话答复数据确定模块,用于基于所述待答复的语音数据的语义确定会话答复数据,并输出所述会话答复数据至所述会话对端。

[0019] 一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现以下步骤:获取待进行断句检测的目标语音数据;对所述目标语音数据进行语音帧划分,得到目标语音帧序列;提取所述目标语音帧序列中各个目标语音帧对应的声学特征,得到目标声学特征序列,所述目标声学特征序列包括各个所述目标语音帧分别对应的目标声学特征;将所述目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列,所述端点检测值序列包括各个所述目标语音帧分别对应的端点检测值;根据所述端点检测值序列得到所述目标语音数据对应的语音端点。

[0020] 在一些实施例中,所述将所述目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列包括:将所述目标声学特征序列输入到端点检测模型中,所述端点检测模型结合所述目标声学特征序列输出各个所述目标语音帧对应的端点检测概率;当所述目标语音帧对应的端点检测概率大于预设概率时,获取第一预设值作为端点检测值;将所述目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列。

[0021] 在一些实施例中,所述根据所述端点检测值序列得到所述目标语音数据对应的语音端点包括:获取所述端点检测值序列中,所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域;将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点。

[0022] 在一些实施例中,所述将所述目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列还包括:当所述目标语音帧对应的端点检测概率小于预设概率时,

获取第二预设值作为端点检测值;所述第一检测值区域所对应的前向语音点为语音结束点,所述根据所述端点检测值序列得到所述目标语音数据对应的语音端点还包括:获取所述端点检测值序列中,所述第二预设值的连续排列数量大于第二数量阈值的第二检测值区域,所述第一数量阈值大于所述第二数量阈值;将所述第二检测值区域所对应的检测区域语音点作为所述目标语音数据对应的语音起始点;从所述语音起始点开始,进入获取所述端点检测值序列中,所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域;将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点的步骤。

[0023] 在一些实施例中,所述计算机程序被处理器执行时还实现以下步骤:将语音结束点与语音起始点之间的语音数据作为噪声语音数据,去除所述目标语音数据中的噪声语音数据。

[0024] 在一些实施例中,所述端点检测模型的训练步骤包括:获取训练语音数据;获取所述训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧,作为正样本语音帧;获取所述训练语音数据中同一说话者对应的暂时停顿的语音帧,作为负样本语音帧;根据所述正样本语音帧以及所述负样本语音帧进行模型训练,得到所述端点检测模型。

[0025] 在一些实施例中,所述获取待进行断句检测的目标语音数据包括:获取会话对端发送的当前语音,当所述当前语音所对应的语音帧到达预设数量时,将当前语音作为待进行断句检测的目标语音数据;所述计算机程序被处理器执行时还实现以下步骤:当检测到所述当前语音包括语音结束点时,获取所述目标语音数据中当前语音的开始点到所述语音结束点的语音数据,作为待答复的语音数据;基于所述待答复的语音数据的语义确定会话答复数据,并输出所述会话答复数据至所述会话对端。

[0026] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现以下步骤:获取待进行断句检测的目标语音数据;对所述目标语音数据进行语音帧划分,得到目标语音帧序列;提取所述目标语音帧序列中各个目标语音帧对应的声学特征,得到目标声学特征序列,所述目标声学特征序列包括各个所述目标语音帧分别对应的目标声学特征;将所述目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列,所述端点检测值序列包括各个所述目标语音帧分别对应的端点检测值;根据所述端点检测值序列得到所述目标语音数据对应的语音端点。

[0027] 在一些实施例中,所述将所述目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列包括:将所述目标声学特征序列输入到端点检测模型中,所述端点检测模型结合所述目标声学特征序列输出各个所述目标语音帧对应的端点检测概率;当所述目标语音帧对应的端点检测概率大于预设概率时,获取第一预设值作为端点检测值;将所述目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列。

[0028] 在一些实施例中,所述根据所述端点检测值序列得到所述目标语音数据对应的语音端点包括:获取所述端点检测值序列中,所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域;将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点。

[0029] 在一些实施例中,所述将所述目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列还包括:当所述目标语音帧对应的端点检测概率小于预设概率时,

获取第二预设值作为端点检测值;所述第一检测值区域所对应的前向语音点为语音结束点,所述根据所述端点检测值序列得到所述目标语音数据对应的语音端点还包括:获取所述端点检测值序列中,所述第二预设值的连续排列数量大于第二数量阈值的第二检测值区域,所述第一数量阈值大于所述第二数量阈值;将所述第二检测值区域所对应的检测区域语音点作为所述目标语音数据对应的语音起始点;从所述语音起始点开始,进入获取所述端点检测值序列中,所述第一预设值的连续排列数量大于第一数量阈值的第一检测值区域;将所述第一检测值区域所对应的前向语音点作为所述目标语音数据对应的语音端点的步骤。

[0030] 在一些实施例中,所述计算机程序被处理器执行时还实现以下步骤:将语音结束点与语音起始点之间的语音数据作为噪声语音数据,去除所述目标语音数据中的噪声语音数据。

[0031] 在一些实施例中,所述端点检测模型的训练步骤包括:获取训练语音数据;获取所述训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧,作为正样本语音帧;获取所述训练语音数据中同一说话者对应的暂时停顿的语音帧,作为负样本语音帧;根据所述正样本语音帧以及所述负样本语音帧进行模型训练,得到所述端点检测模型。

[0032] 在一些实施例中,所述获取待进行断句检测的目标语音数据包括:获取会话对端发送的当前语音,当所述当前语音所对应的语音帧到达预设数量时,将当前语音作为待进行断句检测的目标语音数据;所述计算机程序被处理器执行时还实现以下步骤:当检测到所述当前语音包括语音结束点时,获取所述目标语音数据中当前语音的开始点到所述语音结束点的语音数据,作为待答复的语音数据;基于所述待答复的语音数据的语义确定会话答复数据,并输出所述会话答复数据至所述会话对端。

