

(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



(10) 国际公布号
WO 2022/134894 A1

(43) 国际公布日
2022 年 6 月 30 日 (30.06.2022)

- (51) 国际专利分类号:
G10L 15/16 (2006.01)
- (21) 国际申请号: PCT/CN2021/129223
- (22) 国际申请日: 2021 年 11 月 8 日 (08.11.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202011536771.4 2020年12月23日 (23.12.2020) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。
- (72) 发明人: 孙思宁 (SUN, Sining); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。
- (74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路 6 号 4501 房 (部位: 自编 01-03 和 08-12 单元) (仅限办公用途), Guangdong 510623 (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: SPEECH RECOGNITION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 语音识别方法、装置、计算机设备及存储介质

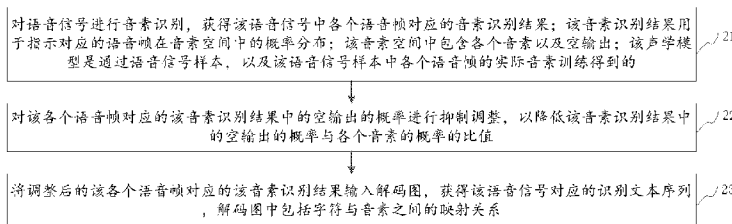


图 2

- 21 Perform phoneme recognition on a speech signal to obtain a phoneme recognition result corresponding to each speech frame in the speech signal, the phoneme recognition result being used for indicating a probability distribution of the corresponding speech frame in a phoneme space, the phoneme space containing phonemes and null outputs, and an acoustic model being obtained by training a speech signal sample and an actual phoneme of each speech frame in the speech signal sample
- 22 Perform suppression adjustment on a probability of the null outputs in the phoneme recognition result corresponding to each speech frame, so as to reduce a ratio of the probability of the null outputs in the phoneme recognition result to a probability of each phoneme
- 23 Input the adjusted phoneme recognition result corresponding to each speech frame into a decoding graph to obtain a recognition text sequence corresponding to the speech signal, the decoding graph containing a mapping relationship between each character and each phoneme

(57) Abstract: A speech recognition method and apparatus, a computer device, and a storage medium. The method comprises: performing phoneme recognition on a speech signal to obtain a phoneme recognition result corresponding to each speech frame in the speech signal (21); performing suppression adjustment on a probability of null outputs in the phoneme recognition result corresponding to each speech frame, so as to reduce a ratio of the probability of the null outputs in the phoneme recognition result to a probability of each phoneme (22); and inputting the adjusted phoneme recognition result corresponding to each speech frame into a decoding graph to obtain a recognition text sequence corresponding to the speech signal (23). The method can improve the recognition accuracy of a model in a speech recognition scenario in the field of artificial intelligence.

(57) 摘要: 一种语音识别方法、装置、计算机设备及存储介质, 该方法包括: 对语音信号进行音素识别, 获得语音信号中各个语音帧对应的音素识别结果 (21); 对各个语音帧对应的音素识别结果中的空输出的概率进行抑制调整, 以降低音素识别结果中的空输出的概率与各个音素的概率的比值 (22); 将调整后的各个语音帧对应的音素识别结果输入解码图, 获得语音信号对应的识别文本序列 (23)。该方法能够在人工智能领域的语音识别场景中, 提高模型的识别准确性。

WO 2022/134894 A1

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

发明名称：语音识别方法、装置、计算机设备及存储介质

本申请要求于 2020 年 12 月 23 日提交中国专利局，申请号为 202011536771.4，申请名称为“语音识别方法、装置、计算机设备及存储介质”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

本申请涉及语音识别技术领域，特别涉及一种语音识别方法、装置、计算机设备及存储介质。

背景技术

语音识别是一种将语音识别为文本的技术，其在各种人工智能（Artificial Intelligence, AI）场景中具有广泛的应用。

语音识别框架通常包含声学模型部分和解码部分，其中，声学模型部分用于识别输入的语音信号中的各个语音帧的音素，解码部分通过识别出的各个语音帧的音素，输出语音信号的文本序列。在相关技术中，通过递归神经网络转移机（Recurrent Neural Network Transducer, RNN-T）来实现声学模型，是业内研究的重点之一。

然而，RNN-T 模型在音素识别过程中引入了空输出的概念，即预测某个语音帧上不包含有效的音素，空输出的引入，在一些应用场景中会导致后续解码过程的错误率上升，尤其是导致删除错误的增多，影响语音识别的准确性。

发明内容

根据本申请提供的各种实施例，提供了一种语音识别方法、装置、计算机设备及存储介质

一种语音识别方法，由计算机设备执行，所述方法包括：

对语音信号进行音素识别，获得所述语音信号中各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；

对所述各个语音帧对应的所述音素识别结果中的空输出的概率进行抑制调整，以降低所述音素识别结果中的空输出的概率与各个音素的概率的比值；及

将调整后的所述各个语音帧对应的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

一种语音识别装置，所述装置包括：

语音信号处理模块，用于对语音信号进行音素识别，获得所述语音信号中各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；

概率调整模块，用于对所述各个语音帧对应的所述音素识别结果中的空输出的概率进行抑制调整，以降低所述音素识别结果中的空输出的概率与各个音素的概率的比值；及

解码模块，用于将调整后的所述各个语音帧对应的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

一种语音识别方法，所述方法包括：

获取语音信号，所述语音信号包括对原始语音进行切分获得的各个语音帧；

对语音信号进行音素识别，获得所述各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；及

将所述各个语音帧对应的所述音素识别结果中，空输出的概率满足指定条件的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

一种语音识别装置，所述装置包括：

语音信号获取模块，用于获取语音信号，所述语音信号包括对原始语音进行切分获得的各个语音帧；

音素识别结果获得模块，用于对语音信号进行音素识别，获得所述各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；及

识别文本序列获取模块，用于将所述各个语音帧对应的所述音素识别结果中，空输出的概率满足指定条件的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

一种计算机设备，所述计算机设备包含处理器和存储器，所述存储器中存储有至少一条计算机指令，所述至少一条计算机指令由所述处理器加载并执行以实现上述的语音识别方法。

一种计算机可读存储介质，所述存储介质中存储有至少一条计算机指令，所述至少一条计算机指令由处理器加载并执行以实现上述语音识别方法。

一种计算机程序产品或计算机程序，该计算机程序产品或计算机程序包括计算机指令，该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令，处理器执行该计算机指令，使得该计算机设备执行上述计算机程序方法。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1是本申请各个实施例涉及的一种语音识别系统的系统构成图；

图2是根据一示例性实施例示出的语音识别方法的流程示意图；

图3是根据一示例性实施例示出的语音识别方法的流程示意图；

图4是图3所示实施例涉及的对齐过程示意图；

图5是图3所示实施例涉及的一种声学模型的结构示意图；

- 图 6 是图 3 所示实施例涉及的预测器的网络结构图；
图 7 是图 3 所示实施例涉及的模型训练及应用流程图；
图 8 是根据一示例性实施例示出的语音识别系统的框架图；
图 9 是根据一示例性实施例示出的视频中的对象标注装置的结构方框图；
图 10 是根据一示例性实施例示出的一种计算机设备的结构框图。

具体实施方式

这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。

在对本申请所示的各个实施例进行说明之前，首先对本申请涉及到的几个概念进行介绍：

1) 人工智能 (Artificial Intelligence, AI)

AI 是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说，人工智能是计算机科学的一个综合技术，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。

人工智能技术是一门综合学科，涉及领域广泛，既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

2) 语音技术 (Speech Technology, ST)

语音技术的关键技术有自动语音识别技术 (Automatic Speech Recognition, ASR) 和语音合成技术 (Text To Speech, TTS) 以及声纹识别技术。让计算机能听、能看、能说、能感觉，是未来人机交互的发展方向，其中语音成为未来最被看好的人机交互方式之一。

3) 机器学习 (Machine Learning, ML)

机器学习是一门多领域交叉学科, 涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为, 以获取新的知识或技能, 重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心, 是使计算机具有智能的根本途径, 其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、示教学习等技术。

本申请实施例提供的方案应用在涉及人工智能的语音技术和机器学习技术等场景, 以实现将用户语音准确的识别为对应的文本。

参考图 1, 展示了本申请各个实施例涉及的一种语音识别系统的系统构成图。如图 1 所示, 该系统包括声音采集组件 120、以及语音识别设备 140。

其中, 声音采集组件 120 和语音识别设备 140 之间通过有线或者无线方式相连。

声音采集组件 120 可以实现为麦克风、麦克风阵列或者拾音器等。声音采集组件 120 用于采集用户说话时的语音数据。

语音识别设备 140 用于对声音采集组件 120 采集到的语音数据进行识别, 获得识别出的文本序列。

可选的, 语音识别设备 140 还可以对识别出的文本序列进行自然语义处理, 以对用户语音做出响应。

其中, 声音采集组件 120 和语音识别设备 140 可以实现为相互独立的两个硬件设备。例如, 声音采集组件 120 是设置在车辆方向盘上的麦克风, 语音识别设备 140 可以是车载智能设备; 或者, 声音采集组件 120 是设置在遥控器上的麦克风, 语音识别设备 140 可以是遥控器控制的智能家居设备 (比如智能电视、机顶盒、空调等等)。

或者, 声音采集组件 120 和语音识别设备 140 可以实现为同一个硬件设备。例如, 语音识别设备 140 可以是智能手机、平板电脑、智能手表、智能眼镜等智能设备, 而声音采集组件 120 可以是语音识别设备 140 内置的麦克风。

在一些实施例中, 上述语音识别系统还可以包含服务器 160。

其中, 该服务器 160 可以用于对语音识别设备 140 中的语音识别模型进行部署和更新。或者, 服务器 160 也可以向语音识别设备 140 提供云端语音识别

的服务，即接收语音识别设备 140 发送的语音数据，对语音数据进行语音识别后，将识别结果返回给语音识别设备 140。或者，服务器 160 也可以与语音识别设备 140 协作完成对语音数据的识别以及对语音数据的响应等操作。

服务器 160 是一台服务器，或者由若干台服务器，或者是一个虚拟化平台，或者是一个云计算服务中心。

服务器可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络（Content Delivery Network, CDN）、以及大数据和人工智能平台等基础云计算服务的云服务器。

