(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 7 月 9 日 (09.07.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/140374 A1

(51) 国际专利分类号:
G10L 17/00 (2013.01)

(21) 国际申请号: PCT/CN2019/088976

(22) 国际申请日: 2019 年 5 月 29 日 (29.05.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910018423.9 2019年1月4日 (04.01.2019) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: 刘博卿 (LIU, Boqing); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 贾雪丽 (JIA, Xueli); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 王健宗 (WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 广州三环专利商标代理有限公司 (SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路80号汇华商贸大厦1508室, Guangdong 510070 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,

(54) **Title:** VOICE DATA PROCESSING METHOD, APPARATUS AND DEVICE AND STORAGE MEDIUM

(54) **发明名称:** 语音数据处理方法、装置、设备及存储介质

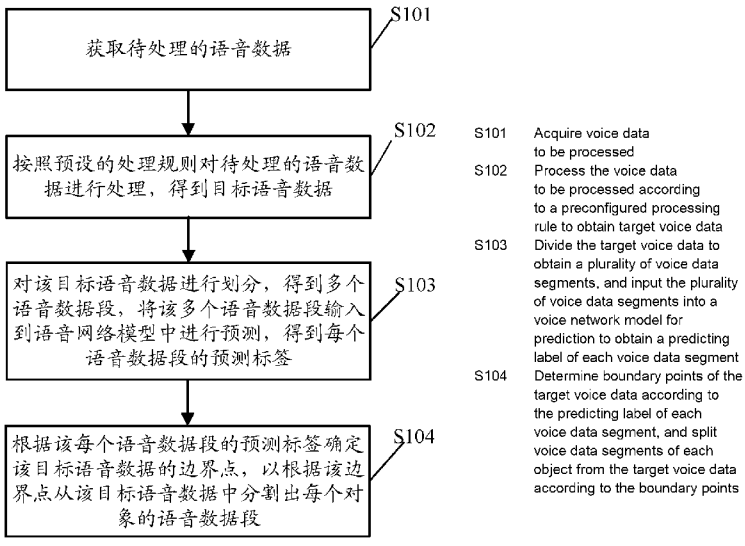

图 1

(57) **Abstract:** A voice data processing method, apparatus and device and a storage medium. The method comprises: acquiring voice data to be processed, the voice data to be processed consisting of voice data segments of a plurality of objects (S101); processing the voice data to be processed according to a preconfigured processing rule to obtain target voice data (S102); dividing the target voice data to obtain a plurality of voice data segments, and inputting the plurality of voice data segments into a voice network model for prediction to obtain a predicting label of each voice data segment (S103); and determining boundary points of the target voice data according to the predicting label of each voice data segment, and splitting voice data segments of each object from the target voice data according to the boundary points (S104). The method can automatically acquire boundary points of voice data and can improve the accuracy in acquiring the boundary points of voice data.

WO 2020/140374 A1

[见续页]

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种语音数据处理方法、装置、设备及储存介质，该方法包括：获取待处理的语音数据，待处理的语音数据由多个对象的语音数据段组成（S101）；按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据（S102）；对目标语音数据进行划分，得到多个语音数据段；将多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签（S103）；根据每个语音数据段的预测标签确定目标语音数据的边界点，根据边界点从目标语音数据中分割出每个对象的语音数据段（S104）。该方法能够自动获取语音数据的边界点，并且能够提高获取语音数据的边界点的准确度。

语音数据处理方法、装置、设备及存储介质

本申请要求于 2019 年 01 月 04 日提交中国专利局、申请号为 201910018423.9、申请名称为“语音数据处理方法、装置、设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机技术领域，尤其涉及一种语音数据处理方法、装置、设备及存储介质。

背景技术

通常呼叫中心接收到的语音数据很多都混杂有多人的语音片段，这时需要先对语音数据进行语音分割(speaker diarization)，才能进一步对目标语音片段进行语音分析。语音分割是指：通过获取每两个人的语音边界点（边界点可以是指每两个人说话的转换点）来分割出每个人的语音段。实践中，需要人工对语音进行分析，以获取每两个人的语音边界点，导致语音分割的效率及准确度较低。

发明内容

本申请实施例提供一种语音数据处理方法、装置、设备及存储介质，可自动检测出语音数据的边界点，提高语音分割的效率及准确度。

第一方面，本申请实施例提供了一种　语音数据处理方法，包括：

获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成；

按照预设的处理规则对所述待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则；

对所述目标语音数据进行划分，得到多个语音数据段；将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率；

根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

第二方面，本申请实施例提供了一种语音数据处理装置，包括：

获取单元，用于获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成；

处理单元，用于按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则；

预测单元，用于对所述目标语音数据进行划分，得到多个语音数据段；将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率；

分割单元，用于根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

第三方面，本申请实施例提供了另一种电子设备，包括：

处理器，适于实现一条或一条以上指令；以及，

计算机可读存储介质，所述计算机可读存储介质存储有一条或一条以上指令，所述一条或一条以上指令适于由所述处理器加载并执行如下步骤：

获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成；

按照预设的处理规则对所述待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则；

对所述目标语音数据进行划分，得到多个语音数据段；将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率；

根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

第四方面，本申请实施例提供了一种计算机可读存储介质，包括：所述计算机可读存储介质存储有一条或一条以上指令，所述一条或一条以上指令适于由处理器加载并执行如下步骤：

获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成；

按照预设的处理规则对所述待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则；

对所述目标语音数据进行划分，得到多个语音数据段；将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率；

根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

本申请实施例中，可通过语音网络模型自动获取目标语音数据的边界点，不需要人工操作，可节省大量人力，并可提高获取语音数据的边界点的准确度及效率；满足用户对语音数据处理的智能化、自动化需求。

附图说明

图 1 是本申请实施例提供的一种语音数据处理方法的流程示意图；

图 2 是本申请实施例提供的一种语音数据处理方法的流程示意图；

图 3 是本申请实施例提供的一种语音数据处理装置的结构示意图；

图 4 是本申请实施例提供的一种电子设备的结构示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

基于现有技术中需要人工分析方式确定语音数据的边界点，导致语音分割的效率及准确度较低的问题，本申请实施例提供一种自动语音数据处理的方法，该方法可以由电子设备来执行，该电子设备可以是指智能终端、服务器、电脑或检测仪等设备；该方法可以通过语音网络模型对语音数据进行预测，得到预测标签，根据预测标签确定语音数据的边界点，根据语音数据的边界点分割出每个对象的语音数据段，可节省大量人力，并可提高检测准确度，满足用户语音分割智能化、

自动化需求。

请参见图 1，是本申请实施例提供的一种语音数据处理方法的流程示意图，本申请实施例的所述方法可以由上述提及的电子设备来执行。本实施例中，该语音数据处理方法包括以下步骤。