[0033] 上述语音检测方法、装置、计算机设备和存储介质,获取待进行断句检测的目标语音数据,对目标语音数据进行语音帧划分,得到目标语音帧序列,提取目标语音帧序列中各个目标语音帧对应的声学特征,得到目标声学特征序列,目标声学特征序列包括各个目标语音帧分别对应的目标声学特征,将目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列,端点检测值序列包括各个目标语音帧分别对应的端点检测值,根据端点检测值序列得到目标语音数据对应的语音端点。由于在端点检测时,是通过进行语音帧划分,基于语音帧的声学特征进行检测得到的,而且在确定语音端点时,基于端点检测值序列确定得到语音端点,故可以准确得到语音端点,提高了得到语音端点的准确度。

附图说明

[0034] 图1为一些实施例中语音检测方法的应用环境图;

[0035] 图2为一些实施例中语音检测方法的流程示意图;

[0036] 图3为一些实施例中得到MFCC特征的示意图;

[0037] 图4为一些实施例中将目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列步骤的流程示意图;

[0038] 图5为一些实施例中端点检测模型的训练步骤的流程示意图;

[0039] 图6为一些实施例中语音检测装置的结构框图;

[0040] 图7为一些实施例中端点检测值序列得到模块的结构框图;

[0041] 图8为一些实施例中计算机设备的内部结构图。

具体实施方式

[0042] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0043] 本申请提供的语音检测方法，可以应用于如图1所示的应用环境中。其中，终端102通过网络与服务器104进行通信。终端可以进行语音采集，得到待进行断句检测的目标语音数据，服务器执行本申请实施例提供的语音检测方法，得到目标语音数据对应的语音端点。服务器得到语音数据的端点之后，可以对该语音数据进行进一步的处理，例如进行切分，切分之后对语音片段进行语音识别，识别到其中的文本，基于文本进行语义理解，基于语义理解的结果进行会话。其中，终端102可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

[0044] 可以理解，本申请实施例提供的方法还可以是在终端执行的。

[0045] 在一些实施例中，如图2所示，提供了一种语音检测方法，以该方法应用于图1中的服务器为例进行说明，包括以下步骤：

[0046] 步骤S202，获取待进行断句检测的目标语音数据。

[0047] 其中，断句是指将目标语音数据切分为多段语音，每段语音表示一段完整的句子。

[0048] 具体地，可以通过终端实时采集目标语音数据，上传到服务器中，服务器中也可以预先存储待进行断句检测的目标语音数据。例如，服务器中可以存储有大量的语音数据，需要对这些语音数据进行端点检测，以确定用户是否说完了，以获取到完整语义的语音，基于该语音识别得到该段语句所表示的含义。因此可以获取这些未进行端点检测的语音数据，作为待进行断句检测的目标语音数据。

[0049] 在一些实施例中，对于人机交互，例如智能机器人与人进行电话会话，由于是为了及时、准确的识别、响应或者回复用户的语音内容，因此可以通过本申请实施例的方法检测用户语音的断句情况。因此，获取的待检测的语音数据可以是用户语音通道的单通道语音数据，语音数据包括用户的说话内容。语音数据具体可以是实时的流数据，按照预定的一帧语音数据的长度，例如可以为50毫秒一帧或者10毫秒一帧等，随着时间一帧一帧的流入得到序列的音频数据。即待检测的语音数据可以是一段音频数据，例如待检测的语音数据可以是包括预设帧数量的音频数据，比如一段待检测的语音数据包括20帧的音频数据。

[0050] 步骤S204，对目标语音数据进行语音帧划分，得到目标语音帧序列。

[0051] 具体地，服务器可以按照预设时长，对目标语音数据进行划分，每个语音帧的时间长度为预设时长，例如预设时长可以为10毫秒。将划分得到的语音帧按照语音的顺序进行排列，得到目标语音帧序列，目标语音帧序列包括多帧按照语音顺序排列的语音帧。

[0052] 步骤S206，提取目标语音帧序列中各个目标语音帧对应的声学特征，得到目标声学特征序列，目标声学特征序列包括各个目标语音帧分别对应的目标声学特征。

[0053] 其中，声学特征是表示语音的声学特性的特征，声学特征可以是指表示语音声学特性的物理量，如表示音色的能量集中区、共振峰频率、共振峰强度、带宽、表示语音韵律特

性的时长、基频或者平均语声功率的至少一种等。声学特征可以是梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient,MFCC)。

[0054] 具体地,服务器可以对每个目标语音帧进行声学特征提取,得到每个目标语音帧对应的声学特征,声学特征按照语音的顺序进行排列,得到目标声学特征序列。

[0055] 在一些实施例中,每一帧语音数据对应的声学特征具体可以包括MFCC或者pitch(音高)特征,通过音频的音高、音调、频率以及能量等表示用户说话的特征。

[0056] MFCC的提取步骤可如图3所示,可以将音频数据通过一系列的倒谱向量来描述,每个向量就每帧音频数据对应的MFCC特征向量。即对于语音数据,可以先进行预加重、分帧和加窗处理,然后再进行傅里叶变换(FFT),取变换后的平均值或者绝对值之后,进行梅尔谱滤波(mel滤波),再进行取对数处理以及DCT变换,得到动态特征。

[0057] 步骤S208,将目标声学特征序列输入到端点检测模型中进行处理,得到端点检测值序列,端点检测值序列包括各个目标语音帧分别对应的端点检测值。

[0058] 其中,语音端点可以包括语音起始点以及结束点。一个目标语音数据中可以包括多个语音端点。例如,一个目标语音数据中,假设A与B在进行对话,A在说完之后,间隔了1秒,B开始说话。则目标语音数据中,包括A开始说话的起始点以及结束说话的结束点,B开始说话的起始点以及结束说话的结束点。语音端点检测可以称为语音活动检测(Voice Activity Detection,VAD)。