服务器 160 与语音识别设备 140 之间通过通信网络相连。可选的，该通信网络是有线网络或无线网络。

可选的，该系统还可以包括管理设备（图 1 未示出），该管理设备与服务器 160 之间通过通信网络相连。可选的，通信网络是有线网络或无线网络。

可选的，上述的无线网络或有线网络使用标准通信技术和/或协议。网络通常为因特网、但也可以是任何网络，包括但不限于局域网（Local Area Network, LAN）、城域网（Metropolitan Area Network, MAN）、广域网（Wide Area Network, WAN）、移动、有线或者无线网络、专用网络或者虚拟专用网络的任何组合。在一些实施例中，使用包括超文本标记语言（Hyper Text Mark-up Language, HTML）、可扩展标记语言（Extensible Markup Language, XML）等的技术和/或格式来代表通过网络交换的数据。此外还可以使用诸如安全套接字层（Secure Socket Layer, SSL）、传输层安全（Transport Layer Security, TLS）、虚拟专用网络（Virtual Private Network, VPN）、网际协议安全（Internet Protocol Security, IPsec）等常规加密技术来加密所有或者一些链路。在另一些实施例中，还可以使用定制和/或专用数据通信技术取代或者补充上述数据通信技术。

参考图 2，展示了一种语音识别方法的流程示意图，该语音识别方法可以由计算机设备执行。该计算机设备可以是上述图 1 所示系统中的语音识别设备 140 或者服务器 160，或者，该计算机设备可以同时包括上述图 1 所示系统中的语音识别设备 140 和服务器 160。如图 2 所示，该语音识别方法可以包括如下步骤：

步骤 21，对语音信号进行音素识别，获得该语音信号中各个语音帧对应的

音素识别结果；该音素识别结果用于指示对应的语音帧在音素空间中的概率分布；该音素空间中包含各个音素以及空输出；

其中，音素识别结果可以通过声学模型对语音信号进行音素识别所得到的结果。该声学模型是通过语音信号样本，以及该语音信号样本中各个语音帧的实际音素训练得到的。

音素（phone），是根据语音的自然属性划分出来的最小语音单位，依据音节里的发音动作来分析，一个动作构成一个音素。音素分为元音与辅音两大类。比如，汉语音节啊（ā）只有一个音素，爱（ài）有两个音素，代（dài）有三个音素等。

音素是构成音节的最小单位或最小的语音片段，是从音质的角度划分出来的最小的线性的语音单位。音素是具体存在的物理现象。国际音标（由国际语音学会制定，用来统一标示各国语音的字母。也称为“国际语音学字母”、“万国语音学字母”）的音标符号与全人类语言的音素一一对应。

音素空间中包括的空输出的数量可以大于或者等于 1，例如包括一个空输出。

在本申请实施例中，对于语音信号中的每个语音帧，声学模型可以识别该语音帧对应的音素，得到该语音帧的音素属于各个预先设置的音素以及空输出的概率。

例如，在一些实施例中，上述音素空间包含 212 种音素以及一个空输出（表示对应的语音帧没有用户发音），也就是说，对于一个输入的语音帧，本申请实施例所示的声学模型，能够输出该语音帧分别对应 212 种音素以及空输出的概率。

步骤 22，对该各个语音帧对应的该音素识别结果中的空输出的概率进行抑制调整，以降低该音素识别结果中的空输出的概率与各个音素的概率的比值。

步骤 23，将调整后的该各个语音帧对应的该音素识别结果输入解码图，获得该语音信号对应的识别文本序列。

其中，解码图用于基于音素识别结果确定语音帧对应的音素，解码图中可以包括字符与音素之间的映射关系，一个字符可以是一个汉字，也可以是一个词。

在本申请实施例中，音素识别结果输入解码图后，基于解码图根据音素识别结果中的音素空间中各个音素以及空输出的概率，确定该音素识别结果对应

的是某一个音素或者对应的是空输出，并根据确定音素确定对应的文本，如果音素识别结果对应的是空输出，则确定该音素识别结果对应的语音帧中不包含用户发音，即没有对应的文本。

语音信号中的语音帧可以是对声音采集组件采集到的原始语音进行切分获得的各个语音帧。语音信号中的各个语音帧有序排列，例如可以按照语音帧在原始语音中的位置进行排列，例如，语音帧在原始语音中的位置越靠前，则该语音帧在语音信号中的位置越靠前。语音信息中的各个语音帧依次排列，将解码图得到的各个语音帧的文本，按照语音帧在语音信号中的位置进行排列，得到识别文本序列。

由于上述音素识别结果中包含空输出，可能会导致识别错误率上升，例如，有可能会出现某个有发音的语音帧被误识别为空输出的情况（这种情况也被称为删除错误），从而影响语音识别的准确性，对此，本申请实施例所示的方案，在声学模型输出音素识别结果之后，对音素识别结果中的空输出的概率进行抑制，而随着音素识别结果中的空输出的概率被抑制，该音素识别结果被识别为某一个音素的可能性也随之上升，从而能够有效的减少有发音的语音帧被误识别为空输出的情况。

综上所述，本申请实施例所示的方案，对于包含语音帧在各个音素以及空输出上的概率分布的音素识别结果，在将该音素识别结果输入解码图之前，先对音素识别结果中的空输出的概率进行抑制，降低语音帧被识别为空输出的几率，从而降低语音帧被误识别为空输出的可能性，也就是降低模型的删除错误，从而提高模型的识别准确性。

参考图 3，展示了一种语音识别方法的流程示意图，该语音识别方法可以由计算机设备执行。比如，该计算机设备可以是上述图 1 所示系统中的语音识别设备 140 或者服务器 160，或者，该计算机设备可以同时包括上述图 1 所示系统中的语音识别设备 140 和服务器 160。如图 3 所示，该语音识别方法可以包括如下步骤：

步骤 301，获取语音信号，该语音信号包括对原始语音进行切分获得的各个语音帧。

在本申请实施例中，声音采集组件采集到的用户说话过程中的原始语音后，将采集到的原始语音发送给计算机设备，比如，发送给语音识别设备，语音识

别设备对原始语音进行切分，获得若干个语音帧。

在一些实施例中，语音识别设备可以将原始语音切分成带有重叠的短时语音片段，例如，一般对于采样率为 16K 的语音，切分后一帧语音长度为 25ms，帧间重叠为 15ms，此过程也称为“分帧”。

步骤 302，对语音信号进行音素识别，获得该语音信号中各个语音帧对应的音素识别结果。

其中，该音素识别结果用于指示对应的语音帧在音素空间中的概率分布；该音素空间中包含各个音素以及一个空输出；该声学模型是通过语音信号样本，以及该语音信号样本中各个语音帧的实际音素训练得到的。

在本申请实施例中，声学模型为端到端的机器学习模型，其输入数据包括语音信号中的语音帧（例如，输入包括语音帧的特征向量），输出数据为预测出的该语音帧的音素在音素空间中的分布概率，即音素识别结果。

例如，上述音素识别结果可以表示为如下所示的一个概率向量：

$$(p_0, p_1, p_2, \dots, p_{212})$$

上述概率向量中， p_0 表示语音帧为空输出的概率， p_1 表示语音帧对应第 1 种音素的概率，整个音素空间中包含 212 种音素，外加一个空输出。

在一些实施例中，对语音信号进行音素识别，获得该语音信号中各个语音帧对应的音素识别结果，包括：

通过已训练的声学模型对目标语音帧进行特征提取，获得该目标语音帧的特征向量；该目标语音帧是该各个语音帧中的任意一个；

将该目标语音帧输入该声学模型中的编码器，获得该目标语音帧的声学隐层表示向量；

将该目标语音帧的历史识别文本的音素信息输入该声学模型中的预测器，获得该目标语音帧的文本隐层表示向量；该目标语音帧的历史识别文本，是该解码器对该目标语音帧的前 n 个非空输出的语音帧的音素识别结果进行识别得到的文本； n 为大于或者等于 1 的整数；

将该目标语音帧的声学隐层表示向量，以及该目标语音帧的文本隐层表示向量输入联合网络，获得该目标语音帧的该音素识别结果。

本申请实施例中，可以通过转换器（Transducer）模型，实现上述声学模型。Transducer 模型介绍如下：

给定输入序列：

$$\mathbf{x} = (x_1, x_2, \dots, x_T) \in \mathcal{X}^*$$

和输出序列:

$$\mathbf{y} = (y_1, y_2, \dots, y_U) \in \mathcal{Y}^*$$

其中, \mathcal{X}^* 表示所有的输入序列的集合, \mathcal{Y}^* 表示所有输出序列的集合, $x_t \in \mathcal{X}$, $y_u \in \mathcal{Y}$ 均为实数向量, \mathcal{X} 和 \mathcal{Y} 分别表示输入和输出空间。例如, 在本方案中, Transducer 模型用来进行音素识别, 输入序列 \mathbf{x} 为特征向量序列, 比如滤波器组 (Filter Bank, FBank) 特征, 或者, 梅尔倒谱系数 (Mel Frequency Cepstrum Coefficient, MFCC) 特征, x_t 表示 t 时刻的特征向量; 输出序列 \mathbf{y} 则为音素序列, y_u 表示第 u 步的音素。

定义一个扩展输出空间 $\bar{\mathcal{Y}} = \mathcal{Y} \cup \emptyset$, \emptyset 表示空输出符号, 代表模型没有输出。引入空输出符号之后, 序列 $(y_1, \emptyset, \emptyset, y_2, \emptyset, \emptyset, y_3) \in \bar{\mathcal{Y}}^*$ 就等价于 $(y_1, y_2, y_3) \in \mathcal{Y}^*$ 。在本方案中, 因为空输出的引入, 输出序列与输入序列将有相同的长度, 因此, 也将集合 $\bar{\mathcal{Y}}^*$ 中的元素 $\mathbf{a} \in \bar{\mathcal{Y}}^*$ 称为“对齐”。给定任意输入序列, Transducer 模型定义了一个条件分布 $Pr(\mathbf{a} \in \bar{\mathcal{Y}}^* | \mathbf{x})$, 此条件分布将用来计算给定输入序列 \mathbf{x} 后输出序列 \mathbf{y} 的概率:

$$Pr(\mathbf{y} \in \mathcal{Y}^* | \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{B}^{-1}(\mathbf{y})} Pr(\mathbf{a} | \mathbf{x}) \quad (1)$$

