

(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
2025 年 1 月 23 日 (23.01.2025)



(10) 国际公布号
WO 2025/016325 A1

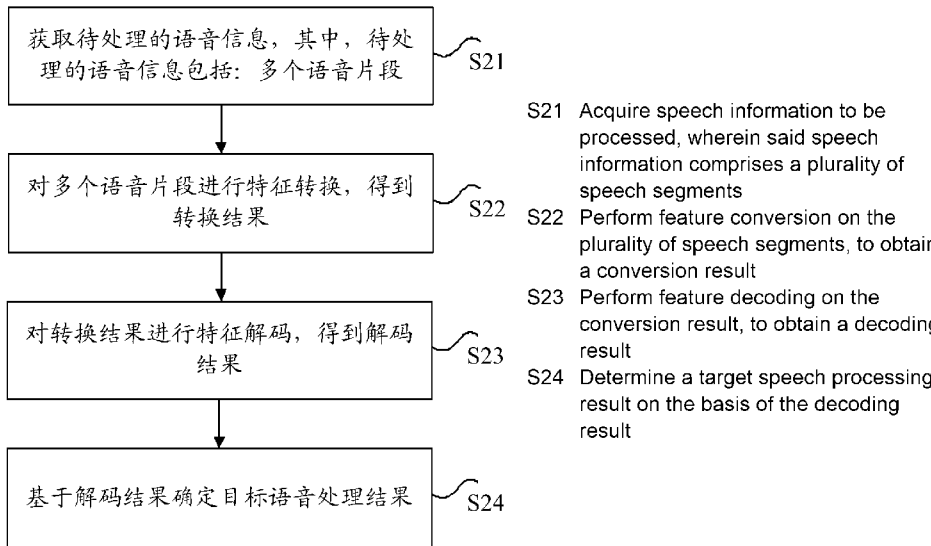
- (51) 国际专利分类号:
G10L 15/02 (2006.01) *G10L 25/30* (2013.01)
G10L 15/16 (2006.01) *G10L 25/03* (2013.01)
- (21) 国际申请号: PCT/CN2024/105264
- (22) 国际申请日: 2024 年 7 月 12 日 (12.07.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202310868220.5 2023年7月14日 (14.07.2023) CN
- (71) 申请人: 阿里巴巴(中国)有限公司(**ALIBABA (CHINA) CO., LTD**) [CN/CN]; 中国浙江省杭州市滨江区长河街道网商路 699 号 4 号楼 5 楼 508 室, Zhejiang 310052 (CN)。
- (72) 发明人: 陈谦(**CHEN, Qian**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号, Zhejiang 310024 (CN)。郑斯奇(**ZHENG, Siqi**); 中国浙江省杭州市西湖区

三墩镇灯彩街 1008 号, Zhejiang 310024 (CN)。张仕良(**ZHANG, Shiliang**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号, Zhejiang 310024 (CN)。张庆林(**ZHANG, Qinglin**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号, Zhejiang 310024 (CN)。王雯(**WANG, Wen**); 美国加利福尼亚州桑尼维尔市阿尔马诺大道 525 号, California 94085 (US)。

- (74) 代理人: 北京博浩百睿知识产权代理有限公司(**KSK IP AGENCY**); 中国北京市海淀区知春路甲 48 号盈都大厦 C 座 1 单元 10E, Beijing 100098 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,

(54) **Title:** SPEECH PROCESSING METHOD AND SYSTEM, AND STORAGE MEDIUM

(54) 发明名称: 语音处理方法、存储介质及系统



(57) **Abstract:** A speech processing method and system, and a storage medium. The method comprises: acquiring speech information to be processed, wherein said speech information comprises a plurality of speech segments (S21); performing feature conversion on the plurality of speech segments, to obtain a conversion result (S22); performing feature decoding on the conversion result, to obtain a decoding result (S23); and determining a target speech processing result on the basis of the decoding result (S24). By means of the method, the technical problems of low processing efficiency and poor recognition accuracy when a speech data stream is processed in the relevant art are solved.