[0059] 端点检测值是用于判断是否为端点的值,例如可以是0或者1。端点检测模型是用于检测是否为端点的模型,可以是深度神经网络模型。端点检测模型是预先采用有监督的训练得到的,在训练时,用于预训练神经网络模型的训练数据是包括语音部分和静音部分的一整段音频数据,整段音频数据中的静音部分可能属于真正断句结束后的静音,也可能属于暂时性停顿(即非真正断句结束)时的静音。因此,可以将训练数据包括非真正断句结束的语音部分和静音部分例如用户犹豫、思考或者磕绊等暂时性停顿静音作为负样本,真正断句结束的语音部分和静音部分例如真正结束说话后等待回复时的静音作为正样本,神经网络模型可以基于训练数据学习得到整段语音数据对应声学特征的上下文关系,基于上下文(context)语义综合得到是否为端点的概率。

[0060] 具体地,端点检测模型可以输出目标语音帧为端点的概率,服务器可以根据概率得到端点检测值。端点检测值按照语音顺序排列,得到端点检测值序列。即端点检测模型在检测端点时,深度神经网络可以基于多帧音频数据的上下文确定每一帧音频数据的检测结果。

[0061] 在一些实施例中,在提取声学特征输入神经网络模型进行端点检测之前,可以对待检测的语音数据进行预处理。具体的,可以检测语音数据包括的静音时长,当语音数据的静音时长超过预设时长阈值时,则确定真正断句结束。例如,预设时长阈值可以设置为300毫秒或者500毫秒,当语音数据中的静音时长超过预设时长阈值时,由于空白静音时间太长可能已经无法捕捉前面的声学特征了,输入神经网络模型可能产生负面效果,因此可以直接确定真正的断句结束。当语音数据中的静音时长未超过预设时长阈值时,再通过神经网络模型进行检测。

[0062] 步骤S210,根据端点检测值序列得到目标语音数据对应的语音端点。

[0063] 具体地,得到端点检测值序列后,可以获取端点检测值序列中满足端点条件的区

域，获取该区域所对应的语音点，例如时间点或者语音帧的序号，作为语音端点。端点条件可以包括起点条件以及结束点条件。可以根据一帧的检测结果判断是否断句，也可以根据连续多帧的检测结果判断是否断句。例如当连续出现两帧音频数据的检测结果为“0”时，确定是真正断句结束了，例如“11000”。而例如“10111”这种情况，只有一个“0”，则确定用户还在说话，确定语音数据中不存在语音终止点。

[0064] 上述语音检测方法中，获取待进行断句检测的目标语音数据，对目标语音数据进行语音帧划分，得到目标语音帧序列，提取目标语音帧序列中各个目标语音帧对应的声学特征，得到目标声学特征序列，目标声学特征序列包括各个目标语音帧分别对应的目标声学特征，将目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列，端点检测值序列包括各个目标语音帧分别对应的端点检测值，根据端点检测值序列得到目标语音数据对应的语音端点。由于在端点检测时，是通过进行语音帧划分，基于语音帧的声学特征进行检测得到的，而且在确定语音端点时，基于端点检测值序列确定得到语音端点，故可以准确得到语音端点，提高了得到语音端点的准确度。

[0065] 在一些实施例中，如图4所示，将目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列包括：

[0066] 步骤S402，将目标声学特征序列输入到端点检测模型中，端点检测模型结合目标声学特征序列输出各个目标语音帧对应的端点检测概率。

[0067] 其中，一个目标语音帧对应的端点检测概率是结合整个目标声学特征序列得到的。例如端点检测模型可以是深度神经模型。

[0068] 具体地，服务器将目标声学特征序列输入到端点检测模型中，端点检测模型输出每个目标语音帧为端点的概率。

[0069] 步骤S404，当目标语音帧对应的端点检测概率大于预设概率时，获取第一预设值作为端点检测值。

[0070] 其中，预设概率可以根据需要设置，一般而言大于0.5，例如可以是0.8。第一预设值可以是预先设置的，例如可以是0。

[0071] 具体地，对于每个端点检测概率，如果端点检测概率大于预设概率，则将第一预设值作为端点检测值。否则，可以将第二预设值例如1作为端点检测值。

[0072] 步骤S406，将目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列。

[0073] 具体地，服务器然后按照语音帧的顺序对端点检测值进行排列，得到端点检测值序列。举个例子，假设有5个语音帧，假设得到该语音帧对应的端点检测概率分别为0.20、0.30、0.85、0.99以及0.10。预设概率为0.8，第一预设值为0，第二预设值为1，则端点检测值序列为1、1、0、0以及1。

[0074] 在一些实施例中，结束点条件包括第一预设值的连续排列数量大于第一数量阈值，根据端点检测值序列得到目标语音数据对应的语音端点包括：获取端点检测值序列中，第一预设值的连续排列数量大于第一数量阈值的第一检测值区域；将第一检测值区域所对应的前向语音点作为目标语音数据对应的语音端点。

[0075] 其中，第一检测值区域对应的前向语音点是指第一检测值区域对应的前一个语音点。第一数量阈值可以根据需要设置，可以是大于2的数值，例如可以是3。检测值区域所对

应的语音点可以用时间表示，也可以用语音帧的序号表示。检测值区域所对应的语音点可以是检测值区域的起点、中间点或者结束点。举个例子，假设第一数量阈值为2，第一预设值为0，端点检测值序列为1、1、0、1、0、0、0、1、1、1、0、0、0。第5个语音帧到第7个语音帧均为0，即0的连续排列数量为3，因此第5个语音帧到第7个语音帧为满足0的连续排列数量大于第一数量阈值的检测值区域，因此可以将第一检测值区域的前一个语音帧作为目标语音数据对应的语音端点，例如第4个语音帧作为语音结束点。通过将第一检测值区域对应的前向语音点作为语音端点，可以减少噪声，以及减少在线识别场景中造成一定的等待延迟的情况。

[0076] 在一些实施例中，起始点条件包括第二预设值的连续排列数量大于第二数量阈值。因此可以获取端点检测值序列中，第二预设值的连续排列数量大于第二数量阈值的第二检测值区域；将该第二检测值区域所对应的检测区域语音点作为目标语音数据对应的语音端点，得到语音端点集合。即语音端点包括起始点与结束点组成的语音端点集合。