S101、获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成。

本申请实施例中，该处理的语音数据在不同应用场景下所包括的具体内容不同。例如，在会议场景中，该待处理的语音数据可以是指对多个人所讲的话进行录音得到的语音数据；在呼叫应用场景中，该待处理的语音数据可以是指呼叫中心接收到的多个人的呼叫数据。这里的对象主要是指说话的人，对象也可以是指动物。

S102、按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则。

本申请实施例中，由于待处理的语音数据中包括冗余数据，如静音数据或非人的语音等等，非人的语音包括设备的声音（如拍照声）、环境声音（如车辆的声音）；为了避免对冗余数据进行处理，提高处理效率，可以对待处理的语音数据进行预处理。具体的，电子设备可以采用滤波器对待处理的语音数据进行过滤处理，得到目标语音数据，其中，滤波器可以为高通滤波器或带通滤波器等。和/或者，为了便于语音网络模型对语音数据进行预测，电子设备可以对待处理的语音数据进行格式转换处理。

在一个实施例中，所述预设的处理规则包括数据过滤规则，步骤 S102 包括：将该待处理的语音数据进行划分，得到多个原始语音数据段，获取该多个原始语音数据段中每个原始语音数据段的能量值，将该多个原始语音数据段中能量值小于或等于预设能量值的原始语音数据段删除，将该多个原始语音数据段中能量值大于该预设能量值的原始语音数据段进行合并，得到该目标语音数据。

通常有用语音数据的能量大于干扰语音数据（如静音或非人的声语音）的能量，因此，电子设备可以根据语音数据的能量对待处理的语音数据进行过滤处理。具体的，电子设备可以按照预设长度将该待处理的语音数据进行划分，得到多个原始语音数据段，对该多个原始语音数据段中每个原始语音数据段进行时频变换，得到每个原始语音数据段的频域信息，每个原始语音数据段的频域信息用于描述每个原始语音数据段的频率与能量之间的关系。根据每个原始语音数据段的频域信息获取对应原始语音数据段的能量值，由于能量值较小的原始语音数据段为干扰语音数据的概率较高，能量值较高的原始语音数据段为有用语音数据的概率较高；因此可以将该多个原始语音数据段中能量值小于或等于预设能量值的原始语音数据段删除，将该多个原始语音数据段中能量值大于该预设能量值的原始语音数据段进行合并，得到该目标语音数据。

在另一个实施例中，该预设的处理规则包括数据格式处理规则，步骤 S102 包括：获取该待处理的语音数据的数据格式，当该待处理的语音数据的数据格式与预设数据格式不相同时，按照该预设数据格式该待处理的语音数据进行格式转换处理，得到该目标语音数据。

为了便于语音网络模型对语音数据进行预测，电子设备可以获取该待处理的语音数据的数据格式，当该待处理的语音数据的数据格式与预设数据格式不相同时，表明该待处理语音数据的数据格式不适用于语音网络模型预测处理，该预设

数据格式可以为适用于语音网络模型对语音数据预测的数据格式，例如该格式可以为脉冲编码调制（Pulse Code Modulation，PCM），或语音交互格式（Audio Interchange File Format，AIFF）等等。按照该预设数据格式对该待处理的语音数据进行格式转换处理，得到该目标语音数据，该目标语音数据的数据格式为预设数据格式。

S103、对该目标语音数据进行划分，得到多个语音数据段，将该多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，该预测标签包括语音数据段为边界点的概率。

本申请实施例中，为了提高预测的效率，电子设备可以将该目标语音数据划分为相互之间不重叠的多个语音数据段，或者，为了提高预测的准确度，电子设备可以将该目标语音数据划分为相互之间重叠的多个语音数据段。进一步，将多个语音数据段输入到语音网络模型中，语音网络模型对每个语音数据段中的特征点进行分析并进行预测，得到每个语音数据段的预测标签。

在一个实施例中，电子设备可以将该目标语音数据划分为相互之间不重叠的多个语音数据段。具体的，电子设备可以按照预设语音长度将多个语音数据段划分为多个语音数据段。例如，该预设语音长度为 10s，待处理的语音数据的长度为 100s，电子设备可以将该待处理的语音数据划分为 10 个语音数据段，如，第一语音数据段为 0~10s，第二语音数据段为 10~20s，第三语音数据段为 20s~30s，以此类推得到 10 个语音数据段。

在另一个实施例中，若各个语音数据段为不重叠的语音数据段，由于各个对象之间说话转换很快，容易导致检测不到边界点。例如，第一语音数据段和第二语音数据段为非重叠且相邻的语音数据段，若第一语音数据段为对象 A 的语音数据，若第二语音数据段为对象 B 的语音数据。由于第一语音数据内的特征点都具有相似性，第二语音数据内的特征点都具有相似性，则第一语音数据、第二语音数据都不会被预测为边界点，导致检测不到边界点。为了避免错过边界点，电子设备可以将该目标语音数据划分为相互之间重叠的多个语音数据段。具体的，按照预设步长对所述目标语音数据进行划分，得到多个语音数据段，该多个语音数据段中每个语音数据段的长度大于该预设步长。其中，每个语音数据段的长度可以相同，也可以不相同，下面以每个语音数据段的长度相同为例，例如，预设步长为 8s，每个语音数据的长度为 10s，则第一语音数据段为 0~10s，第二语音数据段为 8~18s，第三语音数据段为 16~26s。每两个相邻的语音数据段之间包括重叠的语音数据。

在一个实施例中，上述将该多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，包括：语音网络模型分析每个语音数据段中各个特征点的相似度，根据各个特征点的相似度，计算每个语音数据段的相似度总和，根据每个语音数据段的相似度总和确定对应语音数据段的预测标签。其中，语音数据段的相似度总和越大，表明该语音数据段为同一个对象的语音数据的概率较大，该语音数据段为边界点的概率较小；语音数据段的相似度总和越小，表明该语音数据段为同一个对象的语音数据的概率较小，该语音数据段为边界点的概率较大。例如，对于第一语音数据段，该语音数据段包括 5 个特征点，分别为 x1，x2，x3，x4 及 x5。电子设备可以分别计算每个特征点与其他特征点之间的相似度，如，分别计算 x1 与 x2，x3，x4 及 x5 的相似度，对各个相似度进行累加得到 x1 的相似度总和，同理计算得到 x2-x5 的相似度总和，将各个特征点的相似度总和进行累加，得到该第一语音数据段的相似度总和，根据第一语音数

据段的相似度总和确定该第一语音数据段的预测标签，并输出该预设标签。该预测标签包括语音数据段为边界点的概率，该概率的值为[0,1]，概率的值越大，表明对应语音数据段（或语音数据段内的特征点）为边界点的概率越大，概率的值越小，表明对应语音数据段（或语音数据段内的特征点）为边界点的概率越小。特别的，若概率的值为 0，则对应语音数据段（或语音数据段内的特征点）不为边界点；若概率的值为 1，则对应语音数据段（或语音数据段内的特征点）为边界点。