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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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: 一种语音处理方法、存储介质及系统, 该方法包括: 获取待处理的语音信息, 其中, 待处理的语音信息包括: 多个语音片段 (S21); 对多个语音片段进行特征转换, 得到转换结果 (S22); 对转换结果进行特征解码, 得到解码结果 (S23); 基于解码结果确定目标语音处理结果 (S24)。该方法解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

语音处理方法、存储介质及系统

技术领域

本公开涉及计算机技术领域，具体而言，涉及一种语音处理方法、存储介质及系统。

背景技术

语音处理是使用计算机对人类语言进行分析、识别、合成和理解的技术，随着智能化时代的到来，语音处理的应用领域越来越广泛，其对于提供更加智能化的人机交互体验具有重要意义。相关技术中采用不同的语音处理模型对语音数据流进行处理，而各种不同的语音处理模型由于构造原理不同，在进行语音处理时，均存在不同程度的缺陷，从而导致语音数据流的处理效率低下、识别准确性差。

针对上述的问题，目前尚未提出有效的解决方案。

发明内容

本公开实施例提供了一种语音处理方法、存储介质及系统，以至少解决相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

根据本公开实施例的一个方面，提供了一种语音处理方法，包括：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到转换结果；对转换结果进行特征解码，得到解码结果；基于解码结果确定目标语音处理结果。

根据本公开实施例的又一方面，还提供了一种语音处理方法，包括：获取待识别的语音信息，其中，待识别的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到多个语音特征向量；对多个语音特征向量进行特征解码，得到多个文本片段；基于多个文本片段确定目标语音识别结果。

根据本公开实施例的又一方面，还提供了一种语音处理方法，包括：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果；基于解码结果确定目标语音处理结果。

根据本公开实施例的又一方面，还提供了一种计算机可读存储介质，计算机可读存储介质包括存储的可执行程序，其中，在可执行程序运行时控制计算机可读存储介质所在设备执行本公开实施例中任意一项所述的语音处理方法。

根据本公开实施例的又一方面，还提供了一种语音处理系统，包括：处理器；以及存储器，与处理器连接，被设置为处理器提供处理以下处理步骤的指令：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到转换结果；对转换结果进行特征解码，得到解码结果；基于解码结果确定目标语音处理结果。

在本公开实施例中，通过获取待处理的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音处理结果的目的，从而实现了提升对于语音数据流进行处理时的处理效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

容易注意到的是，上面的通用描述和后面的详细描述仅仅是为了对本公开进行举例和解释，并不构成对本公开的限定。

附图说明

此处所说明的附图用来提供对本公开的进一步理解，构成本公开的一部分，本公开的示意性实施例及其说明用于解释本公开，并不构成对本公开的不当限定。在附图中：

图 1 示出了一种用于实现语音处理方法的计算机终端（或移动设备）的硬件结构框图；

图 2 是根据本公开实施例 1 的一种语音处理方法的流程图；

图 3 是根据本公开实施例 1 的一种语音处理方法的示意图；

图 4 是根据本公开实施例 1 的又一种语音处理方法的示意图；

图 5 是根据本公开实施例 2 的一种语音处理方法的流程图；

图 6 是根据本公开实施例 3 的一种语音处理方法的流程图；

图 7 是根据本公开实施例 4 的一种语音处理方法的流程图；

图 8 是根据本公开实施例 1 的一种语音处理装置的结构框图；

图 9 是根据本公开实施例 2 的一种语音处理装置的结构框图；

图 10 是根据本公开实施例 3 的一种语音处理装置的结构框图；

图 11 是根据本公开实施例 4 的一种语音处理装置的结构框图；

图 12 是根据本公开实施例 6 的一种计算机终端的结构框图。

具体实施方式

为了使本技术领域的人员更好地理解本公开方案，下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本公开一部分的实施例，而不是全部的实施例。基于本公开中的实施例，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，都应当属于本公开保护的范围。

需要说明的是，本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

本公开提供的技术方案主要采用大模型技术实现，此处的大模型是指具有大规模模型参数的深度学习模型，通常可以包含上亿、上百亿、上千亿、上万亿甚至十万亿以上的模型参数。大模型又可以称为基石模型/基础模型（Foundation Model），通过大规模无标注的语料进行大模型的预训练，产出亿级以上参数的预训练模型，这种模型能适应广泛的下游任务，模型具有较好的泛化能力，例如大规模语言模型（Large Language Model, LLM）、多模态预训练模型（multi-modal pre-training model）等。