[0077] 在一些实施例中，检测区域语音点是第二检测值区域所在的语音点，可以是第二检测值区域的初始语音点。针对语音的起始点和终止点，可以分别设置不同的判断阈值，例如第一数量阈值大于第二数量阈值。例如，语音的起始点是从没说话转换为开始说话，音频数据的特征变化比较明显，因此可以将起始点对应的阈值设置较小，例如2帧，即出现两帧音频数据对应的检测结果为“11”时则确定语音的起始点，为第一个“1”所对应的语音帧，从而提高了语音端点的检测效率。而说话过程转换为检测语音的终止点时，是从正在说话转换为不说话，音频数据的特征变化相较没有那么明显，为了保证端点检测的准确性，可以将终止点对应的阈值设置较大，例如4帧，即出现4帧音频数据对应的检测结果为“0000”时则确定为真正断句结束，将“0000”之前的“1”对应的语音帧作为语音结束点。

[0078] 本申请实施例中，通过连续多帧的端点检测值判断是否断句，从而提高准确度。例如，模型的检测结果的准确率不是100％的，假设一帧的错误率为0.1，若根据连续三帧的检测结果确定是否断句，连续三帧的错误率就只有0.001。

[0079] 在一些实施例中，将目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列还包括：当目标语音帧对应的端点检测概率小于预设概率时，获取第二预设值作为端点检测值；第一检测值区域所对应的前向语音点为语音结束点，根据端点检测值序列得到目标语音数据对应的语音端点还包括：获取端点检测值序列中，第二预设值的连续排列数量大于第二数量阈值的第二检测值区域，第一数量阈值大于第二数量阈值；将第二检测值区域所对应的检测区域语音点作为目标语音数据对应的语音起始点；从语音起始点开始，进入获取端点检测值序列中，第一预设值的连续排列数量大于第一数量阈值的第一检测值区域；将第一检测值区域所对应的前向语音点作为目标语音数据对应的语音端点的步骤。

[0080] 其中，第二预设值例如可以为1。在由第一预设值以及第二预设值组成的端点值序列中，可以先检测语音开始点，将端点检测值序列中，第二预设值的连续排列数量大于第二数量阈值的第二检测值区域所对应的检测区域语音点作为语音起始点，然后从该语音起始点开始，获取语音结束点。得到语音结束点之后，继续进入获取语音起始点的步骤。这样，服务器可以持续检测语音数据，当检测到语音起始点后，采用语音终止点的判断策略检测和判断终止点。当检测到终止点后，转换为语音起始点的判断策略检测和判断语音的起始点，以此反复检测用户语音的端点。

[0081] 在一些实施例中，对于神经网络模型的输出可以是每帧音频数据对应的端点检测概率，在根据预设概率进行二分类时，针对起始点和终止点也可以分别设置不同的概率阈值，即检测起始点时的概率阈值和检测终止点时的预设概率可以设置为不同的，以此提高端点检测的效率和准确性。例如，起始点的概率阈值可以大于终止点的概率阈值，由于在检测时，不确定是否存在新的一段语音，因此可以设置相对较高的概率阈值，即预设概率，以确保检测得到的语音起始点的准确度较高，而当存在语音起始点时，由于一般会存在语音结束点，因此可以设置相对降低的概率，以确保可以检测到语音结束点。

[0082] 本申请实施例中，每一帧音频数据对应得到二分类的检测结果。例如，属于真正断句结束(终止点)的音频数据对应的检测结果表示为“0”，属于非真正断句结束的音频数据对应的检测结果表示为“1”。虽然是每一帧音频数据会得到一个检测结果，但是每帧音频数据的检测结果是综合了整段语音数据的上下文的。比如待检测的语音数据对应的检测结果可以表示为“1011100000”。其中，表示为“1”的音频数据是表示非真正断句结束的音频数据，即表示为“1”的音频数据可能是用户正在说话的语音数据，也可能是用户暂时性停顿时的静音数据，从而避免将用户暂时性停顿时的静音数据错误判断为断句结束的静音数据，在会话时可以避免出现错误打断用户说话的情况。

[0083] 在一些实施例中，可以将语音结束点与语音起始点之间的语音数据作为噪声语音数据，去除目标语音数据中的噪声语音数据。

[0084] 具体地，服务器可以将结束点到起始点之间的数据进行滤除，从而去除目标语音数据中的真正静音的部分。这样，后续进行语音识别时，可以去除静音部分的干扰。即对于一段语音数据，假设其第一个语音帧为语音终止点，最后一个语音帧为语音起始点，则说明该段语音数据为静音数据，删除该段语音数据。

[0085] 在一些实施例中，如图5所示，端点检测模型的训练步骤包括：

[0086] 步骤S502，获取训练语音数据。

[0087] 其中，训练语音数据是用于训练端点检测模型的数据，在训练端点检测模型时，训练语音数据可以有多个，例如1万个。

[0088] 具体地，服务器可以从训练语料库中获取用于进行模型训练的训练样本，即训练语音数据。

[0089] 步骤S504，获取训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧，作为正样本语音帧。

[0090] 其中，正样本语音帧是指真正断句的语音帧，为正样本。断句结束的语音帧以及说话者切换的语音帧可以是人工标注的。说话者可以是目标语音中说话的人。说话者切换点是由一个说话者切换到另一个说话者的点。假设4-7秒为A在说话，7-10秒为B在说话，则第7秒为说话者切换点。可以理解，说话者切换点可以用语音帧的序号表示。例如，可以用第7秒所对应的语音帧例如第100帧表示。

[0091] 具体地，对于训练语音数据中，真正断句结束的语音帧以及说话者切换点对应的语音帧，服务器可以确定其对应的标签为正样本对应的标签。否则，确定其对应的标签为负样本对应的标签。