在一个实施例中，语音数据段的特征点可以是指以下任一项或多项能量、音调及音色等等，能量是指声音的强度（即声音的响度），音调是指声音的高低，音色是指声音的特性。当语音数据段的特征点可以包括能量、音调及音色时，上述计算每个语音数据段的相似度总和，包括：根据每个语音数据段的能量特征点之间的相似度计算得到对应语音数据段的第一相似度总和，根据每个语音数据段的音调特征点之间的相似度计算得到对应语音数据段的第二相似度总和，根据每个语音数据段的音色特征点之间的相似度计算得到对应语音数据段的第三相似度总和；对第一相似度总和、第二相似度总和及第三相似度总和进行加权求和，得到每个语音数据段的相似度总和。其中，第一相似度总和、第二相似度总和及第三相似度总和对应的权重可以是用户设置的，也可以是电子设备根据应用场景设置的。例如，在各个对象的声音响度差别较大的场景中，可以将第一相似度总和的权重设置一个较大值，以突出各个对象的声音能量的差别。在各个对象的音调差别较大的场景中，可以将第二相似度总和的权重设置一个较大值，以突出各个对象的音调的差别。在各个对象的音色差别较大的场景中，可以将第三相似度总和的权重设置一个较大值，以突出各个对象的音色的差别。

在一个实施例中，该语音网络模型可以由两个长短期记忆网络模型（即 Bi-LSTMs）及一个多层神经网络模型，该多层神经网络模型可以与其中一个长短期记忆网络模型连接。每个 Bi-LSTMs 包括前向长短期记忆网络层 forward LSTM 和后向长短期记忆网络层 backward LSTM，两者的输出连接到一起，然后输出给下一层；以便 Bi-LSTMs 可以通过前向和反向两个方向来处理语音数据段，从而可以同时利用以前的和将来的上下文信息。多层神经网络模型由三个全连接层组成，前两层的激活方程可以是 tanh 函数，最后一层的激活方程可以 sigmoid 函数；以便可以输出一个范围在 0 和 1 之间的概率（预测标签）。

S104、根据该每个语音数据段的预测标签确定该目标语音数据的边界点，以根据该边界点从该目标语音数据中分割出每个对象的语音数据段。

本申请实施例中，为了分割各个对象的语音数据段，电子设备可以根据该语音数据段的预测标签确定该目标语音数据的边界点，例如，电子设备可以将概率大于预设概率的语音数据段作为目标语音数据的边界点，或将概率大于预设概率的语音数据段中的部分特征点作为目标语音数据的边界点。并根据该边界边从该目标语音数据中分割出每个对象的语音数据段，以便对某个对象的语音数据段进行分析。目标语音数据的边界点可以是指不同对象的语音数据段的转化点，例如，假设目标语音数据中第四语音数据段为边界点，则第一语音数据段到第四语音数据段为第一对象（如第一个人）的语音数据段，第四语音数据段以后的语音数据段为第二对象（如第二个人）的语音数据段。

本申请实施例中，可通过语音网络模型自动获取目标语音数据的边界点，不需要人工操作，可节省大量人力，并可提高获取语音数据的边界点的准确度及效率；满足用户对语音数据处理的智能化、自动化需求。

请参见图 2，是本申请实施例提供的另一种语音数据处理方法的流程示意图，本申请实施例的该方法可以由上述提及的电子设备来执行。本实施例中，该语音数据处理方法包括以下步骤。

S201、获取训练样本集合，该训练样本集合包括样本音频数据的多个样本语音数据段，及每个样本语音数据段的标注标签，该标注标签包括样本语音数据段为边界点的概率。

S202、以该多个样本语音数据段为该语音网络模型的输入，以该每个样本语音数据段的标注标签为该语音网络模型的训练目标，对该语音网络模块进行迭代训练。

在步骤 S201 和 S202 中，为了提高语音网络模型的预测准确度，可以对语音网络模型进行优化训练。具体的，电子设备可以采集训练样本集，该训练样本集包括样本音频数据的多个样本语音数据段，以及每个样本语音数据段的标注标签。其中，该标注标签包括样本语音数据段为边界点的概率，该标注标签可以是指人工对特征样本进行标注的。为了提高语音网络模型的适用范围，该样本音频数据可以是由处于不同区域的对象的语音数据段组成，或/和，该样本音频数据可以是由不同年龄的对象的语音数据段组成。进一步，以该多个样本语音数据段为该语音网络模型的输入，以该每个样本语音数据段的标注标签为该语音网络模型的训练目标；当该语音网络模型输出的语音数据段的预测标签与该语音数据段的标注标签相同，或者这两者相似时，表明语音网络模型的预测准确度较高，则结束对语音网络模型的迭代训练；当该语音网络模型输出的语音数据段的预测标签与该语音数据段的标注标签差异较大，表明该语音网络模型的预测准确度较低，则调整语音网络模型的网络参数，继续对该语音网络模型进行迭代训练。

在一个实施例中，步骤 S202 包括：将该多个样本语音数据段输入到该语音网络模型进行预测，得到每个样本语音数据段的预测标签；根据该每个样本语音数据段的预测标签及对应的样本语音数据段的预测标签确定该语音网络模型的预测误差；若该语音网络模型的预测误差大于预设误差值，则调整该语音网络模型的网络参数；若该语音网络模型的预测误差小于或等于该预设误差值，则结束对该语音网络模型的迭代训练。

电子设备可以通过调整语音网络模型的网络参数来优化语音网络模型，具体的，将该多个样本语音数据段输入到该语音网络模型进行预测，得到每个样本语音数据段的预测标签，根据该每个样本语音数据段的预测标签及对应的样本语音数据段的标注标签确定该语音网络模型的预测误差；若该语音网络模型的预测误差大于预设误差值，表明语音网络模型的预测准确度较低，则调整该语音网络模型的网络参数，并继续对语音网络模型进行迭代训练；若该语音网络模型的预测误差小于或等于该预设误差值，则表明语音网络模型的预测准确度较高，可以结束对该语音网络模型的迭代训练。例如，假设目标语音数据段包括 T 个语音数据段，可表示为 $X=(\mathrm{X1},\ \mathrm{X2},\ldots XT)$，其中，每个语音数据段可以是由多个特征点组成，第 i 个语音数据段的标注标签为 y_i，第 i 个语音数据段的预测标签为 $f(X)_i$，则语音网络模型的预测误差 L 可以表示为公式（1）所示。

$$L=\frac{1}{T}\sum_{i=1}^{T}y_i\log\left(f(X)_i\right)+(1-y_i)\log\left(1-f(X)_i\right) \qquad (1)$$

在一个实施例中，接收将目标样本语音数据段的标注标签设置为第一类标注标签的指令，该第一类标注标签为样本语音数据段为边界点的概率大于预设概率

的标签，该目标样本语音数据段为样本音频数据的边界点所在的语音数据段；将与该目标样本语音数据段之间的时间间隔小于或等于预设时间间隔的样本语音数据段的标注标签设置为第一类标注标签；将与该目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段的标注标签设置为第二类标注标签，该第二类标注标签为样本语音数据段为边界点的概率小于或等于该预设概率的标签。