需要说明的是，大模型在实际应用时，可以通过少量样本对预训练模型进行微调，使得大模型可以应用于不同的任务中。例如，大模型可以广泛应用于自然语言处理（Natural Language Processing, NLP）、计算机视觉等领域，具体可以应用于如视觉问答（Visual Question Answering, VQA）、图像描述（Image Caption, IC）、图像生成等计算机视觉领域任务，也可以广泛应用于基于文本的情感分类、文本摘要生成、机器翻译等自然语言处理领域任务。因此，大模型主要的应用场景包括但不限于数字助理、智能机器人、搜索、在线教育、办公软件、电子商务、智能设计等。在本公开实施例中，以语音处理场景下通过语音处理大模型进行数据处理为例进行解释说明。

首先，在对本公开实施例进行描述的过程中出现的部分名词或术语适用于如下解释：

连续时序分类（Connectionist Temporal Classification, CTC）模型：是一种用于序列分类的深度学习模型，采用端到端的训练方式，可以处理变长输入序列和输出序列。CTC 模型通过将输入序列映射到输出序列上来进行分类任务，在推断阶段，CTC 模型使用前向算法计算每个时间步上可能出现的所有标记组合及其概率，并从中选择输出结果。

基于注意力的编码器-解码器（Attention-based Encoder-Decoder, AED）模型：是一种序列到序列模型，用于将一个不定长的输入序列映射到另一个不定长的输出序列，该模型主要由编码器和解码器组成，编码器将输入序列转换为固定长度的向量表示，并能够通过这个向量向解码器传递信息，解码器使用该向量作为初始状态生成输出序列。

循环神经网络-变换器（Recurrent Neural Network-Transducer, RNN-T）模型：是一种端到端的语音识别模型，其采用序列建模和转录技术来将语音信号转化为文本。

相关技术中采用不同的语音处理模型对语音数据流进行处理，而各种不同的语音处理模型由于构造原理不同，在进行语音处理时，均存在不同程度的缺陷，从而导致语音数据流的处理效率低下、识别准确性差。具体的，CTC 流式模型框架存在独立帧假设，因此其建模上下文能力相对较弱；AED 模型需要使用额外设计的分块注意力机制来实现流式识别，而这并不适合处理语音数据流任务；RNN-T 模型的长时间依赖性较差，随着输入音频长度的增加，其识别精度可能会下降。

由此可知，相关技术中存在对语音数据流进行处理时的处理效率低、识别准确性差的技术问题，针对上述的问题，目前尚未提出有效的解决方案。

实施例 1

根据本公开实施例，还提供了一种语音处理的方法实施例，需要说明的是，在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行，并且，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。

考虑到大模型的模型参数量庞大，且移动终端的运算资源有限，本公开实施例提供的上述语音处理方法可以应用于如图 1 所示的应用场景，但不仅限于此。在如图 1 所示的应用场景中，语音处理大模型部署在服务器 10 中，服务器 10 可

以通过局域网连接、广域网连接、因特网连接，或者其他类型的数据网络，连接一个或多个客户端设备 20，此处的客户端设备 20 可以包括但不限于：智能手机、平板电脑、笔记本电脑、掌上电脑、个人计算机、智能家居设备、车载设备等。客户端设备 20 可以通过图形用户界面与用户进行交互，实现对大模型的调用，进而实现本公开实施例所提供的方法。

在本公开实施例中，客户端设备和服务器构成的系统可以执行如下步骤：服务器获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到转换结果；对转换结果进行特征解码，得到解码结果；基于解码结果确定目标语音处理结果。在客户端设备中，响应作用于图形用户界面的第一触控操作，从多种候选语音处理类型中选取目标语音处理类型，其中，多种候选语音处理类型包括：语音识别类型、语音翻译类型、语音交互类型；响应作用于图形用户界面的第二触控操作，确定目标语音处理类型对应的多个语音片段；响应作用于图形用户界面的第三触控操作，将多个语音片段转换为多个语音特征向量，对多个语音特征向量进行特征解码以得到多个解码片段，以及对多个解码片段进行拼接以得到目标语音处理结果；在图形用户界面内显示目标语音处理结果。需要说明的是，在客户端设备的运行资源能够满足语音处理大模型的部署和运行条件的情况下，本公开实施例可以在客户端设备中进行。