[0092] 步骤S506，获取训练语音数据中同一说话者对应的暂时停顿的语音帧，作为负样本语音帧。

[0093] 具体地，暂时性停顿可以是用户犹豫、思考或者磕绊等暂时性停顿静音。暂时停顿的语音帧可以是人工标注的。

[0094] 本申请实施例的训练数据可以分为两类，一类是真正断句结束，一类是非真正断句结束的，传统方式可能会认为这两种都是断句结束。但是，例如说话过程中因为思考、犹豫或者结巴等出现的暂时性停顿，并不是真正的断句，即此时用户说话并未结束，暂时性停顿的前后是可以连接起来表达完整的一句话。因此，对于一段音频数据，人工将真正断句结束的标记为正样本，将非真正断句结束的音频数据标记为负样本，从而得到用于训练深度神经网络模型的训练数据，通过将标记好的训练数据输入神经网络模型中，调节神经网络模型中的参数，以此得到语音数据的端点检测模型。

[0095] 步骤S508，根据正样本语音帧以及负样本语音帧进行模型训练，得到端点检测模型。

[0096] 具体地，训练时，可以进行多轮迭代训练，直至模型满足收敛条件。模型收敛条件可以是模型损失值小于预设损失值或者迭代次数达到预设次数的至少一个。模型损失值可以根据端点检测模型输出的端点检测概率与样本的标签值的差异得到，其中模型损失值与差异成正相关关系，即差异越大，则模型损失值越大。

[0097] 本申请实施例中，用于训练神经网络模型的训练数据是与传统方式不同的，训练数据包括非真正断句结束的语音部分和静音部分例如用户犹豫、思考或者磕绊等暂时性停顿静音作为负样本，真正断句结束的语音部分和静音部分例如真正结束说话后等待回复时的静音作为正样本，神经网络模型可以基于训练数据学习得到整段语音数据对应声学特征的上下文关系，基于上下文语义综合得到检测结果，从而提高了端点检测的准确度。

[0098] 在一些实施例中，步骤S202即获取待进行断句检测的目标语音数据包括：获取会话对端发送的当前语音，当当前语音所对应的语音帧到达预设数量时，将当前语音作为待进行断句检测的目标语音数据。

[0099] 会话对端是会话的另一端，本申请实施例提供的方法可以是由会话机器人执行的，会话对端可以是与会话机器人进行电话的另一端。预设数量可以根据需要设置，例如可以是2秒所对应的语音帧的数量。

[0100] 具体地，在会话时，音频流一帧一帧的流入，获取到预设长度的音频数据后，作为一段待检测的语音数据，例如每10毫秒一帧的流入，获取到20帧后作为一段待检测的语音数据进行检测。将待检测的语音数据划分为多帧音频数据，提取多帧音频数据各自对应的声学特征，将多帧音频数据对应的声学特征输入至预训练的神经网络模型，得到神经网络模型输出的多帧音频数据各自对应的检测结果，从而可以根据多帧音频数据各自对应的检测结果确定语音端点，判断是否存在或者哪些帧属于起始点或者终止点。

[0101] 语音检测方法还包括：当检测到当前语音包括语音结束点时，获取当前语音的开始点到语音结束点的语音数据，作为待答复的语音数据。基于待答复的语音数据的语义确定会话答复数据，并输出会话答复数据至会话对端。

[0102] 具体地，当得到语音结束点时，说明会话对端的用户已经说完，获取到的当前语音数据是包含完整语义的数据，因此可以获取从该用户开始说话到结束说话的语音数据，基于该语音数据进行语义理解，例如可以将该语音数据对应的文本作为问句，将该问句输入到答案确定模型中，答案确定模型输出该问句的答案，会话机器人将该答案作为会话答复

数据，并转化为语音，向会话对端发送该语音数据，从而可以实现智能对话。

[0103] 本申请实施例中，通过预训练的神经网络模型能够利用上下文信息更加准确的检测出语音数据中的语音端点，提高语音端点检测的效率和准确性，从而能够及时响应和回复客户，同时也避免了在用户还未真正结束说话时提前打断用户说话。

[0104] 本申请实施例的方案通过提取每一帧语音数据的声学特征，采用预先训练的深度神经网络检测该帧语音数据是否属于真正的断句，深度神经网络是通过包括真正断句和非真正断句的语音数据训练得到的。对于语句中间的暂时性停顿的静音数据，也可以检测得到非真正断句的结果。通过深度神经网络将每一帧语音数据进行二分类，根据连续多帧语音数据的检测结果检测是否为语音的起始点或者终止点，从而有效的提高了语音端点检测的准确性和效率。

[0105] 应该理解的是，虽然上述流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，上述流程图中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0106] 在一些实施例中，如图6所示，提供了一种语音检测装置，包括：一种语音检测装置，装置包括：

[0107] 目标语音数据获取模块602，用于获取待进行断句检测的目标语音数据；

[0108] 语音帧划分模块604，用于对目标语音数据进行语音帧划分，得到目标语音帧序列；

[0109] 声学特征提取模块606，用于提取目标语音帧序列中各个目标语音帧对应的声学特征，得到目标声学特征序列，目标声学特征序列包括各个目标语音帧分别对应的目标声学特征；

[0110] 端点检测值序列得到模块608，用于将目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列，端点检测值序列包括各个目标语音帧分别对应的端点检测值；

[0111] 语音端点得到模块610，用于根据端点检测值序列得到目标语音数据对应的语音端点。

[0112] 在一些实施例中，如图7所示，端点检测值序列得到模块608包括：

[0113] 端点检测概率得到单元702，用于将目标声学特征序列输入到端点检测模型中，端点检测模型结合目标声学特征序列输出各个目标语音帧对应的端点检测概率；

[0114] 第一预设值获取单元704，用于当目标语音帧对应的端点检测概率大于预设概率时，获取第一预设值作为端点检测值；

[0115] 端点检测值序列得到单元706，用于将目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列。

[0116] 在一些实施例中，语音端点得到模块包括：第一检测值区域获取单元，用于获取端点检测值序列中，第一预设值的连续排列数量大于第一数量阈值的第一检测值区域；语音端点得到单元，用于将第一检测值区域所对应的前向语音点作为目标语音数据对应的语音