其中 $\mathcal{B}: \bar{\mathcal{Y}}^* \rightarrow \mathcal{Y}^*$ 表示移除对齐序列中的空输出, $\mathcal{B}^{-1}: \mathcal{Y}^* \rightarrow \bar{\mathcal{Y}}^*$ 表示对输出序列添加空输出生成对齐序列。从公式 (1) 中可以看出, 为了计算输出序列 \mathbf{y} 的概率, 需要将序列 \mathbf{y} 所对应的所有可能的对齐 \mathbf{a} 的条件概率进行求和, 参考图 4, 其示出了本申请实施例涉及的对齐过程示意图。该图 4 给出一个示例用来说明公式 (1)。

在图 4 中, $U=3$, $T=5$, 从左下角到右上角所有可能的路径, 都是一个对齐。加粗箭头标出的为其中一条可能路径, 当模型沿着纵向前进一步时, 将输出一个非空符号 (音素); 当模型沿着横向前进一步时, 将输出空符号 (即上述空输出), 表示没有输出产生。在同一时刻, 模型允许多个输出产生。

为了建模 $Pr(\mathbf{a} \in \bar{\mathcal{Y}}^* | \mathbf{x})$, 一般采用三个子网络进行联合建模, 参考图 5, 其示出了本申请实施例涉及的一种声学模型的结构示意图。如图 5 所示。声学模型包含编码器 51、预测器 52、以及联合网络 53。

其中, 编码器 51 (Encoder), 可以为递归神经网络, 例如长短期记忆 (Long Short-Term Memory, LSTM) 网络, 接受 t 时刻的音频特征输入, 输出声学隐层表示 h_t^{enc} ;

预测器 52 (Predictor)，可以为递归神经网络，例如 LSTM，接受模型历史的非空输出标签 y_{u_i-1} ，输出为文本隐层表示 $h_{u_i}^{pre}$ ；

联合网络 53 (Joint Network)，可以为全连接神经网络，如线性层加激活单元，用于将 h_t^{enc} 和 $h_{u_i}^{pre}$ 经过线性变换之后求和，输出隐单元表示 z_i ；最后，经过一个 softmax 函数，将其转换成概率分布。

上述图 5 中， $\hat{y}_i \in \bar{Y}$ ，且对齐 $\mathbf{a} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{T+U})$ ，最终，公式 (1) 的计算为：

$$Pr(\mathbf{y} \in \mathcal{Y}^* | \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{B}^{-1}(\mathcal{Y})} Pr(\mathbf{a} | \mathbf{x}) = \prod_{i=1}^{T+U} Pr(\hat{y}_i | x_1, \dots, x_{t_i}, y_0, \dots, y_{u_i-1}) \quad (2)$$

公式 (2) 的计算，需要遍历计算所有可能的对齐路径，直接使用该算法将导致大量的计算量，在模型训练过程中，可以采用前后向算法来进行公式 (2) 的概率计算。

在一些实施例中，该编码器为前向序列记忆网络 (Feedforward Sequential Memory Networks, FSMN)。

在一些实施例中，该预测器为一维卷积网络。

本申请实施例所示的方案，可以应用在车载离线语音识别系统等计算能力有限的场景。车载设备对模型参数量和计算量要求高，中央处理器 (Central Processing Unit, CPU) 算力有限，因此，对模型参数量和模型结构要求较高。为了降低计算量，适应此类计算能力有限的应用场景，本申请所示的方案采用全前向神经网络 FSMN 作为模型的 Encoder (编码器)，并采用一维卷积网络替代常用的长短时记忆网络 LSTM 作为 Predictor (预测器)。

上述的 Transducer 模型，为了刻画模型的历史信息，Encoder 和 Predictor 网络一般采用循环神经网络 (Recurrent Neural Network, RNN) 结构，例如 LSTM 或者门控循环单元 (Gated Recurrent Unit, GRU)。但是在计算资源有限的嵌入式设备上，递归神经网络会带来大量的计算量，会带来大量的 CPU 资源的占用。另一方面，车载离线语音识别的内容主要为查询和控制类的指令，句子相对较短，无需太长的历史信息。对此，本方案使用了基于 FSMN 的 Encoder 和基于一维卷积的 Predictor 网络。一方面，可以压缩模型参数，另一方面，可以极大的节省计算资源，提高计算速度，保证语音识别的实时性。

本方案中，采用了基于 FSMN 的 Encoder 结构。FSMN 网络应用于大词汇量语音识别任务中。本方案中所采用的 FSMN 结构，可以为带有投影层和残差

连接的结构。

对于 Predictor 网络，本方案中采用了一维卷积网络，根据有限的历史预测输出，来产生当前的输出。参考图 6，其示出了本申请实施例涉及的预测器的网络结构图。如图 6 所示，Predictor 网络使用了 4 个非空的历史输出来预测当前输出的框架。也就是，将当前输入对应的 4 个非空的历史输出 61 经过向量映射后，输入到以为一维卷积网络 62 中，得到文本隐层表示向量。

在本申请实施例中，上述声学模型可以通过预先设置的语音样本、以及该语音信号样本中各个语音帧的实际音素训练得到。例如，在训练过程中，将语音样本中的一个语音帧输入到声学模型中基于 FSMN 的 Encoder 网络，并将该语音帧的前 4 个非空语音帧的实际音素（训练开始时刻不存在历史的非空语音帧，或者历史非空语音帧不足时，可以通过预先设置的音素来代替），输入到基于一维卷积的 Predictor 网络，在声学模型对输入数据进行处理的过程中，对声学模型中三个部分（Encoder、Predictor 以及联合网络）的参数进行更新，使得所有可能的对齐路径上的概率之和，也就是上述公式（2）的结果最大化，从而实现声学模型的训练。

步骤 303，对该各个语音帧对应的该音素识别结果中的空输出的概率进行抑制调整，以降低该音素识别结果中的空输出的概率与各个音素的概率的比值。

在一些实施例中，该对该各个语音帧对应的该音素识别结果中的空输出的概率进行抑制调整，包括：

通过以下调整方式中的至少一种，对该各个语音帧对应的该音素识别结果进行调整：

降低该各个语音帧对应的该音素识别结果中的空输出的概率；

以及，提高该各个语音帧对应的该音素识别结果中的各个音素的概率。

在一些实施例中，该降低该各个语音帧对应的该音素识别结果中的空输出的概率，包括：

将该各个语音帧对应的该音素识别结果中的空输出的概率乘以第一权重，该第一权重小于 1 且大于 0。

在本申请实施例中，对音素识别结果中的空输出的概率进行抑制，可以只降低音素识别结果中的空输出的概率，例如，在音素识别结果中的空输出的概率上乘以一个介于 0 到 1 之间的数。这样在音素识别结果中各个音素的概率不变的情况下，能够降低空输出的概率与各个音素的概率之间的比值。

在一些实施例中，该降低该各个语音帧对应的该音素识别结果中的空输出的概率，包括：

将该各个语音帧对应的该音素识别结果中的各个音素的概率乘以第二权重，该第二权重大于1。

在本申请实施例中，对音素识别结果中的空输出的概率进行抑制，可以只提升低音素识别结果中的空输出的概率，例如，在音素识别结果中的各个音素的概率上乘以一个大于1的数。这样在音素识别结果中的空输出的概率不变的情况下，能够降低空输出的概率与各个音素的概率之间的比值。

在另一示例性的方案中，计算机设备也可以在降低音素识别结果中的空输出的概率的同时，提高该音素识别结果中的各个音素的概率。例如，在音素识别结果中的空输出的概率上乘以一个介于0到1之间的数，同时，在音素识别结果中的各个音素的概率上乘以一个大于1的数。

在本方案中，上述声学模型为了得到输入和输出之前的对齐，在输入的音素序列中，需要插入空输出符号，即 \emptyset 。 \emptyset 符号和其他音素一样，使用模型进行预测。假设非空音素总数为P，则最终模型的输出维度为P+1，通常第0维表示空输出 \emptyset 。实验发现，空输出的引入，使模型的删除错误大幅提高，这说明大量的音素被错误的识别为空输出，为了解决空输出概率过高的问题，本申请在Transducer解码过程中，通过调节空输出的概率权重，来减少删除错误的产生。

以将该各个语音帧对应的该音素识别结果中的空输出的概率乘以第一权重为例，假设空输出的概率为 $Pr(\emptyset)$ ，为了降低空输出的概率，本方案在原始的空输出概率值上，除以一个大于1的权重 α ， $\alpha > 1$ ， α 称为折扣因子（discount factor），调整之后的空输出概率值为：

$$Pr_{dis}(\emptyset) = Pr(\emptyset)/\alpha \quad (3)$$

一般来说，采用对数概率作为最终的数值参与最终的解码分数计算，因此，对公式(3)两边取对数之后，可以得到：

$$\log Pr_{dis}(\emptyset) = \log Pr(\emptyset) - \log \alpha \quad (4)$$

上述公式(4)的结果可以作为空输出的调整后概率，以进行后续的解码。

在一些实施例中，上述第一权重或第二权重，是由开发人员或者管理人员预先设置在计算机设备中的，例如，上述第一权重或第二权重可以由开发人员预先设置在语音识别模型中。

步骤304，将各个语音帧对应的该音素识别结果中，空输出的概率满足指定

条件的该音素识别结果输入解码图，获得该语音信号对应的识别文本序列。

在一些实施例中，该将调整后的该各个语音帧对应的该音素识别结果输入解码图，获得该语音信号对应的识别文本序列，包括：

当目标音素识别结果中的空输出的概率满足指定条件时，将该目标音素识别结果输入该解码图，获得该目标音素识别结果对应的识别文本；

其中，该目标音素识别结果是该各个语音帧对应的该音素识别结果中的任意一个。

在一些实施例中，该指定条件包括：

该目标音素识别结果中的空输出的概率小于概率阈值。

实验中发现，相比于 DNN-HMM 模型，Transducer 模型的输出具有明显的尖峰效应，即在某一时刻，模型会以极高的置信度输出某一个预测结果。利用模型的尖峰效应，我们可以在解码过程中，跳过模型预测为空输出的概率，即这些概率将不会参与到解码图的解码过程中，由于本专利以音素为建模单元，同时，解码时跳过空输出，则解码图搜索的步数只与音素的个数有关，本方案中称之为“音素同步解码（Phone Synchronous Decoding, PSD）。下图给出了本方案所提出的 PSD 算法和空输出权重调整的整个流程：