在上述运行环境下，本公开提供了如图 2 所示的语音处理方法。图 2 是根据本公开实施例 1 的一种语音处理方法的流程图，如图 2 所示，该方法包括以下步骤：

步骤 S21，获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；

步骤 S22，对多个语音片段进行特征转换，得到转换结果；

步骤 S23，对转换结果进行特征解码，得到解码结果；

步骤 S24，基于解码结果确定目标语音处理结果。

上述待处理的语音信息可以通过终端设备上的音频组件对用户语音进行采集而获得，如采用手机麦克风组件获得用户的多个语音片段；待处理的语音信息还可以从已经录制好的音频文件中进行提取而得到，如应用软件中的语音消息；待处理的语音信息还可以从互联网上的音频资源获取，如网络直播、网络音乐电台等；待处理的语音信息还可以为从智能设备中获取语音指令或会话信息。以上渠道所获取的待处理的语音信息都可以进行语音识别、语音合成、情感分析、说话

人识别等应用。

需要说明的是，上述待处理语音信息中的多个语音片段中可以包括连续的多对话内容或者离散的个人语音内容，本公开实施例不予限制。

进一步的，对待处理的语音信息中的多个语音片段进行特征转换，得到多个语音特征向量。其中，特征转换指的是将待处理的语音数据通过一定方式转换成新的特征表示的过程，通过特征转换可以减少数据的维度、去除冗余信息、提取有用的特征，并提高算法的鲁棒性和泛化能力。

对多个语音特征向量进行特征解码，得到解码结果，基于解码结果能够确定目标语音处理结果。在对多个语音特征向量进行特征解码的过程中，可以采用自回归解码、分类解码、生成解码、序列到序列解码、路径解码和注意力解码等方式，这些解码方式可以根据具体的任务和应用场景选择使用。

其中，自回归解码是一种常见的解码方式，通常应用于自然语言处理、语音识别等序列生成任务中。自回归解码的基本思想是在一次生成的过程中，每一步依次生成序列中的每一个元素，每次生成的结果会影响下一次的生成。这可以用一个条件概率模型来表示，如在使用 RNN 生成文本序列时，一次生成一个字符，每个字符的出现概率取决于前面已经生成的字符以及生成的条件。分类解码则是将语音特征向量映射到一个固定的类别集合中，每个类别代表一个音素、音节或单词，这种解码方式常用于语音识别任务。生成解码则是使用生成模型，例如生成对抗网络、变分自编码器，将语音特征向量映射到声音波形空间中，从而生成与输入特征相对应的语音信号。序列到序列解码则是使用循环神经网络，例如长短时记忆网络、门控循环单元，将语音特征向量序列映射到另一个语音特征向量序列，实现语音转换、语音合成等任务；路径解码则是将语音特征向量映射到一条特定的路径上，该路径可以表示音素、音节、单词或语音的时长和边界。注意力解码则是使用注意力机制，将输入的语音特征向量序列与另一个语音特征向量序列进行对齐，从而生成对应的输出序列。

基于上述步骤 S21 至步骤 S24，通过获取待处理的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音处理结果的目的，从而实现了提升对于语音数据流进行处理时的处理效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

需要说明的是，本公开实施例可以但不限于应用于健康、科学、社会、语言和艺术等领域中任何涉及语音交互的场景中，尤其涉及智能家居、智能客服、智能会议等领域内的语音识别、语音翻译等应用场景下，本公开实施例中的语音处理方法可以更加准确、高效地实现语音转换成文本的功能，进一步为用户提供更加智能化的交互体验。

下面对本公开实施例中的语音处理方法进行进一步介绍。

在一种可选的实施例中，在步骤 S22，对多个语音片段进行特征转换，得到转换结果包括：在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应。

在一种可选的实施例中，语音处理大模型包括：特征转换模型与特征解码模型。特征转换模型的主要功能是将语音信号从时域或频域转换到特征域。常见的特征转换模型包括傅里叶变换、离散余弦变换、小波变换等，它们可以提取出语音信号的频谱特征、能量特征等。特征解码模型（是语音处理中的另一个重要模型，其主要功能是将特征域的语音信号解码为原始的时域或频域信号。特征解码模型可以通过逆变换的方式将特征信号恢复为原始的语音信号，从而实现对语音信号的还原。

特征转换模型和特征解码模型通常配合使用，前者用于将语音信号转换为特征表示，后者用于将特征表示解码为原始信号。这样的处理过程可以提取语音信号中的有用信息，并减少信号的冗余，从而方便后续的语音识别、语音合成、语音增强等任务的进行。