语音数据中包括边界点和非边界点，由于语音数据中边界点相对比较少，即只有极少的语音数据段为边界点的概率大于预设阈值，极多数的语音数据段为边界点的概率小于预设阈值；这两类数据的不平衡（相差太多）特征在对语音网络模型进行训练时导致很多问题，如导致语音网络模型的预测准确度较低。因此，可以将为真正边界点的附近增加正样本的数量，即增加概率大于预设概率的语音数据段。具体的，电子设备可以接收将目标样本语音数据段的标注标签设置为第一类标注标签的指令，该目标样本语音数据段为样本音频数据的边界点所在的语音数据段，即目标样本语音数据段为真正的边界点所在的语音数据段。为了解决不平衡问题，可以将与该目标样本语音数据段之间的时间间隔小于或等于预设时间间隔的样本语音数据段的标注标签设置为第一类标注标签，即将目标样本语音数据段附近的语音数据段都标注为边界点（即正样本）所在的语音数据段。将与该目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段的标注标签设置为第二类标注标签，即将与该目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段标注为非边界点所在的语音数据段。例如，预设时间间隔为 50ms，电子设备可以将与目标样本数据段的时间间隔为 50ms 内的所有样本语音数据段标注为第一类标注标签。

S203、获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成。

S204、按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则。

S205、对该目标语音数据进行划分，得到多个语音数据段，将该多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，该预测标签包括语音数据段为边界点的概率。

S206、根据该每个语音数据段的预测标签确定该目标语音数据的边界点，以根据该边界点从该目标语音数据中分割出每个对象的语音数据段。

在一个实施例中，该每个语音数据段包括多个特征点，每相邻两个语音数据段包括相同的特征点，该每个语音数据段的预测标签为对应语音数据段内每个特征点的预测标签，步骤 S206 包括：根据该每个语音数据段的预测标签统计每个特征点为边界点的平均概率，将平均概率大于预设概率的所有特征点作为该目标语音数据的边界点。

由于相邻的每个语音数据段具有重叠性，因此，每相邻两个语音数据段包括相同的特征点，针对同一个特征点存在有两个概率的情况下，电子设备可以根据各个特征点的平均概率来确定语音数据的边界点。例如，假设目标语音数据包括第一语音数据段及第二语音数据段，第一语音数据段包括特征点 x1，x2，x3，x4 及 x5，第一语音数据段的预测标签为 0.2，则特征点 x1，x2，x3，x4 及 x5 的预测标签（概率）均为 0.2；第二语音数据段包括特征点 x4，x5，x6，x7 及 x8。第二语音数据段的预测标签为 0.5，则特征点 x4，x5，x6，x7 及 x8 的预测标签均为 0.5；根据预测标签确定每个特征点的平均标签（即平均概率），即特

征点 x1，x2，x3 的平均预测标签均为 0.2，特征点 x4 及 x5 的平均预测标签为 0.35，特征点 x6，x7 及 x8 的平均预测标签需要根据第三语音数据段的预测标签进行计算。进一步，将平均概率大于预设概率的所有特征点作为该目标语音数据的边界点。

本申请实施例中，可通过语音网络模型自动获取目标语音数据的边界点，不需要人工操作，可节省大量人力，并可提高获取语音数据的边界点的准确度及效率；满足用户对语音数据处理的智能化、自动化需求。

请参见图 3，是本申请实施例提供的一种语音数据处理装置的结构示意图，本申请实施例的该装置可以设置在上述提及的电子设备中。本实施例中，该装置包括:

获取单元 301，用于获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成。

处理单元 302，用于按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则。

预测单元 303，用于对所述目标语音数据进行划分，得到多个语音数据段；将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率。

分割单元 304，用于根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

可选的，处理单元 302，用于按照预设步长对所述目标语音数据进行划分，得到多个语音数据段，所述多个语音数据段中每个语音数据段的长度大于所述预设步长。

可选的，训练单元 305，用于获取训练样本集合，所述训练样本集合包括样本音频数据的多个样本语音数据段，及每个样本语音数据段的标注标签，所述标注标签包括样本语音数据段为边界点的概率；以所述多个样本语音数据段为所述语音网络模型的输入，以所述每个样本语音数据段的标注标签为所述语音网络模型的训练目标，对所述语音网络模块进行迭代训练。

可选的，训练单元 305，用于将所述多个样本语音数据段输入到所述语音网络模型进行预测，得到每个样本语音数据段的预测标签；根据所述每个样本语音数据段的预测标签及对应的样本语音数据段的标注标签确定所述语音网络模型的预测误差；若所述语音网络模型的预测误差大于预设误差值，则调整所述语音网络模型的网络参数；若所述语音网络模型的预测误差小于或等于所述预设误差值，则结束对所述语音网络模型的迭代训练。

可选的，训练单元 305，用于接收将目标样本语音数据段的标注标签设置为第一类标注标签的指令，所述第一类标注标签为样本语音数据段为边界点的概率大于预设概率的标签，所述目标样本语音数据段为样本音频数据的边界点所在的语音数据段；将与所述目标样本语音数据段之间的时间间隔小于或等于预设时间间隔的样本语音数据段的标注标签设置为第一类标注标签；将与所述目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段的标注标签设置为第二类标注标签，所述第二类标注标签为样本语音数据段为边界点的概率小于或等于所述预设概率的标签。

可选的，所述每个语音数据段包括多个特征点，每相邻两个语音数据段包括相同的特征点，所述每个语音数据段的预测标签为对应语音数据段内每个特征点的预测标签；分割单元 304，用于根据所述每个语音数据段的预测标签统计每个特征点为边界点的平均概率；将平均概率大于预设概率的所有特征点作为所述目标语音数据的边界点。

可选的，所述预设的处理规则包括数据过滤规则，处理单元 302，用于将所述待处理的语音数据进行划分，得到多个原始语音数据段；获取所述多个原始语音数据段中每个原始语音数据段的能量值；将所述多个原始语音数据段中能量值小于或等于预设能量值的原始语音数据段删除；将所述多个原始语音数据段中能量值大于所述预设能量值的原始语音数据段进行合并，得到所述目标语音数据。

本申请实施例中，可通过语音网络模型自动获取目标语音数据的边界点，不需要人工操作，可节省大量人力，并可提高获取语音数据的边界点的准确度及效率；满足用户对语音数据处理的智能化、自动化需求。

请参见图 4，是本申请实施例提供的一种电子设备的结构示意图，如图所示的本实施例中的电子设备可以包括：一个或多个处理器 401；一个或多个输入装置 402，一个或多个输出装置 403 和存储器 404。上述处理器 401、输入装置 402、输出装置 403 和存储器 404 通过总线 404 连接。