算法 1: PSD 算法；

输入：特征序列 (x_1, x_2, \dots, x_T) ；空输出权重调节系数 β_{blank} ；

空输出阈值 γ_{blank} ；历史输出大小 M；解码图 LG；

输出：预测的词序列 w^* ；

1 $y_{in} = \text{Zeros}(M), u = 0, Q_{posterior} = \{\}, w^* = \{\}$

2 预测器前向计算 $h_0^{pred} = \text{Predictor}(y_{in})$

3 For 所有时间 t:

4 编码器前向计算 $h_t^{enc} = \text{Encoder}(x_t)$

5 联合网络前向计算 $p_{t,u} = \text{Joint}(h_t^{enc}, h_{u-1}^{pred})$

6 空输出概率调整 $p_{t,u}(blank) = p_{t,u}(blank) * \beta_{blank}$

7 获取当前预测 $y_{t,u} = \text{arg max}(p_{t,u})$

8 IF 当前预测不为空 $y_{t,u} \neq blank$; Then

9 输出计数器加 1 $u = u + 1$

10 更新预测器输入 $y_{in} = [y_{in}[1:], y_{t,u}]$

11 预测器前向计算 $h_u^{pred} = \text{Predictor}(y_{in})$

```

12      EndIF
13      IF 空输出概率小于阈值  $p_{t,u}(blank) < \gamma_{blank}$ 
14          声学模型后验概率入队  $Enqueue(Q_{posterior}, p_{t,u})$ 
15          LG 解 码 器 进 行 解 码
16           $w_{t,u} = WFSTDecoding(LG, Q_{posterior})$ 
           解码结果入队  $Enqueue(w^*, w_{t,u})$ 
17      EndIF
18  EndFor
19  Return  $w^*$ 

```

其中，上述算法中的第 6 行进行了权重的调整，算法中的 β_{blank} 即为公式(3)中的 $1/\alpha$ ，上述算法中的第 13-17 行，为本方案所提出的 PSD 算法，即只有当空输出的概率小于一定的阈值 γ_{blank} 时，网络输出的概率分布才会参与到后续的解码图的解码中。

在一些实施例中，上述概率阈值是由开发人员或者管理人员预先设置在计算机设备中的，例如，上述概率阈值可以由开发人员预先设置在语音识别模型中。

在一些实施例中，该将调整后的该各个语音帧对应的该音素识别结果输入解码图，获得该语音信号对应的识别文本序列之前，还包括：

获取阈值影响参数，该阈值影响参数包括环境音强度、指定时间段内语音识别失败的次数、以及用户设置信息中的至少一种；

基于该阈值影响参数，确定该概率阈值。

在本申请实施例中，上述概率阈值也可以由计算机设备在进行语音识别过程中进行调整。也就是说，计算机设备可以获取可能影响概率阈值的取值的相关参数，并通过相关参数来灵活设置概率阈值。

例如，环境音强度可以会对用户发出的语音造成干扰，因此，当环境音强度较强时，计算机设备可以将概率阈值设置的较高，使得更多的音素识别结果被输入解码图进行解码，从而保证识别的准确性；反之，当环境音强度较弱时，计算机设备可以将概率阈值设置的较低，使得更多的音素识别结果被跳过，从而保证识别的效率。

再例如，基于解码图对音素识别结果进行解码的准确性会影响语音识别的成功率，当指定时间段（比如当前时刻之前的一段时间，例如 5 分钟）内语音

识别失败的次数过多时，计算机设备可以将概率阈值设置的较高，使得更多的音素识别结果被输入解码图进行解码，从而保证识别的准确性；反之，当指定时间段内语音识别失败的次数较少或者未失败时，计算机设备可以将概率阈值设置的较低，使得更多的音素识别结果被跳过，从而保证识别的效率。

在一些实施例中，该解码图由音素词典和语言模型复合构成。

本方案采用的解码图是由音素词典和语言模型两个子加权有限自动机（Weighted Finite State Transducer, WFST）图复合而成。

音素词典 WFST：汉字或词到音素序列的映射。输入音素序列串，WFST 可以输出对应的汉字或者词；通常，此 WFST 与文本领域无关，在不同的识别任务中为通用部分；

语言模型 WFST：此 WFST 通常由 n-gram 语言模型转换而来，语言模型用来计算一个句子出现的概率，利用训练数据和统计学方法训练而来。通常，不同领域的文本，例如新闻和口语对话的文本，常用词和词间搭配存在较大的差异，因此，当进行不同领域的语音识别时，可以通过改变语言模型 WFST，来实现适配。

参考图 7，其示出了本申请实施例涉及的模型训练及应用流程图。如图 7 所示，以应用于车载设备为例，本申请实施例所示的模型训练完成之后，采用 libtorch 进行模型的量化和部署。libtorch 的安卓版本，采用 QNNPACK 库进行 INT8 的矩阵计算，大大加速的矩阵运算速度。模型在 Python 环境 71 中，采用 pytorch 进行训练，然后对模型进行训练后量化，即把模型参数量化为 INT8，并采用 INT8 的矩阵乘法来加速计算，将量化后的模型导出后，用于 C++环境 72 的前向推断，以通过测试数据进行测试。

通过本申请所示的方案，一方面，基于 Transducer 的端到端模型在训练过程中，无需帧级别的对齐信息，极大的简化了建模过程；其次，简化解码图，减小搜索空间。而本方案中提出的方法，由于采用音素建模，解码图只需要 L 和 G 复合，搜索空间大为降低。最后，采用音素建模，结合自定义的解码图，可以实现灵活的定制需求，可以根据不同的业务场景，在不改变声学模型的前提下，只需要定制语言模型，就可以适配各自的业务场景。

相比于相关技术中的离线识别系统，本方案在识别率和 CPU 占用率上都具有优势：

在识别率方面，相比于 DNN 结合隐马尔可夫模型（Hidden Markov Model,

HMM) 系统模型 (DNN-HMM 模型), 本方案所示的系统模型取得了大幅的提高;

在 CPU 占用方面, 本方案所示的系统模型在模型参数量是 DNN-HMM 系统的 4 倍的情况下, 仍与 DNN-HMM 系统模型具有相似的 cpu 占用率。

语音识别率对比如下:

下表 1 示出了在 3 个数据集合上, 现有的 DNN-HMM 系统和本方案所提出的 Transducer 系统的字错误率 (Character Error Rate, CER) 对比。

表 1

模型	参数量	测试集 1 CER(%)	测试集 2 CER(%)
DNN-HMM	0.7M	14.88	19.77
Transducer1	0.8M	12.1	16.09
Tansducer2	1.9M	9.76	13.4
Tansducer3	2.1M	8.93	13.18

从表 1 中可以发现, 相似参数量下, 在两个测试集合上, Transducer1 模型分别取得了相对 18.7%和 18.6%的 CER 的下降。同时, 当增加模型参数量之后, 利用 Transducer3, 分别取得了 8.93%和 13.18%的词错误率。

CPU 占用率对比:

表 2

模型	参数量	CPU 占用 (峰值)
DNN-HMM	0.7M	16%
Transducer1	0.8M	18%
Tansducer2	1.9M	20%
Tansducer3	2.1M	20%

通过表 2 对比 Transducer1 和 DNN-HMM, 两个模型在同等参数量时, Transducer1 模型比 DNN-HMM 模型的峰值高了 2%, 但是当模型参数量增加, Transducer 模型的峰值并没有明显变化, 在大幅增加模型参数量和降低识别错误率的条件下, CPU 占用率仍然保持在较低的水平。

综上所述, 本申请实施例所示的方案, 对于包含语音帧在各个音素以及空输出上的概率分布的音素识别结果, 在将该音素识别结果输入解码图之前, 先对音素识别结果中的空输出的概率进行抑制, 降低语音帧被识别为空输出的几率, 从而降低语音帧被误识别为空输出的可能性, 也就是降低模型的删除错误,

从而提高模型的识别准确性。

本申请上述图 3 所示实施例中的方案，以同时应用空输出权重调节（步骤 303）和解码跳帧（对应步骤 304）为例进行说明，在其它实现方案中，空输出权重调节和解码跳帧也可以独立应用。例如，在本申请一个示例性的实施例中，当独立应用上述解码跳帧时，本申请所示的方案可以如下：

获取语音信号，该语音信号包括对原始语音进行切分获得的各个语音帧；

对语音信号进行音素识别，获得该各个语音帧对应的音素识别结果；该音素识别结果用于指示对应的语音帧在音素空间中的概率分布；该音素空间中包含各个音素以及一个空输出；

将该各个语音帧对应的该音素识别结果中，空输出的概率满足指定条件的该音素识别结果输入解码图，获得该语音信号对应的识别文本序列。

综上所述，本申请实施例所示的方案，对于包含语音帧在各个音素以及空输出上的概率分布的音素识别结果，在将该音素识别结果输入解码图时，可以对其中空输出的概率满足条件的音素识别结果进行解码，减少了需要解码的因素识别结果的数量，跳过不必要的解码步骤，从而有效的提高了语音识别效率。

参考图 8，其是根据一示例性实施例示出的一种语音识别系统的框架图。如图 8 所示，音频采集设备 81 与语音识别设备 82 相连，语音识别设备 82 中包含声学模型 82a、概率调整单元 82b、解码图输入单元 82c、解码图 82d、以及特征提取单元 82e。其中，解码图 82d 由音素词典和语言模型组成。

在应用过程中，音频采集设备 81 采集用户的原始语音后，将原始语音传输给语音识别设备 82 中的特征提取单元 82e，由特征提取单元进行切分并对各个语音帧进行特征提取后，将一个语音帧的语音特征，以及解码图 82d 对该语音帧的前 4 个非空语音帧识别出的文本的音素，分别输入到声学模型 82a 中的 FSMN 和一维卷积网络中，得到声学模型 82a 输出该语音帧的音素识别结果。