具体的，在连续多个时刻，采用语音处理大模型依次对多个语音片段采用语音编码器进行语音编码或者采用语音分词器进行语音分词，得到多个语音特征向量。

对多个语音片段采用语音编码器进行语音编码时，可以对多个语音片段先进行预处理，如截取固定长度的语音段、进行语音信号增强、去除噪声等，进而提取语音特征，即将语音片段转换为可以被语音编码器理解的特征表示，其中，上述语音特征可以包括梅尔频率倒谱系数（Mel-Frequency Cepstral Coefficients, MFCC）、语谱图、线性预测倒谱相关系数（Linear Predictive Cepstral Coding, LPCC）等。随后构建语音编码器模型，通过建立一个神经网络模型，接受语音特征作为输入并生成编码后的语音特征向量表示作为输出，其中，上述语音编码器模型包括自编码器（Autoencoder）、变分自编码器（Variational Auto-Encoder）等。将每

个时刻对应语音片段的语音特征输入到语音编码器模型中，能够得到对应的编码后的语音特征向量，这些语音特征向量可以用于对语音片段进行识别、分类、聚类任务。

需要注意的是，不同的语音编码器模型在输入输出的参数设置和网络结构设计上可能存在差异，具体的实现方式需要根据具体的任务和数据进行调整和优化，本公开不予限制。

对多个语音片段采用语音分词器进行语音分词时，可以对多个语音片段先进行预处理，如截取固定长度的语音段、进行语音信号增强、去除噪声等，进而提取语音特征，即将语音片段转换为可以被语音编码器理解的特征表示，其中，上述语音特征可以包括 MFCC、语谱图、LPCC 等。随后构建语音分词模型，通过建立一个神经网络分词模型，接受语音特征作为输入并生成分词后的输出语音特征向量，常用的语音分词模型包括卷积神经网络、循环神经网络、卷积-循环神经网络等。将每个时刻对应语音片段的语音特征输入到语音分词模型中，得到对应的分词结果，语音分词结果的输出格式可以包括标注文本和时间戳序列，还可以对分词结果进行后处理和优化，如去除重复词汇、合并相邻词汇、进行音素级别的分词等，最后存储分词结果及相关信息，如原始语音数据、分词文本、时间戳序列等，将其作为语音片段对应的语音特征向量。

需要注意的是，对于不同的语音分词任务，要根据数据的特征和需求进行针对性的处理和模型设计。同时，语音分词模型的建立、分词结果的评价和后处理等环节也需要进行精细化调整，才能得到更准确、更符合要求的语音分词结果，本公开不予限制。

基于上述可选实施例，通过在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，从而用于流式语音识别和流式语音翻译等任务，进一步提升语音处理效率。

在一种可选的实施例中，在步骤 S23，对转换结果进行特征解码，得到解码结果包括：在连续多个时刻，采用语音处理大模型依次对多个语音特征向量进行特征解码，得到多个解码片段。

在一种可选的实施例中，特征解码模型为多模态语言模型。多模态语言模型是一种能够处理多种不同模态（如文本、图像、音频等）输入的语言模型。它能够同时理解和生成多种不同类型的数据，并能够将这些不同模态的数据进行有效的集成和交互。多模态语言模型的目标是通过结合不同模态的信息，提高语言模型在理解和生成任务中的性能和表现。

具体的，在多个连续时刻，采用基于解码器结构（Decoder）的多模态语言模型（Multi-modal Language Model）依次对多个语音特征向量进行特征解码，得到多个解码片段。在特征解码过程中，每一次解码输出会被作为下一次解码输入的一部分，因此可以利用解码输出结果，并重复进行解码来获得连续多个时刻的解码片段，直到达到特定终止条件或达到最大轮数。在得到多个解码片段后，还可以对它们进行后处理，如去重、排序、过滤无效信息等。

基于上述可选实施例，通过在连续多个时刻，采用语音处理大模型依次对多个语音特征向量进行特征解码，得到多个解码片段，能够实现语音特征向量的快速解码，进而提升语音处理效率。

在一种可选的实施例中，在连续多个时刻，采用语音处理大模型依次对多个语音特征向量进行特征解码，得到多个解码片段包括：在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

上述上下文信息是指与历史时刻语音片段与历史时刻解码片段相关联的信息。在语音识别或语音合成任务中，上下文信息包括历史时刻的语音片段、解码片段以及其相关特征向量等信息。上下文信息可以提供更多的上下文语境，帮助准确理解当前时刻的语音信息以及生成更加流畅自然的语音片段。