所处理器 401 可以是中央处理单元(Central Processing Unit，CPU)，该处理器还可以是其他通用处理器、数字信号处理器（Digital Signal Processor，DSP)、专用集成电路（Application Specific Integrated Circuit，ASIC)、现成可编程门阵列（Field-Programmable Gate Array，FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

输入装置 402 可以包括触控板、指纹采传感器（用于采集用户的指纹信息和指纹的方向信息）、麦克风等，输出装置 403 可以包括显示器（LCD 等）、扬声器等，输出装置 403 可以输出提示信息，提示信息可用于提示目标语音数据的边界点。

该存储器 404 可以包括只读存储器和随机存取存储器，并向处理器 401 提供指令和数据。存储器 404 的一部分还可以包括非易失性随机存取存储器，存储器 404 用于存储计算机程序，该计算机程序包括程序指令，处理器 401 用于执行存储器 404 存储的程序指令，以用于执行一种语音数据处理方法，即用于执行以下操作：

获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成；

按照预设的处理规则对所述待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则；

对所述目标语音数据进行划分，得到多个语音数据段；将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率；

根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

可选的，处理器 401 用于执行存储器 404 存储的程序指令，用于执行以下操作：

按照预设步长对所述目标语音数据进行划分，得到多个语音数据段，所述多个语音数据段中每个语音数据段的长度大于所述预设步长。

可选的，处理器 401 用于执行存储器 404 存储的程序指令，用于执行以下操作：

获取训练样本集合，所述训练样本集合包括样本音频数据的多个样本语音数据段，及每个样本语音数据段的标注标签，所述标注标签包括样本语音数据段为边界点的概率；

以所述多个样本语音数据段为所述语音网络模型的输入，以所述每个样本语音数据段的标注标签为所述语音网络模型的训练目标，对所述语音网络模块进行迭代训练。

可选的，处理器 401 用于执行存储器 404 存储的程序指令，用于执行以下操作：

将所述多个样本语音数据段输入到所述语音网络模型进行预测，得到每个样本语音数据段的预测标签；

根据所述每个样本语音数据段的预测标签及对应的样本语音数据段的标注标签确定所述语音网络模型的预测误差；

若所述语音网络模型的预测误差大于预设误差值，则调整所述语音网络模型的网络参数；

若所述语音网络模型的预测误差小于或等于所述预设误差值，则结束对所述语音网络模型的迭代训练。

可选的，处理器 401 用于执行存储器 404 存储的程序指令，用于执行以下操作：

接收将目标样本语音数据段的标注标签设置为第一类标注标签的指令，所述第一类标注标签为样本语音数据段为边界点的概率大于预设概率的标签，所述目标样本语音数据段为样本音频数据的边界点所在的语音数据段；

将与所述目标样本语音数据段之间的时间间隔小于或等于预设时间间隔的样本语音数据段的标注标签设置为第一类标注标签；

将与所述目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段的标注标签设置为第二类标注标签，所述第二类标注标签为样本语音数据段为边界点的概率小于或等于所述预设概率的标签。

可选的，处理器 401 用于执行存储器 404 存储的程序指令，用于执行以下操作：

根据所述每个语音数据段的预测标签统计每个特征点为边界点的平均概率；

将平均概率大于预设概率的所有特征点作为所述目标语音数据的边界点。

可选的，处理器 401 用于执行存储器 404 存储的程序指令，用于执行以下操作：

将所述待处理的语音数据进行划分，得到多个原始语音数据段；

获取所述多个原始语音数据段中每个原始语音数据段的能量值；

将所述多个原始语音数据段中能量值小于或等于预设能量值的原始语音数据段删除；

将所述多个原始语音数据段中能量值大于所述预设能量值的原始语音数据段进行合并，得到所述目标语音数据。

本申请实施例中所描述的处理器 401、输入装置 402、输出装置 403 可执行本申请实施例提供的语音数据处理方法的第一实施例和第二实施例中所描述的

实现方式，也可执行本申请实施例所描述的电子设备的实现方式，在此不再赘述。

本申请实施例中提供还了一种计算机可读存储介质，该计算机可读存储介质存储有计算机程序，该计算机程序包括程序指令，该程序指令被处理器执行时实现本申请的图 1 及图 2 实施例中所示的语音数据处理方法。

该计算机可读存储介质可以是前述任一实施例该的医疗管理设备的内部存储单元，例如控制设备的硬盘或内存。该计算机可读存储介质也可以是该控制设备的外部存储设备，例如该控制设备上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。进一步地，该计算机可读存储介质还可以既包括该控制设备的内部存储单元也包括外部存储设备。该计算机可读存储介质用于存储该计算机程序以及该控制设备所需的其他程序和数据。该计算机可读存储介质还可以用于暂时地存储已经输出或者将要输出的数据。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的控制设备和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的控制设备和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例是示意性的，例如，该单元的划分，可以为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。

以上该，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权 利 要 求

1、一种语音数据处理方法，其特征在于，包括:

获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成;

按照预设的处理规则对所述待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则;

对所述目标语音数据进行划分，得到多个语音数据段;将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率;

根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

2、根据权利要求 1 所述的方法，其特征在于，所述对所述目标语音数据进行划分，得到多个语音数据段，包括:

按照预设步长对所述目标语音数据进行划分，得到多个语音数据段，所述多个语音数据段中每个语音数据段的长度大于所述预设步长。

3、根据权利要求 1 所述的方法，其特征在于，所述方法还包括:

获取训练样本集合，所述训练样本集合包括样本音频数据的多个样本语音数据段，及每个样本语音数据段的标注标签，所述标注标签包括样本语音数据段为边界点的概率;

以所述多个样本语音数据段为所述语音网络模型的输入，以所述每个样本语音数据段的标注标签为所述语音网络模型的训练目标，对所述语音网络模块进行迭代训练。

4、根据权利要求 3 所述的方法，其特征在于，所述以所述多个样本语音数据段为所述语音网络模型的输入，以所述每个样本语音数据段的标注标签为所述语音网络模型的训练目标，对所述语音网络模块进行迭代训练，包括:

将所述多个样本语音数据段输入到所述语音网络模型进行预测，得到每个样本语音数据段的预测标签;

根据所述每个样本语音数据段的预测标签及对应的样本语音数据段的标注标签确定所述语音网络模型的预测误差;

若所述语音网络模型的预测误差大于预设误差值，则调整所述语音网络模型的网络参数;

若所述语音网络模型的预测误差小于或等于所述预设误差值，则结束对所述语音网络模型的迭代训练。

5、根据权利要求 3 或 4 所述的方法，其特征在于，所述方法还包括:

接收将目标样本语音数据段的标注标签设置为第一类标注标签的指令，所述第一类标注标签为样本语音数据段为边界点的概率大于预设概率的标签，所述目标样本语音数据段为样本音频数据的边界点所在的语音数据段;