该音素识别结果输入至概率调整单元 82b，进行空输出的概率调整，得到调整后的音素识别结果；该调整后的语音识别结果由解码图输入单元 82c 进行判断，当判断出调整后的空输出概率小于阈值时，确定需要进行解码，解码图输入单元 82c 将该调整后的音素识别结果输入解码图 82d，由解码图 82d 识别出文本；反之，如果判断出调整后的空输出概率不小于阈值时，确定不需要进行解

码，则丢弃该调整后的语音识别结果。

上述解码图对各个语音帧的调整后的音素识别结果进行识别，并输出文本序列后，可以将文本序列输出给自然语言处理组件，由自然语言处理组件对用户输入的语音做出响应。

图 9 是根据一示例性实施例示出的一种语音识别装置的结构方框图。该语音识别装置可以实现图 2 或图 3 所示实施例提供的方法中的全部或者部分步骤。该语音识别装置可以包括：

语音信号处理模块 901，用于对语音信号进行音素识别，获得语音信号中各个语音帧对应的音素识别结果；音素识别结果用于指示对应的语音帧在音素空间中的概率分布；音素空间中包含各个音素以及空输出；

概率调整模块 902，用于对各个语音帧对应的音素识别结果中的空输出的概率进行抑制调整，以降低音素识别结果中的空输出的概率与各个音素的概率的比值；及

解码模块 903，用于将调整后的各个语音帧对应的音素识别结果输入解码图，获得语音信号对应的识别文本序列，解码图中包括字符与音素之间的映射关系。

在一些实施例中，概率调整模块 902，用于通过以下调整方式中的至少一种，对各个语音帧对应的音素识别结果进行调整：

降低各个语音帧对应的音素识别结果中的空输出的概率；

以及，

提高各个语音帧对应的音素识别结果中的各个音素的概率。

在一些实施例中，概率调整模块 902，用于将各个语音帧对应的音素识别结果中的空输出的概率乘以第一权重，第一权重小于 1 且大于 0。

在一些实施例中，概率调整模块 902，用于将各个语音帧对应的音素识别结果中的各个音素的概率乘以第二权重，第二权重大于 1。

在一些实施例中，解码模块 903，用于，

当目标音素识别结果中的空输出的概率满足指定条件时，将目标音素识别结果输入解码图，获得目标音素识别结果对应的识别文本；

其中，目标音素识别结果是各个语音帧对应的音素识别结果中的任意一个。

在一些实施例中，指定条件包括：

目标音素识别结果中的空输出的概率小于概率阈值。

在一些实施例中，装置还包括：

参数获取模块，用于获取阈值影响参数，阈值影响参数包括环境音强度、指定时间段内语音识别失败的次数、以及用户设置信息中的至少一种；

阈值确定模块，用于基于阈值影响参数，确定概率阈值。

在一些实施例中，语音信号处理模块 901，用于，

通过已训练的声学模型对目标语音帧进行特征提取，获得目标语音帧的特征向量；目标语音帧是各个语音帧中的任意一个；

将目标语音帧输入声学模型中的编码器，获得目标语音帧的声学隐层表示向量；及

将目标语音帧的历史识别文本的音素信息输入声学模型中的预测器，获得目标语音帧的文本隐层表示向量；目标语音帧的历史识别文本，是解码图对目标语音帧的前 n 个非空输出的语音帧的音素识别结果进行识别得到的文本； n 为大于或者等于 1 的整数；

将目标语音帧的声学隐层表示向量，以及目标语音帧的文本隐层表示向量输入联合网络，获得目标语音帧的音素识别结果。

在一些实施例中，编码器为前向序列记忆网络 FSMN。

在一些实施例中，预测器为一维卷积网络。

在一些实施例中，解码图由音素词典和语言模型复合构成。

综上所述，本申请实施例所示的方案，对于包含语音帧在各个音素以及空输出上的概率分布的音素识别结果，在将该音素识别结果输入解码图之前，先对音素识别结果中的空输出的概率进行抑制，降低语音帧被识别为空输出的几率，从而降低语音帧被误识别为空输出的可能性，也就是降低模型的删除错误，从而提高模型的识别准确性。

图 10 是根据一示例性实施例示出的一种计算机设备的结构示意图。该计算机设备可以实现为上述各个方法实施例中的计算机设备。所述计算机设备 1000 包括中央处理单元 1001、包括随机存取存储器 (Random Access Memory, RAM) 1002 和只读存储器 (Read-Only Memory, ROM) 1003 的系统存储器 1004，以及连接系统存储器 1004 和中央处理单元 1001 的系统总线 1005。所述计算机设备 1000 还包括帮助计算机内的各个器件之间传输信息的基本输入/输出系统

1006, 和用于存储操作系统 1013、应用程序 1014 和其他程序模块 1015 的大容量存储设备 1007。

所述大容量存储设备 1007 通过连接到系统总线 1005 的大容量存储控制器 (未示出) 连接到中央处理单元 1001。所述大容量存储设备 1007 及其相关联的计算机可读介质为计算机设备 1000 提供非易失性存储。也就是说, 所述大容量存储设备 1007 可以包括诸如硬盘或者光盘只读存储器 (Compact Disc Read-Only Memory, CD-ROM) 驱动器之类的计算机可读介质 (未示出)。

不失一般性, 所述计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其他数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括 RAM、ROM、闪存或其他固态存储其技术, CD-ROM、或其他光学存储、磁带盒、磁带、磁盘存储或其他磁性存储设备。当然, 本领域技术人员可知所述计算机存储介质不局限于上述几种。上述的系统存储器 1004 和大容量存储设备 1007 可以统称为存储器。

计算机设备 1000 可以通过连接在所述系统总线 1005 上的网络接口单元 1011 连接到互联网或者其它网络设备。

所述存储器还包括至少一条计算机指令, 所述至少一条计算机指令存储于存储器中, 处理器通过加载并执行该至少一条计算机指令来实现图 2 或图 3 所示的方法的全部或者部分步骤。

在示例性实施例中, 还提供了一种包括指令的非临时性计算机可读存储介质, 例如包括计算机程序 (指令) 的存储器, 上述程序 (指令) 可由计算机设备的处理器执行以完成本申请各个实施例所示的方法。例如, 所述非临时性计算机可读存储介质可以是只读存储器 (Read-Only Memory, ROM)、随机存取存储器 (Random Access Memory, RAM)、只读光盘 (Compact Disc Read-Only Memory, CD-ROM)、磁带、软盘和光数据存储设备等。

在示例性实施例中, 还提供了一种计算机程序产品或计算机程序, 该计算机程序产品或计算机程序包括计算机指令, 该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令, 处理器执行该计算机指令, 使得该计算机设备执行上述各个实施例所示的方法。

一种计算机程序产品，包括计算机程序，其特征在于，该计算机程序被处理器执行时实现上述各个实施例所示的方法。

本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本申请的真正范围和精神由权利要求指出。

应当理解的是，本申请并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

权利要求书

1、一种语音识别方法，其特征在于，由计算机设备执行，所述方法包括：

对语音信号进行音素识别，获得所述语音信号中各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；

对所述各个语音帧对应的所述音素识别结果中的空输出的概率进行抑制调整，以降低所述音素识别结果中的空输出的概率与各个音素的概率的比值；及

将调整后的所述各个语音帧对应的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

2、根据权利要求1所述的方法，其特征在于，所述对所述各个语音帧对应的所述音素识别结果中的空输出的概率进行抑制调整，包括：

降低所述各个语音帧对应的所述音素识别结果中的空输出的概率。

3、根据权利要求2所述的方法，其特征在于，所述降低所述各个语音帧对应的所述音素识别结果中的空输出的概率，包括：

将所述各个语音帧对应的所述音素识别结果中的空输出的概率乘以第一权重，所述第一权重小于1且大于0。

4、根据权利要求2所述的方法，其特征在于，所述降低所述各个语音帧对应的所述音素识别结果中的空输出的概率，包括：

将所述各个语音帧对应的所述音素识别结果中的各个音素的概率乘以第二权重，所述第二权重大于1。

5、根据权利要求1所述的方法，其特征在于，所述对所述各个语音帧对应的所述音素识别结果中的空输出的概率进行抑制调整，包括：

提高所述各个语音帧对应的所述音素识别结果中的各个音素的概率。

6、根据权利要求1所述的方法，其特征在于，所述将调整后的所述各个语音帧对应的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，包括：

当目标音素识别结果中的空输出的概率满足指定条件时，将所述目标音素识别结果输入所述解码图，获得所述目标音素识别结果对应的识别文本；

其中，所述目标音素识别结果是所述各个语音帧对应的所述音素识别结果中的任意一个。

7、根据权利要求6所述的方法，其特征在于，所述指定条件包括：

所述目标音素识别结果中的空输出的概率小于概率阈值。

8、根据权利要求7所述的方法，其特征在于，所述将调整后的所述各个语音帧对应的所

述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列之前，还包括：

获取阈值影响参数，所述阈值影响参数包括环境音强度、指定时间段内语音识别失败的次数、以及用户设置信息中的至少一种；及

基于所述阈值影响参数，确定所述概率阈值。

9、根据权利要求1所述的方法，其特征在于，所述对语音信号进行音素识别，获得所述语音信号中各个语音帧对应的音素识别结果，包括：

通过已训练的声学模型对目标语音帧进行特征提取，获得所述目标语音帧的特征向量；所述目标语音帧是所述各个语音帧中的任意一个；

将所述目标语音帧输入所述声学模型中的编码器，获得所述目标语音帧的声学隐层表示向量；

将所述目标语音帧的历史识别文本的音素信息输入所述声学模型中的预测器，获得所述目标语音帧的文本隐层表示向量；所述目标语音帧的历史识别文本，是所述解码图对所述目标语音帧的前n个非空输出的语音帧的音素识别结果进行识别得到的文本；n为大于或者等于1的整数；及

将所述目标语音帧的声学隐层表示向量，以及所述目标语音帧的文本隐层表示向量输入所述声学模型中的联合网络，获得所述目标语音帧的所述音素识别结果。

10、根据权利要求9所述的方法，其特征在于，所述编码器为前向序列记忆网络FSMN。

11、根据权利要求9所述的方法，其特征在于，所述预测器为一维卷积网络。

12、根据权利要求1至9任一所述的方法，其特征在于，所述解码图由音素词典和语言模型复合构成。

13、一种语音识别方法，其特征在于，由计算机设备执行，所述方法包括：

获取语音信号，所述语音信号包括对原始语音进行切分获得的各个语音帧；

对语音信号进行音素识别，获得所述各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；及

将所述各个语音帧对应的所述音素识别结果中，空输出的概率满足指定条件的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