端点。

[0117] 在一些实施例中，端点检测值序列得到模块还包括：第二预设值获取单元，用于当目标语音帧对应的端点检测概率小于预设概率时，获取第二预设值作为端点检测值；第一检测值区域所对应的前向语音点为语音结束点，语音端点得到模块还用于：获取端点检测值序列中，第二预设值的连续排列数量大于第二数量阈值的第二检测值区域，第一数量阈值大于第二数量阈值；将第二检测值区域所对应的检测区域语音点作为目标语音数据对应的语音起始点；从语音起始点开始，进入获取端点检测值序列中，第一预设值的连续排列数量大于第一数量阈值的第一检测值区域；将第一检测值区域所对应的前向语音点作为目标语音数据对应的语音端点的步骤。

[0118] 在一些实施例中，装置还包括：去除模块，用于将语音结束点与语音起始点之间的语音数据作为噪声语音数据，去除目标语音数据中的噪声语音数据。

[0119] 在一些实施例中，端点检测模型的训练模块包括：训练语音数据获取单元，用于获取训练语音数据；正样本语音帧得到单元，用于获取训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧，作为正样本语音帧；负样本语音帧得到单元，用于获取训练语音数据中同一说话者对应的暂时停顿的语音帧，作为负样本语音帧；训练单元，用于根据正样本语音帧以及负样本语音帧进行模型训练，得到端点检测模型。

[0120] 在一些实施例中，目标语音数据获取模块用于：获取会话对端发送的当前语音，当当前语音所对应的语音帧到达预设数量时，将当前语音作为待进行断句检测的目标语音数据；装置还包括：待答复的语音数据获取模块，用于当检测到当前语音包括语音结束点时，获取当前语音的开始点到语音结束点的语音数据，作为待答复的语音数据；会话答复数据确定模块，用于基于待答复的语音数据的语义确定会话答复数据，并输出会话答复数据至会话对端。

[0121] 关于语音检测装置的具体限定可以参见上文中对于语音检测方法的限定，在此不再赘述。上述语音检测装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

[0122] 在一些实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图8所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储语音数据处理数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种语音检测方法。

[0123] 本领域技术人员可以理解，图8中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0124] 在一些实施例中，提供了一种计算机设备，包括存储器和处理器，存储器中存储有计算机程序，该处理器执行计算机程序时实现上述语音检测方法。

[0125] 在一些实施例中，提供了一种计算机可读存储介质，其上存储有计算机程序，计算

机程序被处理器执行时实现上述语音检测方法。

[0126] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。

[0127] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0128] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