将与所述目标样本语音数据段之间的时间间隔小于或等于预设时间间隔的样本语音数据段的标注标签设置为第一类标注标签;

将与所述目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段的标注标签设置为第二类标注标签，所述第二类标注标签为样本语音数据段为边界点的概率小于或等于所述预设概率的标签。

6、根据权利要求 1 所述的方法，其特征在于，所述每个语音数据段包括多个特征点，每相邻两个语音数据段包括相同的特征点，所述每个语音数据段的预

测标签为对应语音数据段内每个特征点的预测标签;

所述根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，包括:

根据所述每个语音数据段的预测标签统计每个特征点为边界点的平均概率;

将平均概率大于预设概率的所有特征点作为所述目标语音数据的边界点。

7、根据权利要求 1 所述的方法，其特征在于，所述预设的处理规则包括数据过滤规则，所述按照预设的处理规则对所述待处理的语音数据进行处理，得到目标语音数据， 包括:

将所述待处理的语音数据进行划分，得到多个原始语音数据段;

获取所述多个原始语音数据段中每个原始语音数据段的能量值;

将所述多个原始语音数据段中能量值小于或等于预设能量值的原始语音数据段删除;

将所述多个原始语音数据段中能量值大于所述预设能量值的原始语音数据段进行合并，得到所述目标语音数据。

8、一种语音数据处理装置，其特征在于，包括:

获取单元，用于获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成;

处理单元，用于按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则;

预测单元，用于对所述目标语音数据进行划分，得到多个语音数据段; 将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率;

分割单元，用于根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

9、根据权利要求 8 所述的装置，其特征在于，所述预测单元，具体用于按照预设步长对所述目标语音数据进行划分，得到多个语音数据段，所述多个语音数据段中每个语音数据段的长度大于所述预设步长。

10、根据权利要求 8 所述的装置，其特征在于，所述装置还包括:

训练单元，用于获取训练样本集合，所述训练样本集合包括样本音频数据的多个样本语音数据段，及每个样本语音数据段的标注标签，所述标注标签包括样本语音数据段为边界点的概率; 以所述多个样本语音数据段为所述语音网络模型的输入，以所述每个样本语音数据段的标注标签为所述语音网络模型的训练目标，对所述语音网络模块进行迭代训练。

11、根据权利要求 10 所述的装置，其特征在于，所述训练单元，具体用于将所述多个样本语音数据段输入到所述语音网络模型进行预测，得到每个样本语音数据段的预测标签; 根据所述每个样本语音数据段的预测标签及对应的样本语音数据段的标注标签确定所述语音网络模型的预测误差; 若所述语音网络模型的预测误差大于预设误差值，则调整所述语音网络模型的网络参数; 若所述语音网络模型的预测误差小于或等于所述预设误差值，则结束对所述语音网络模型的迭代训练。

12、根据权利要求 10 或 11 所述的装置，其特征在于，

所述训练单元，还用于接收将目标样本语音数据段的标注标签设置为第一

类标注标签的指令，所述第一类标注标签为样本语音数据段为边界点的概率大于预设概率的标签，所述目标样本语音数据段为样本音频数据的边界点所在的语音数据段；将与所述目标样本语音数据段之间的时间间隔小于或等于预设时间间隔的样本语音数据段的标注标签设置为第一类标注标签；将与所述目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段的标注标签设置为第二类标注标签，所述第二类标注标签为样本语音数据段为边界点的概率小于或等于所述预设概率的标签。

13、根据权利要求 8 所述的装置，其特征在于，所述每个语音数据段包括多个特征点，每相邻两个语音数据段包括相同的特征点，所述每个语音数据段的预测标签为对应语音数据段内每个特征点的预测标签；

所述预测单元，具体用于根据所述每个语音数据段的预测标签统计每个特征点为边界点的平均概率；将平均概率大于预设概率的所有特征点作为所述目标语音数据的边界点。

14、根据权利要求 8 所述的装置，其特征在于，所述预设的处理规则包括数据过滤规则，所述处理单元，具体用于将所述待处理的语音数据进行划分，得到多个原始语音数据段；获取所述多个原始语音数据段中每个原始语音数据段的能量值；将所述多个原始语音数据段中能量值小于或等于预设能量值的原始语音数据段删除；将所述多个原始语音数据段中能量值大于所述预设能量值的原始语音数据段进行合并，得到所述目标语音数据。

15、一种电子设备，其特征在于，包括：

处理器，适于实现一条或一条以上指令；以及，

计算机可读存储介质，所述计算机可读存储介质存储有一条或一条以上指令，所述一条或一条以上指令适于由所述处理器加载并执行如下步骤：获取待处理的语音数据，所述待处理的语音数据由多个对象的语音数据段组成；

按照预设的处理规则对所述待处理的语音数据进行处理，得到目标语音数据，所述预设的处理规则包括数据过滤规则和/或数据格式处理规则；

对所述目标语音数据进行划分，得到多个语音数据段；将所述多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签，所述预测标签包括语音数据段为边界点的概率；

根据所述每个语音数据段的预测标签确定所述目标语音数据的边界点，以根据所述边界点从所述目标语音数据中分割出每个对象的语音数据段。

16、根据权利要求 15 所述的电子设备，其特征在于，所述处理器，具体用于按照预设步长对所述目标语音数据进行划分，得到多个语音数据段，所述多个语音数据段中每个语音数据段的长度大于所述预设步长。

17、根据权利要求 15 所述的电子设备，其特征在于，所述处理器，具体用于获取训练样本集合，所述训练样本集合包括样本音频数据的多个样本语音数据段，及每个样本语音数据段的标注标签，所述标注标签包括样本语音数据段为边界点的概率；以所述多个样本语音数据段为所述语音网络模型的输入，以所述每个样本语音数据段的标注标签为所述语音网络模型的训练目标，对所述语音网络模块进行迭代训练。

18、根据权利要求 17 所述的电子设备，其特征在于，所述处理器，具体用于将所述多个样本语音数据段输入到所述语音网络模型进行预测，得到每个样本语音数据段的预测标签；根据所述每个样本语音数据段的预测标签及对应的样本语音数据段的标注标签确定所述语音网络模型的预测误差；若所述语音网络模型

的预测误差大于预设误差值，则调整所述语音网络模型的网络参数；若所述语音网络模型的预测误差小于或等于所述预设误差值，则结束对所述语音网络模型的迭代训练。

19、根据权利要求 17 或 18 所述的电子设备，其特征在于，所述处理器，还用于接收将目标样本语音数据段的标注标签设置为第一类标注标签的指令，所述第一类标注标签为样本语音数据段为边界点的概率大于预设概率的标签，所述目标样本语音数据段为样本音频数据的边界点所在的语音数据段；将与所述目标样本语音数据段之间的时间间隔小于或等于预设时间间隔的样本语音数据段的标注标签设置为第一类标注标签；将与所述目标样本语音数据段之间的时间间隔大于预设时间间隔的样本语音数据段的标注标签设置为第二类标注标签，所述第二类标注标签为样本语音数据段为边界点的概率小于或等于所述预设概率的标签。