14、一种语音识别装置，其特征在于，所述装置包括：

语音信号处理模块，用于对语音信号进行音素识别，获得所述语音信号中各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；

概率调整模块，用于对所述各个语音帧对应的所述音素识别结果中的空输出的概率进行

抑制调整，以降低所述音素识别结果中的空输出的概率与各个音素的概率的比值；及

解码模块，用于将调整后的所述各个语音帧对应的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

15、根据权利要求 14 所述的装置，其特征在于，所述概率调整模块，还用于：
降低所述各个语音帧对应的所述音素识别结果中的空输出的概率。

16、根据权利要求 15 所述的装置，其特征在于，所述概率调整模块，还用于：
将所述各个语音帧对应的所述音素识别结果中的空输出的概率乘以第一权重，所述第一权重小于 1 且大于 0。

17、一种语音识别装置，其特征在于，所述装置包括：

语音信号获取模块，用于获取语音信号，所述语音信号包括对原始语音进行切分获得的各个语音帧；

音素识别结果获得模块，用于对语音信号进行音素识别，获得所述各个语音帧对应的音素识别结果；所述音素识别结果用于指示对应的语音帧在音素空间中的概率分布；所述音素空间中包含各个音素以及空输出；及

识别文本序列获取模块，用于将所述各个语音帧对应的所述音素识别结果中，空输出的概率满足指定条件的所述音素识别结果输入解码图，获得所述语音信号对应的识别文本序列，所述解码图中包括字符与音素之间的映射关系。

18、一种计算机设备，其特征在于，所述计算机设备包含处理器和存储器，所述存储器中存储有至少一条计算机指令，所述至少一条计算机指令由所述处理器加载并执行以实现如权利要求 1 至 13 任一所述的语音识别方法。

19、一种计算机可读存储介质，其特征在于，所述存储介质中存储有至少一条计算机指令，所述至少一条计算机指令由处理器加载并执行以实现如权利要求 1 至 13 任一所述的语音识别方法。

20、一种计算机程序产品，包括计算机程序，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 13 中任一项所述的方法的步骤。