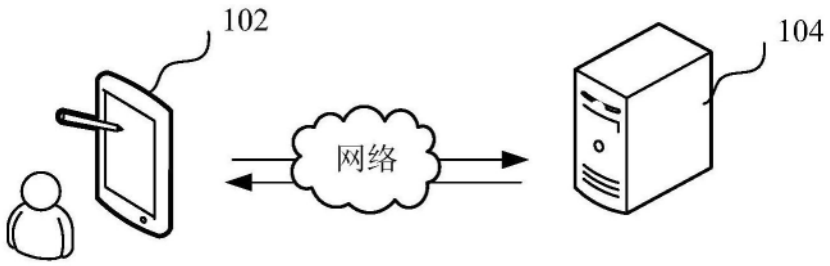

102
104
网络

图1

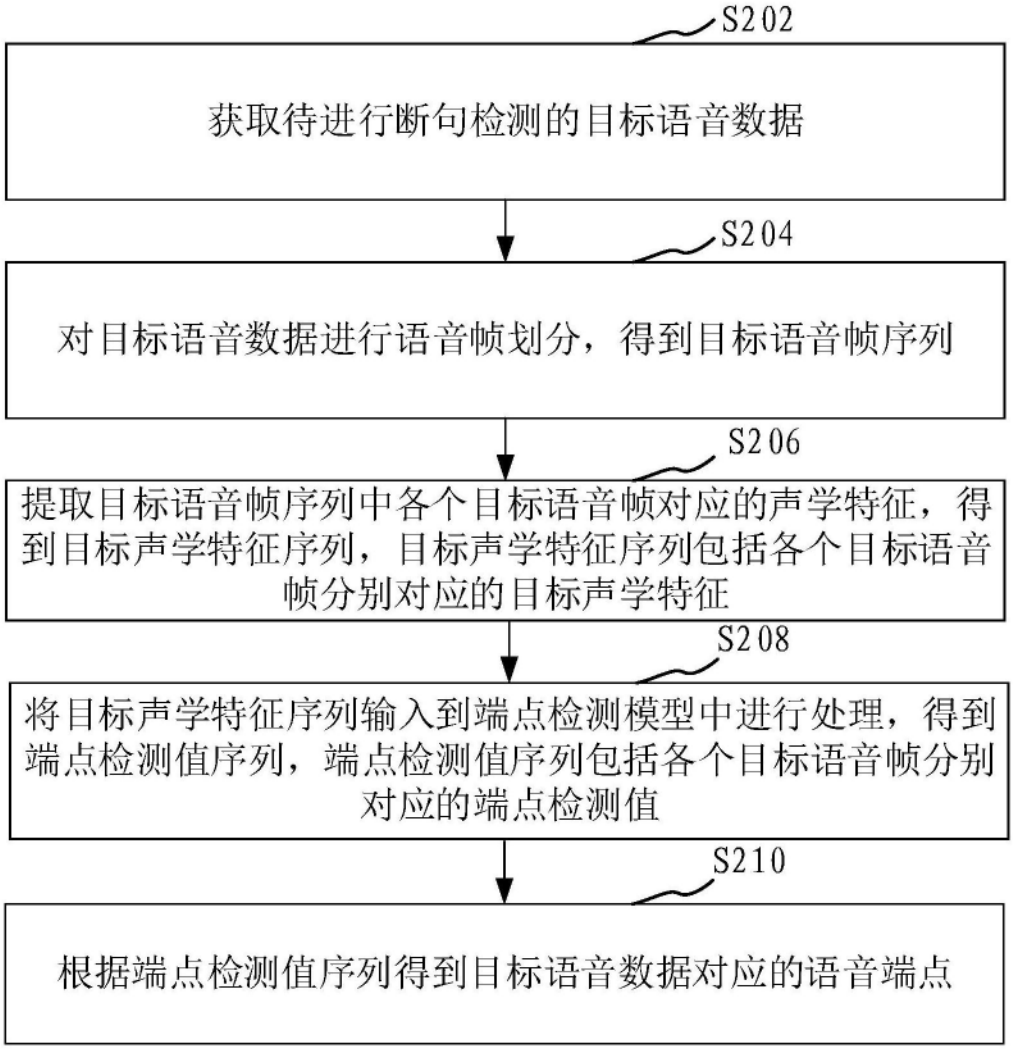

S202
获取待进行断句检测的目标语音数据
S204
对目标语音数据进行语音帧划分，得到目标语音帧序列
S206
提取目标语音帧序列中各个目标语音帧对应的声学特征，得到目标声学特征序列，目标声学特征序列包括各个目标语音帧分别对应的目标声学特征
S208
将目标声学特征序列输入到端点检测模型中进行处理，得到端点检测值序列，端点检测值序列包括各个目标语音帧分别对应的端点检测值
S210
根据端点检测值序列得到目标语音数据对应的语音端点

图2

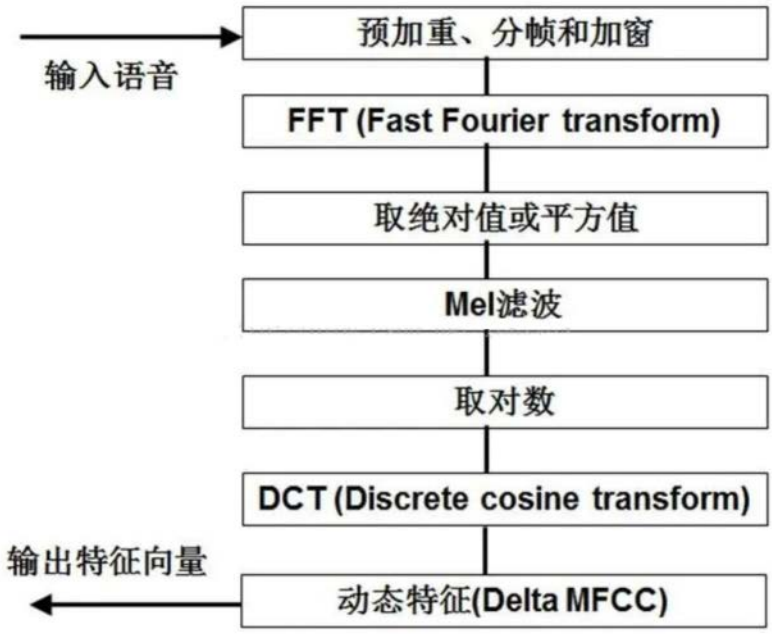

预加重、分帧和加窗
输入语音
FFT (Fast Fourier transform)
取绝对值或平方值
Mel滤波
取对数
DCT (Discrete cosine transform)
输出特征向量
动态特征(Delta MFCC)

图3

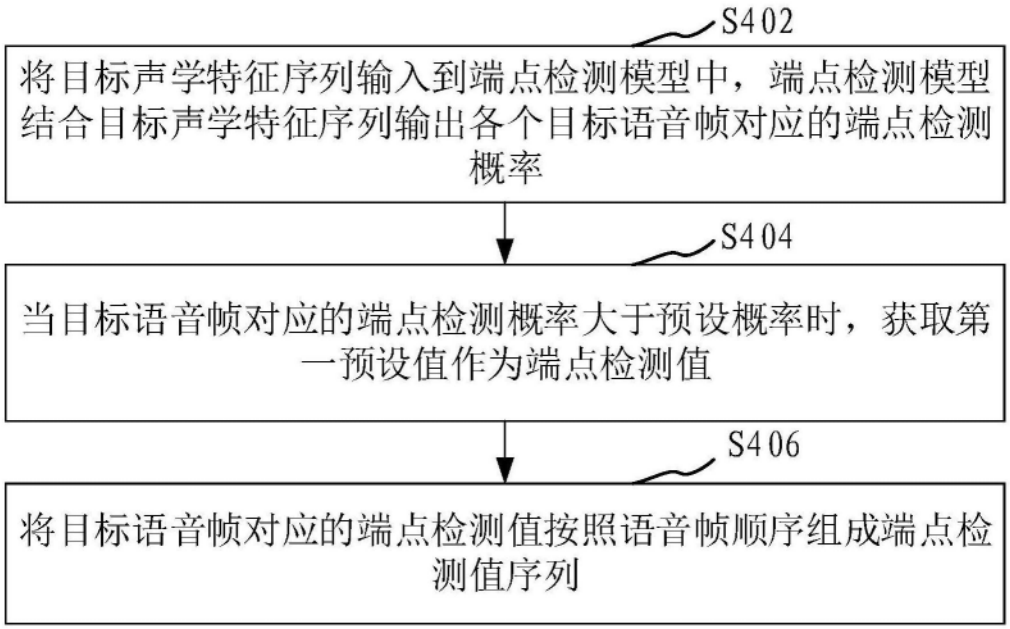

S402
将目标声学特征序列输入到端点检测模型中，端点检测模型结合目标声学特征序列输出各个目标语音帧对应的端点检测概率
S404
当目标语音帧对应的端点检测概率大于预设概率时，获取第一预设值作为端点检测值
S406
将目标语音帧对应的端点检测值按照语音帧顺序组成端点检测值序列