20、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有一条或一条以上指令，所述一条或一条以上指令适于由处理器加载并执行如权利要求 1-7 任一项所述的语音数据处理方法。

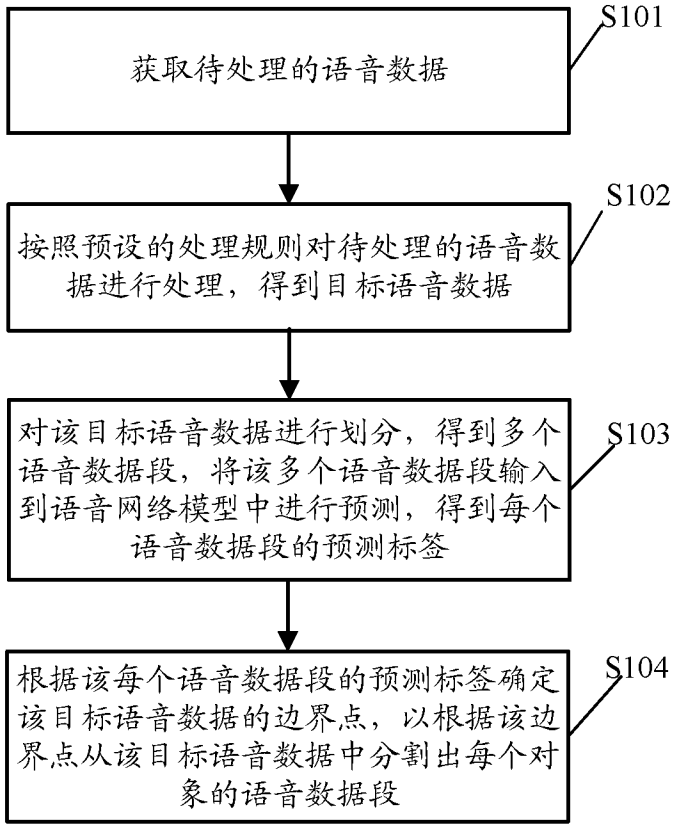
获取待处理的语音数据
S101
按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据
S102
对该目标语音数据进行划分，得到多个语音数据段，将该多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签
S103
根据该每个语音数据段的预测标签确定该目标语音数据的边界点，以根据该边界点从该目标语音数据中分割出每个对象的语音数据段
S104

图 1

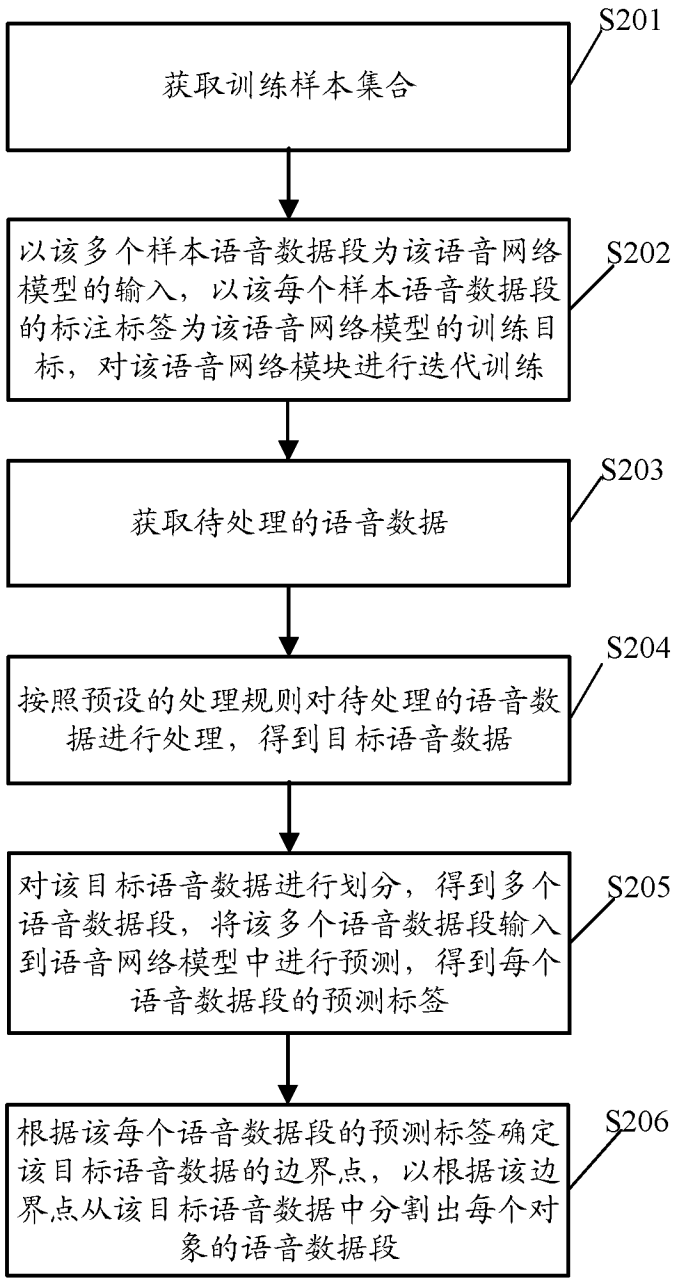
S201
获取训练样本集合
以该多个样本语音数据段为该语音网络模型的输入，以该每个样本语音数据段的标注标签为该语音网络模型的训练目标，对该语音网络模块进行迭代训练
S202
S203
获取待处理的语音数据
S204
按照预设的处理规则对待处理的语音数据进行处理，得到目标语音数据
对该目标语音数据进行划分，得到多个语音数据段，将该多个语音数据段输入到语音网络模型中进行预测，得到每个语音数据段的预测标签
S205
根据该每个语音数据段的预测标签确定该目标语音数据的边界点，以根据该边界点从该目标语音数据中分割出每个对象的语音数据段
S206