基于上述可选实施例，通过在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，进一步提升语音识别的准确性。

在一种可选的实施例中，在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段包括：在连续多个时刻，采用语音处理大模型依次对当前时刻语音特征向量、当前时刻语音特征向量相邻的前一时刻语音片段以及当前时刻语音特征向量相邻的前一时刻解码片段进行特征解码以得到当前时刻解码片段，直至多个语音特征向量全部解码完毕，得到多个解码片段。

具体来说，在每个时刻，当前时刻的语音特征向量是依赖于前面历史时刻的语音片段的，这些历史时刻的语音片段也就是上下文信息。同样地，在语音识别或语音合成的解码过程中，当前时刻的解码片段也会受到历史时刻解码片段的影响，因此历史时刻的解码片段也构成了上下文信息的一部分。

以连续多个时刻中的第一时刻和第二时刻为例，与第二时刻相邻的前一时刻为第一时刻，依次对第二时刻语音特征向量、第一时刻语音片段以及第一时刻解码片段进行特征解码，以得到第二时刻解码片段，直至多个语音特征向量全部解码完毕，得到多个解码片段。

基于上述可选实施例，通过在连续多个时刻，采用语音处理大模型依次对当前时刻语音特征向量、当前时刻语音特征向量相邻的前一时刻语音片段以及当前时刻语音特征向量相邻的前一时刻解码片段进行特征解码以得到当前时刻解码片段，直至多个语音特征向量全部解码完毕，快速得到多个解码片段，能够提高解码效率和解码准确性，进一步提升语音识别的准确性。

在一种可选的实施例中，对转换结果进行特征解码，得到解码结果包括：采用多模态语言模型对转换结果进行特征解码，得到解码结果；或者，采用多模态语言模型对转换结果进行非特征解码，得到解码结果。

具体的，特征解码是指利用多模态语言模型对转换结果进行解码时，使用了语音特征向量，用于辅助解码过程，通过特征解码，可以更准确地预测句子的生成概率，提高解码的准确性和效果。非特征解码则是指在解码过程中不使用语音特征向量，只依靠多模态语言模型对转换结果进行解码。

基于上述可选实施例，可以利用多模态语言模型对转换结果灵活解码，进一步提高语音识别效率和识别准确率。

在一种可选的实施例中，通过终端设备提供一图形用户界面，图形用户界面所显示的内容至少部分地包含一语音处理场景，语音处理方法还包括：

步骤 S31，响应作用于图形用户界面的第一触控操作，从多种候选语音处理类型中选取目标语音处理类型，其中，多种候选语音处理类型包括：语音识别类型、语音翻译类型、语音交互类型；

步骤 S32，响应作用于图形用户界面的第二触控操作，确定目标语音处理类型对应的多个语音片段；

步骤 S33，响应作用于图形用户界面的第三触控操作，将多个语音片段转换为多个语音特征向量，对多个语音特征向量进行特征解码以得到多个解码片段，以及对多个解码片段进行拼接以得到目标语音处理结果；

步骤 S34，在图形用户界面内显示目标语音处理结果。

在上述可选的实施例中，上述图形用户界面为语音处理界面，在语音处理界

面中所显示的语音处理场景可以是语音识别场景、语音翻译场景或者语音交互场景。

上述图形用户界面还包括第一控件（或第一触控区域），当检测到作用于第一控件（或第一触控区域）的第一触控操作时，从多种候选语音处理类型中选取目标语音处理类型，其中，多种候选语音处理类型包括：语音识别类型、语音翻译类型、语音交互类型。

上述图形用户界面还包括第二控件（或第二触控区域），当检测到作用于第二控件（或第二触控区域）的第二触控操作时，确定目标语音处理类型对应的多个语音片段。

上述图形用户界面还包括第三控件（或第三触控区域），当检测到作用于第三控件（或第三触控区域）的第三触控操作时，将多个语音片段转换为多个语音特征向量，对多个语音特征向量进行特征解码以得到多个解码片段，以及对多个解码片段进行拼接以得到目标语音处理结果。