图4

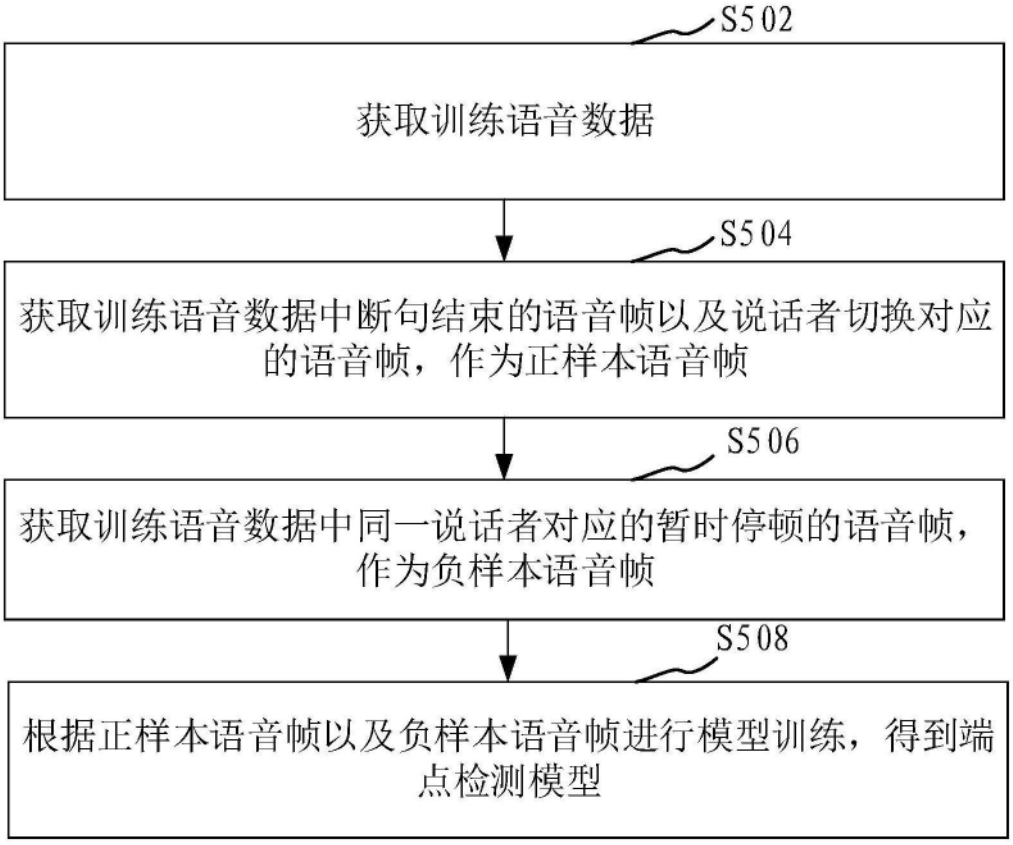

S502
获取训练语音数据
S504
获取训练语音数据中断句结束的语音帧以及说话者切换对应的语音帧，作为正样本语音帧
S506
获取训练语音数据中同一说话者对应的暂时停顿的语音帧，作为负样本语音帧
S508
根据正样本语音帧以及负样本语音帧进行模型训练，得到端点检测模型

图5

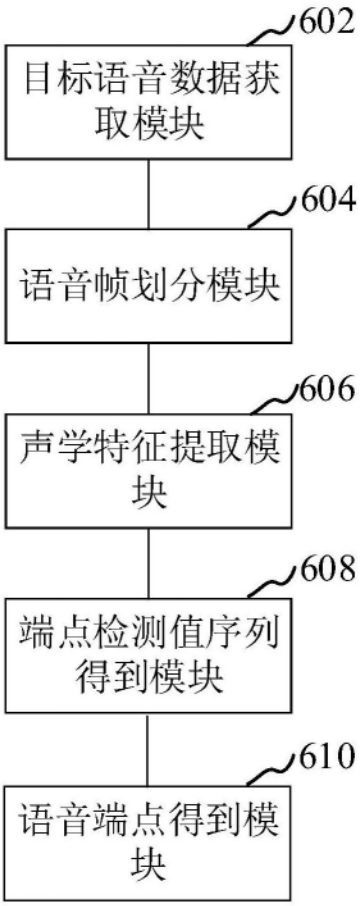

602
目标语音数据获取模块
604
语音帧划分模块
606
声学特征提取模块
608
端点检测值序列得到模块
610
语音端点得到模块

图6

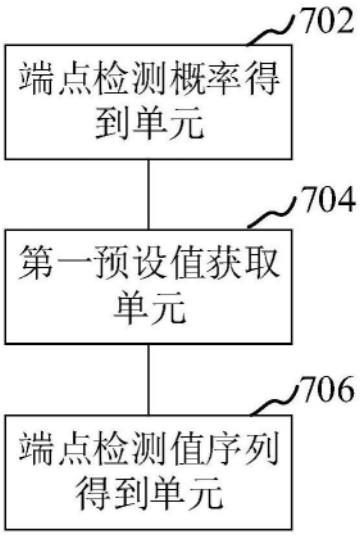

702
端点检测概率得到单元
704
第一预设值获取单元
706
端点检测值序列得到单元

图7

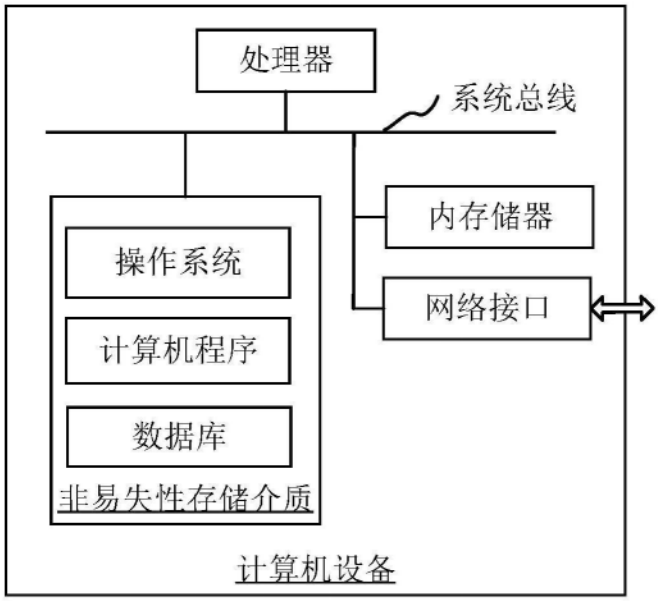


图8