图 2

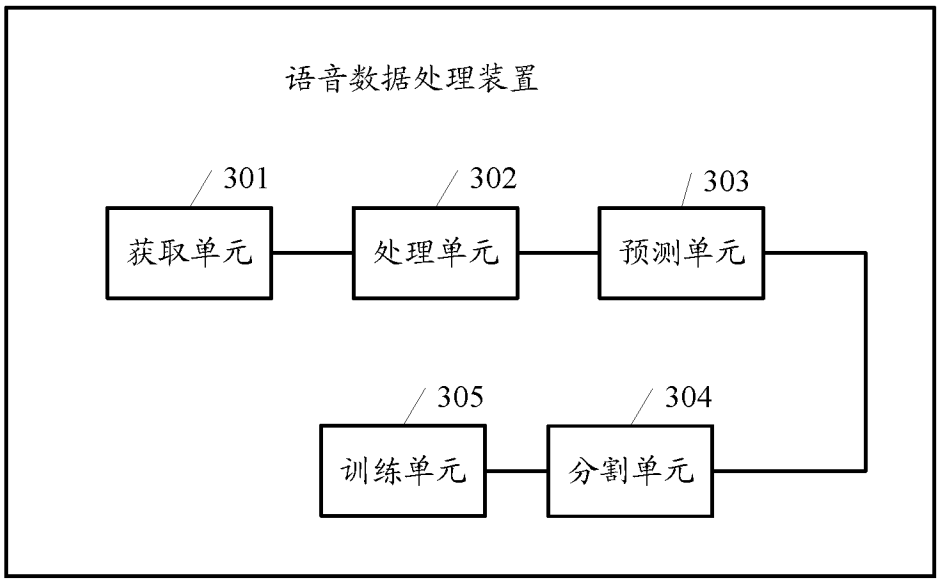
语音数据处理装置
301
获取单元
302
处理单元
303
预测单元
305
训练单元
304
分割单元

图 3

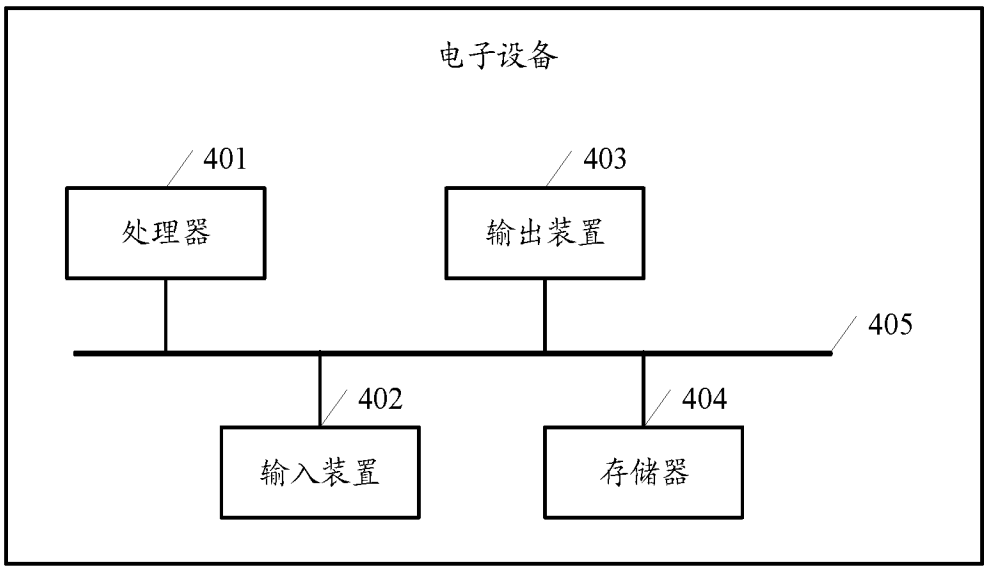
电子设备
401
处理器
403
输出装置
405
402
输入装置
404
存储器

图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/088976

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/00(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 17; G10L 15

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI; CNABS; CNTXT, IEEE, 百度学术, BAIDU: 语音, 音频, 分割, 切分, 分界点, 说话人, 改变, 统计模型, 神经网络, audio, speech, speaker, separate, segment, speaker change, change point, split, speaker diarization

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109616097 A (PING'AN TECHNOLOGY (SHENZHEN) CO., LTD.) 12 April 2019 (2019-04-12) claims 1-10	1-20
Y	YIN, R. et al. "Speaker Change Detection in Broadcast TV Using Bidirectional Long Short-Term Memory Networks" *Proc. Interspeech 2017,* 24 August 2017 (2017-08-24), parts 3 and 5.1, and figures 1 and 2	1-20
Y	CN 106782507 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 31 May 2017 (2017-05-31) description, paragraphs 56 and 60	1-20
A	US 2016111112 A1 (FUJITSU LTD.) 21 April 2016 (2016-04-21) entire document	1-20
A	US 2018039888 A1 (INTERACTIVE INTELLIGENCE GROUP INC.) 08 February 2018 (2018-02-08) entire document	1-20

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 September 2019	**10 October 2019**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/088976

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GE, Zhenhao et al. "Speaker Change Detection using Features through a Neural Network Speaker Classifier" *Intelligent Systems Conference 2017,* 08 September 2017 (2017-09-08), entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/088976

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109616097	A	12 April 2019	None			
CN	106782507	A	31 May 2017	TW	I643184	B	01 December 2018
				CN	106782507	B	06 March 2018
				TW	201824250	A	01 July 2018
				WO	2018113243	A1	28 June 2018
US	2016111112	A1	21 April 2016	JP	6303971	B2	04 April 2018
				JP	2016080916	A	16 May 2016
				US	9536547	B2	03 January 2017
US	2018039888	A1	08 February 2018	None			

国际检索报告

国际申请号
PCT/CN2019/088976

A. 主题的分类

G10L 17/00(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G10L 17; G10L 15

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

DWPI;CNABS;CNTXT, IEEE, 百度学术:语音，音频，分割，切分，分界点，说话人，改变，统计模型，神经网络，audio, speech, speaker, separate, segment,speaker change, change point, split, speaker diarization

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 109616097 A (平安科技深圳有限公司) 2019年 4月 12日 (2019-04-12) 权利要求1-10	1-20
Y	Yin, R.等. "Speaker Change Detection in Broadcast TV Using Bidirectional Long Short-Term Memory Networks" Proc. Interspeech 2017, 2017年 8月 24日 (2017-08-24), 第3部分，第5.1部分，图1，图2	1-20
Y	CN 106782507 A (平安科技深圳有限公司) 2017年 5月 31日 (2017-05-31) 说明书第56，60段	1-20
A	US 2016111112 A1 (FUJITSU LTD.) 2016年 4月 21日 (2016-04-21) 全文	1-20
A	US 2018039888 A1 (INTERACTIVE INTELLIGENCE GROUP INC.) 2018年 2月 8日 (2018-02-08) 全文	1-20
A	Zhenhao Ge 等. "Speaker Change Detection using Features through a Neural Network Speaker Classifier" Intelligent Systems Conference 2017, 2017年 9月 8日 (2017-09-08), 全文	1-20

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2019年 9月 19日	2019年 10月 10日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 游晓梅 电话号码 (86-10)62089539

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/088976

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109616097	A	2019年 4月 12日	无			
CN	106782507	A	2017年 5月 31日	TW	I643184	B	2018年 12月 1日
				CN	106782507	B	2018年 3月 6日
				TW	201824250	A	2018年 7月 1日
				WO	2018113243	A1	2018年 6月 28日
US	2016111112	A1	2016年 4月 21日	JP	6303971	B2	2018年 4月 4日
				JP	2016080916	A	2016年 5月 16日
				US	9536547	B2	2017年 1月 3日
US	2018039888	A1	2018年 2月 8日	无			