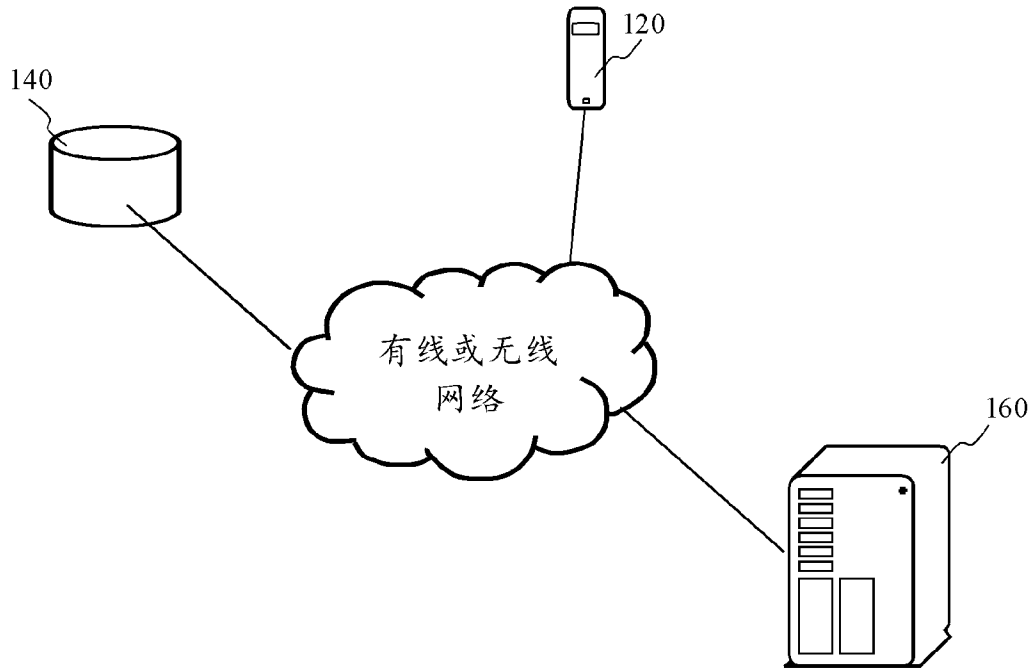


图 1

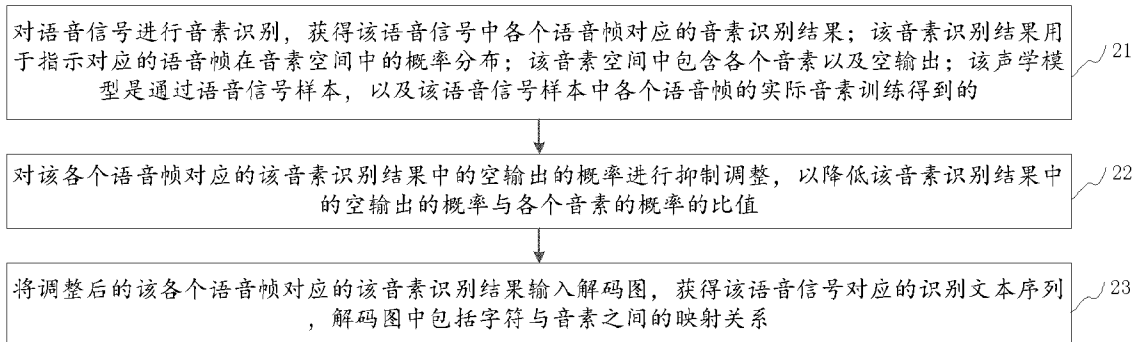


图 2

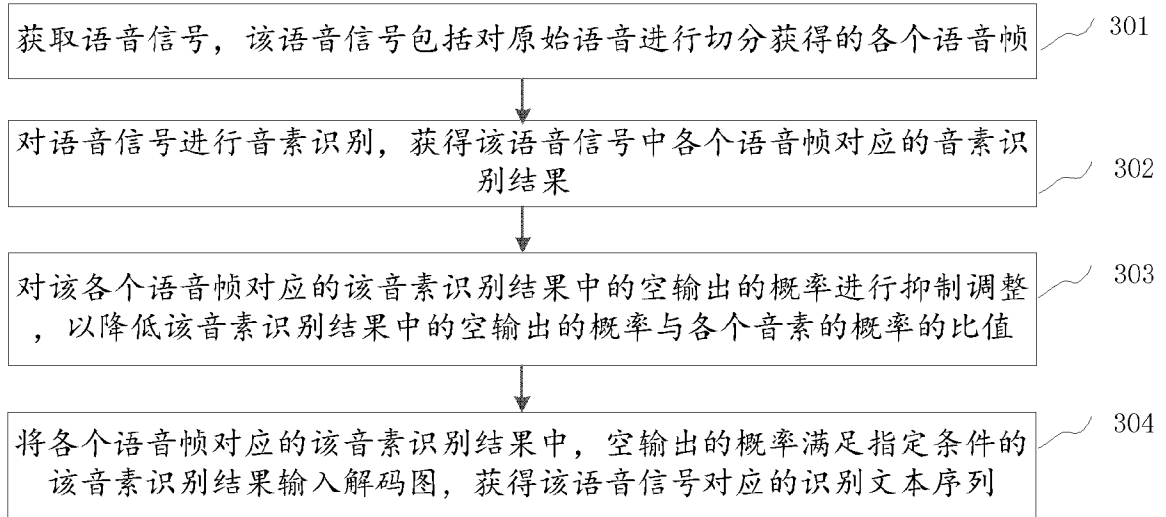


图 3

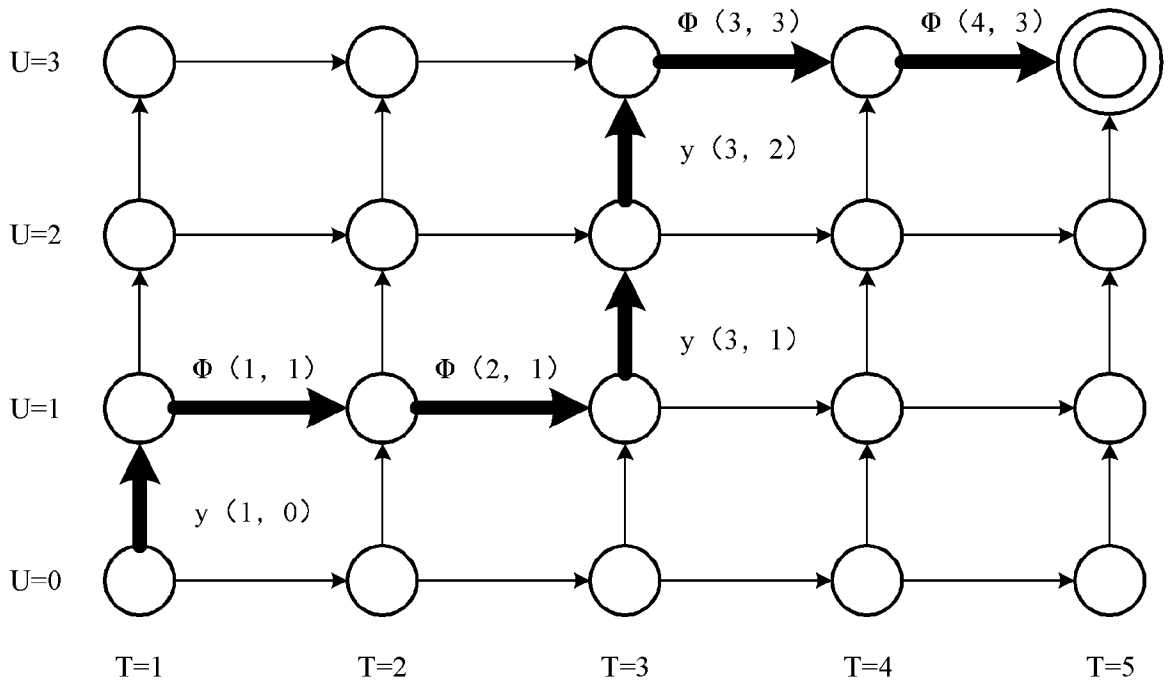


图 4

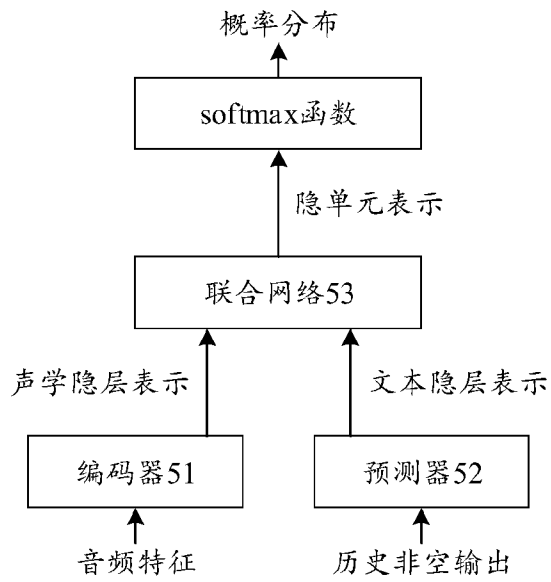


图 5

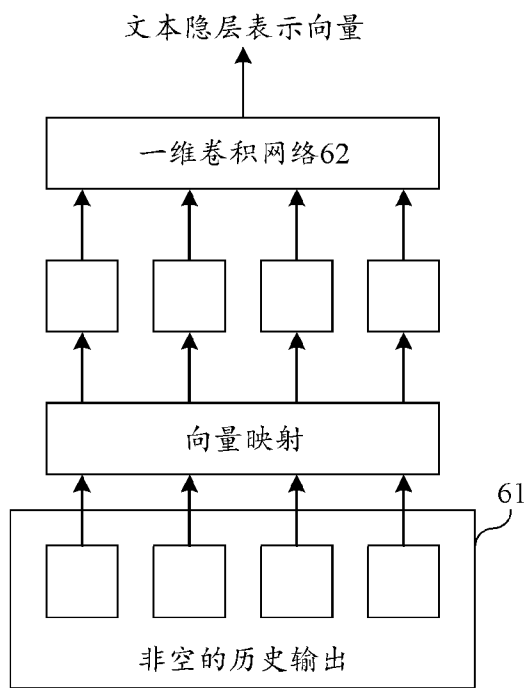


图 6

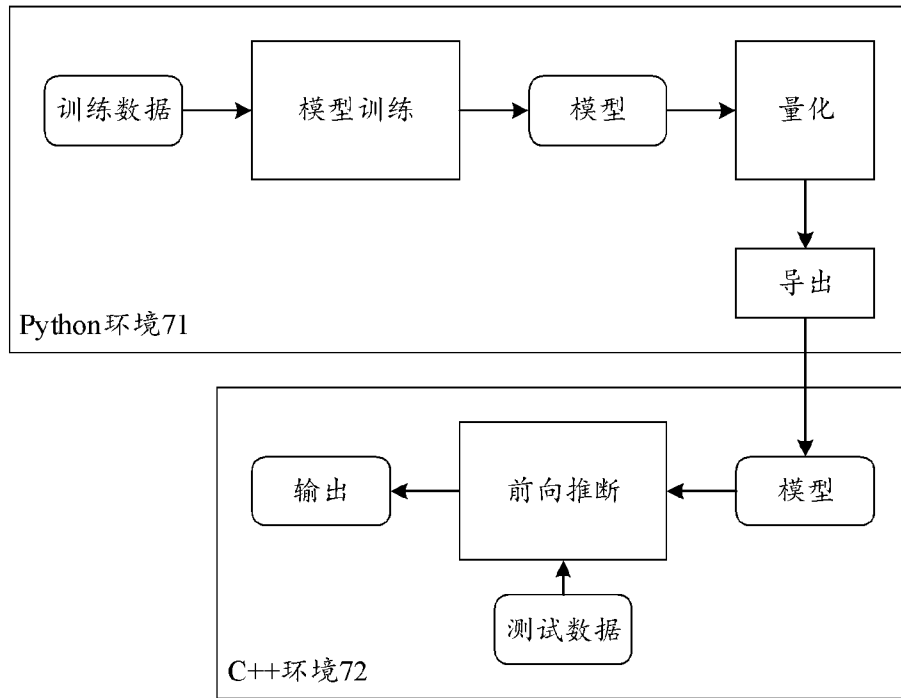


图 7

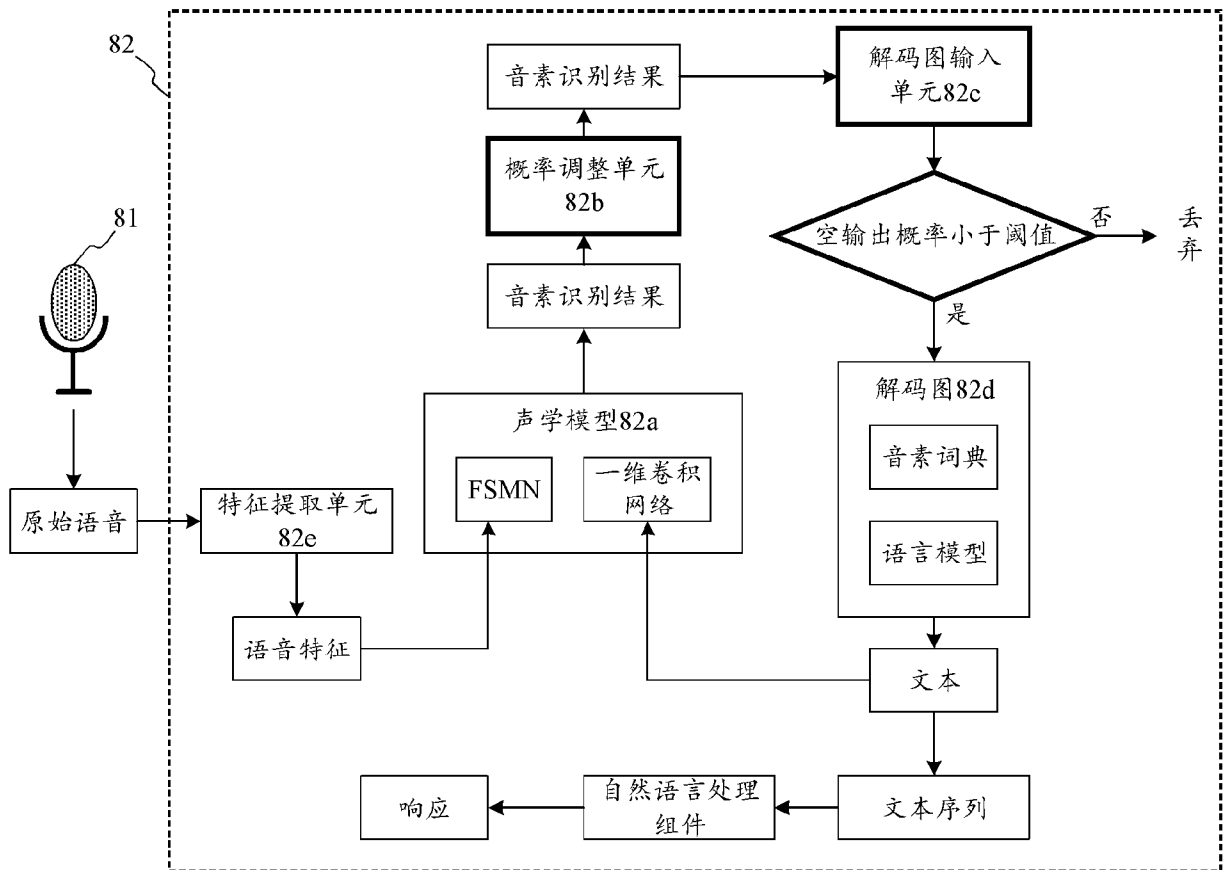


图 8

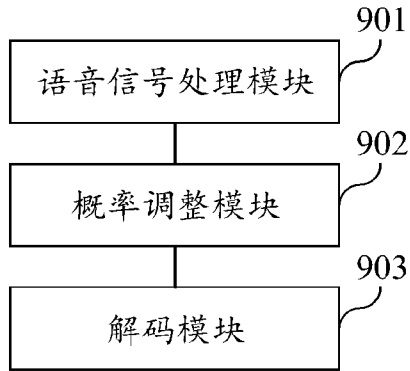


图 9

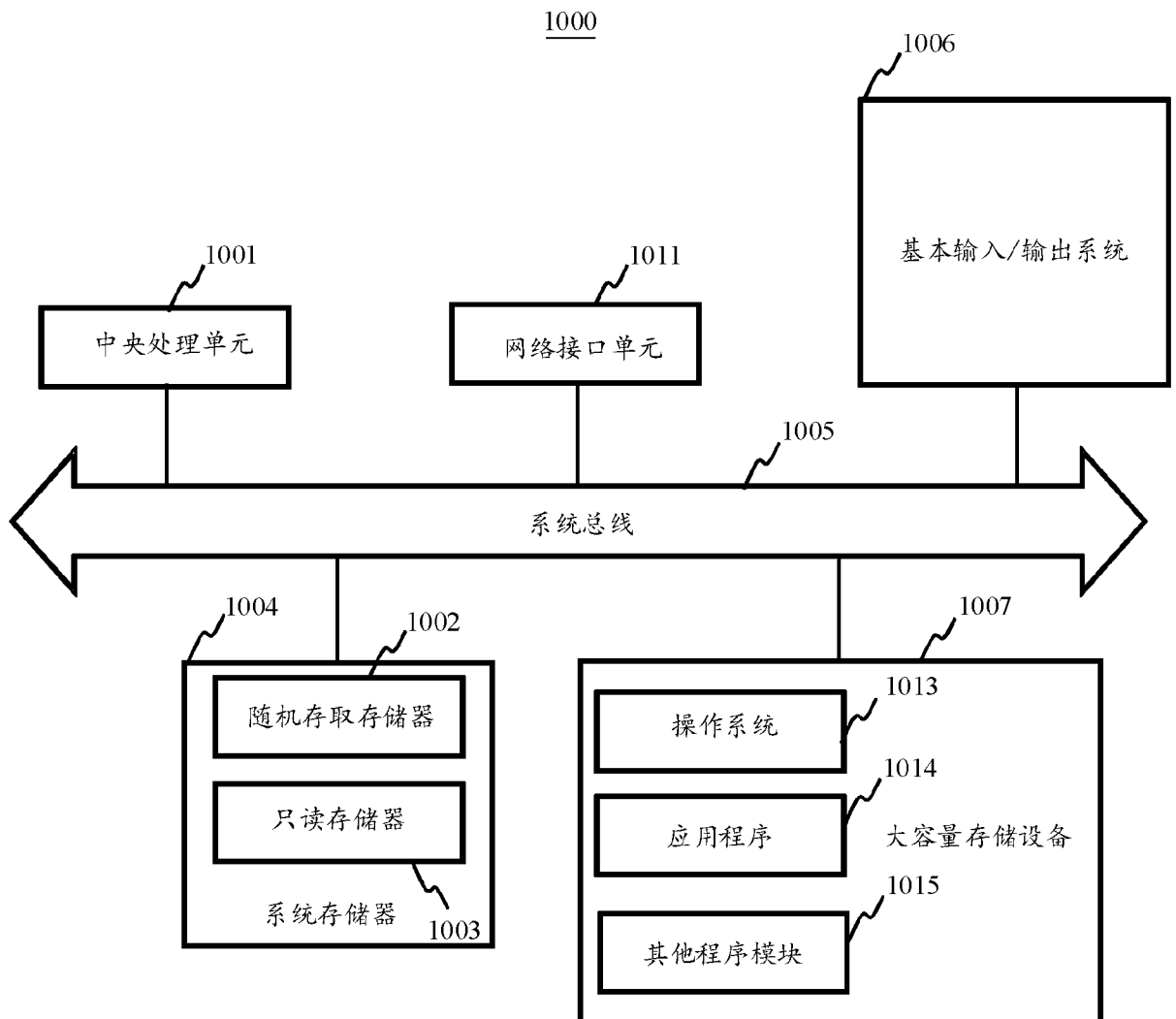


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/129223

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/16(2006.01)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)

G10L15;G10L19

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; WPABS; CNABS; CNTXT; CNKI; 百度学术: 语音识别, 语音, 解码, 音素同步解码, 音素, 同步, 空白, 空输出, 概率, 删除, 错误, 权重, 降低, 减少, 减小, 抑制, blank, phone, synchronous, deletion error, deweight, reduce

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113539242 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 22 October 2021 (2021-10-22) claims 1-15	1-20
X	CN 105895081 A (SUZHOU AISPEECH INFORMATION TECHNOLOGY CO., LTD. et al.) 24 August 2016 (2016-08-24) paragraphs [0033]-[0069]	13, 17-20
X	CN 108389575 A (SUZHOU AISPEECH INFORMATION TECHNOLOGY CO., LTD. et al.) 10 August 2018 (2018-08-10) description, paragraphs [0035]-[0050]	13, 17-20
A	CN 109559735 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 02 April 2019 (2019-04-02) description, paragraphs [0044]-[0069], and figure 1	1-12, 14-16
A	CN 110164421 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 23 August 2019 (2019-08-23) 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

26 January 2022

Date of mailing of the international search report

09 February 2022

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

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/129223**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105529027 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 27 April 2016 (2016-04-27) entire document	1-20
A	CN 108269568 A (INSTITUTE OF ACOUSTICS, CHINESE ACADEMY OF SCIENCES et al.) 10 July 2018 (2018-07-10) entire document	1-20
A	WO 2020195068 A1 (MITSUBISHI ELECTRIC CORP.) 01 October 2020 (2020-10-01) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/129223

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113539242	A	22 October 2021	None			
CN	105895081	A	24 August 2016	US	2019057685	A1	21 February 2019
				EP	3444806	A1	20 February 2019
				EP	3444806	A4	11 December 2019
				WO	2017177484	A1	19 October 2017
CN	108389575	A	10 August 2018	CN	108389575	B	26 June 2020
CN	109559735	A	02 April 2019	WO	2020073509	A1	16 April 2020
CN	110164421	A	23 August 2019	US	2021193123	A1	24 June 2021
				WO	2020119351	A1	18 June 2020
CN	105529027	A	27 April 2016	US	2018254039	A1	06 September 2018
				US	10650809	B2	12 May 2020
				WO	2017101450	A1	22 June 2017
				CN	105529027	B	31 May 2019
CN	108269568	A	10 July 2018	CN	108269568	B	30 July 2021
WO	2020195068	A1	01 October 2020	CN	113574595	A	29 October 2021
				US	2020312306	A1	01 October 2020
				US	11100920	B2	24 August 2021

国际检索报告

国际申请号

PCT/CN2021/129223

<p>A. 主题的分类 G10L 15/16(2006.01)i</p> <p>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</p>																												
<p>B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G10L15;G10L19</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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;WPABS;CNABS;CNTXT;CNKI;百度学术;语音识别, 语音, 解码, 音素同步解码, 音素, 同步, 空白, 空输出, 概率, 删除, 错误, 权重, 降低, 减少, 减小, 抑制, blank, phone ,synchronous ,deletion error, deweight, reduce</p>																												
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 113539242 A (腾讯科技深圳有限公司) 2021年10月22日 (2021 - 10 - 22) 权利要求1-15</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 105895081 A (苏州思必驰信息科技有限公司 等) 2016年8月24日 (2016 - 08 - 24) 第[0033]-[0069]段</td> <td>13, 17-20</td> </tr> <tr> <td>X</td> <td>CN 108389575 A (苏州思必驰信息科技有限公司 等) 2018年8月10日 (2018 - 08 - 10) 说明书第[0035]-[0050]段</td> <td>13, 17-20</td> </tr> <tr> <td>A</td> <td>CN 109559735 A (平安科技深圳有限公司) 2019年4月2日 (2019 - 04 - 02) 说明书第[0044]-[0069]段, 附图1</td> <td>1-12, 14-16</td> </tr> <tr> <td>A</td> <td>CN 110164421 A (腾讯科技深圳有限公司) 2019年8月23日 (2019 - 08 - 23) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 105529027 A (百度在线网络技术北京有限公司) 2016年4月27日 (2016 - 04 - 27) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108269568 A (中国科学院声学研究所 等) 2018年7月10日 (2018 - 07 - 10) 全文</td> <td>1-20</td> </tr> </tbody> </table> <p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <table border="1"> <tr> <td>* 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</td> <td>“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</td> </tr> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 113539242 A (腾讯科技深圳有限公司) 2021年10月22日 (2021 - 10 - 22) 权利要求1-15	1-20	X	CN 105895081 A (苏州思必驰信息科技有限公司 等) 2016年8月24日 (2016 - 08 - 24) 第[0033]-[0069]段	13, 17-20	X	CN 108389575 A (苏州思必驰信息科技有限公司 等) 2018年8月10日 (2018 - 08 - 10) 说明书第[0035]-[0050]段	13, 17-20	A	CN 109559735 A (平安科技深圳有限公司) 2019年4月2日 (2019 - 04 - 02) 说明书第[0044]-[0069]段, 附图1	1-12, 14-16	A	CN 110164421 A (腾讯科技深圳有限公司) 2019年8月23日 (2019 - 08 - 23) 全文	1-20	A	CN 105529027 A (百度在线网络技术北京有限公司) 2016年4月27日 (2016 - 04 - 27) 全文	1-20	A	CN 108269568 A (中国科学院声学研究所 等) 2018年7月10日 (2018 - 07 - 10) 全文	1-20	* 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
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																										
PX	CN 113539242 A (腾讯科技深圳有限公司) 2021年10月22日 (2021 - 10 - 22) 权利要求1-15	1-20																										
X	CN 105895081 A (苏州思必驰信息科技有限公司 等) 2016年8月24日 (2016 - 08 - 24) 第[0033]-[0069]段	13, 17-20																										
X	CN 108389575 A (苏州思必驰信息科技有限公司 等) 2018年8月10日 (2018 - 08 - 10) 说明书第[0035]-[0050]段	13, 17-20																										
A	CN 109559735 A (平安科技深圳有限公司) 2019年4月2日 (2019 - 04 - 02) 说明书第[0044]-[0069]段, 附图1	1-12, 14-16																										
A	CN 110164421 A (腾讯科技深圳有限公司) 2019年8月23日 (2019 - 08 - 23) 全文	1-20																										
A	CN 105529027 A (百度在线网络技术北京有限公司) 2016年4月27日 (2016 - 04 - 27) 全文	1-20																										
A	CN 108269568 A (中国科学院声学研究所 等) 2018年7月10日 (2018 - 07 - 10) 全文	1-20																										
* 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 2022年1月26日	国际检索报告邮寄日期 2022年2月9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	授权官员 游晓梅 电话号码 (86-10)62089539																											

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
A	WO 2020195068 A1 (MITSUBISHI ELECTRIC CORP.) 2020年10月1日 (2020 - 10 - 01) 全文	1-20

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/129223

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	113539242	A	2021年10月22日	无			
CN	105895081	A	2016年8月24日	US	2019057685	A1	2019年2月21日
				EP	3444806	A1	2019年2月20日
				EP	3444806	A4	2019年12月11日
				WO	2017177484	A1	2017年10月19日
CN	108389575	A	2018年8月10日	CN	108389575	B	2020年6月26日
CN	109559735	A	2019年4月2日	WO	2020073509	A1	2020年4月16日
CN	110164421	A	2019年8月23日	US	2021193123	A1	2021年6月24日
				WO	2020119351	A1	2020年6月18日
CN	105529027	A	2016年4月27日	US	2018254039	A1	2018年9月6日
				US	10650809	B2	2020年5月12日
				WO	2017101450	A1	2017年6月22日
				CN	105529027	B	2019年5月31日
CN	108269568	A	2018年7月10日	CN	108269568	B	2021年7月30日
WO	2020195068	A1	2020年10月1日	CN	113574595	A	2021年10月29日
				US	2020312306	A1	2020年10月1日
				US	11100920	B2	2021年8月24日