在图形用户界面内显示目标语音处理结果，以便用户直观查看语音处理结果和语音处理进度，进一步提升用户的智能化交互体验。

需要说明的是，上述第一触控操作、上述第二触控操作和上述第三触控操作均可以是用户用手指接触上述终端设备的显示屏并触控该终端设备的操作。该触控操作可以包括单点触控、多点触控，其中，每个触控点的触控操作可以包括点击、长按、重按、划动等。上述第一触控操作、上述第二触控操作和上述第三触控操作还可以是通过鼠标、键盘等输入设备实现的触控操作。

图 3 是根据本公开实施例 1 的一种语音处理方法的示意图，如图 3 所示，该方法包括以下步骤：

步骤 S301，获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；

步骤 S302，在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应；

步骤 S303，在连续多个时刻，采用所述语音处理大模型依次对当前时刻语音特征向量、当前时刻语音特征向量相邻的前一时刻语音片段以及当前时刻语音特征向量相邻的前一时刻解码片段进行特征解码以得到当前时刻解码片段，直至多个语音特征向量全部解码完毕，得到多个解码片段；

步骤 S304，基于解码结果确定目标语音处理结果。

基于上述步骤 S301 至步骤 S304，通过获取待处理的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音处理结果的目的，从而实现了提升对于语音数据流进行处理时的处理效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

图 4 是根据本公开实施例 1 的又一种语音处理方法的示意图，如图 4 所示，在第一时刻，语音输入片段 1 (Speech chunk1) 通过语音编码器或语音分词器（也称为语音离散化）后，转换成多个语音特征向量，这些向量被送入多模态语言模型以进行特征解码，基于解码结果生成文本片段 1 (Text chunk1)。接下来，在第二时刻，语音输入片段 2 (Speech chunk2) 也通过语音编码器或语音分词器转换成语音特征向量，并输入到多模态语言模型中进行特征解码，生成文本片段 2 (Text chunk2)，以此类推，实现了流式解码，由此本公开实施例的语音识别方法可用于流式语音识别、流式语音翻译以及流式语音交互等任务。

从实际功能价值的角度来看，本公开实施例的语音识别方法在语音识别、语音翻译等场景下可以实现更加准确和高效的语音转换成文本的功能，从而提高使用体验，为用户提供更好的服务。同时，本公开实施例的语音识别方法还可以应用于语音交互领域，例如智能家居、智能客服等领域，为用户提供更加智能化的交互体验。从解决问题的角度来看，本公开实施例的语音识别方法解决了传统流式模型框架中新输入的语音特征和已生成文本之间的矛盾，提高了准确率。同时，基于解码器的流式模型框架在模型结构上更加统一，便于优化和扩展。此外，本公开实施例的语音识别方法提出了一种新型的基于解码器的流式模型框架，与传统的 CTC、AED、RNN-T 等流式模型框架有所不同，具有更好的适用性和准确性。

需要说明的是，本公开所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本公开并不受所描述的动作

顺序的限制，因为依据本公开，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本公开所必须的。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到根据上述实施例的方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件。基于这样的理解，本公开的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质（如ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端设备（可以是手机，计算机，服务器，或者网络设备等等）执行本公开各个实施例所述的方法。

实施例 2

根据本公开实施例，还提供了一种语音处理方法，图 5 是根据本公开实施例 2 的一种语音处理方法的流程图，如图 5 所示，该方法包括以下步骤：

步骤 S51，获取待识别的语音信息，其中，所述待识别的语音信息包括：多个语音片段；

步骤 S52，对所述多个语音片段进行特征转换，得到多个语音特征向量；

步骤 S53，对所述多个语音特征向量进行特征解码，得到多个文本片段；

步骤 S54，基于所述多个文本片段确定目标语音识别结果。

基于上述步骤 S51 至步骤 S54，通过获取待识别的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音识别结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音识别结果的目的，从而实现了提升对于语音数据流进行处理时的识别效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的识别效率低、识别准确性差的技术问题。

需要说明的是，本公开实施例可以但不限于应用于健康、科学、社会、语言和艺术等领域中任何涉及语音交互的场景中，尤其涉及智能家居、智能客服、智能会议等领域内的语音识别、语音翻译等应用场景下，本公开实施例中的语音处理方法可以更加准确、高效地实现语音转换成文本的功能，进一步为用户提供更加智能化的交互体验。

在本公开的上述实施例中沒有详述的部分，可以参见实施例 1 的相关描述，

不予赘述。

实施例 3

根据本公开实施例，还提供了一种语音处理方法，图 6 是根据本公开实施例 3 的一种语音处理方法的流程图，如图 6 所示，该方法包括以下步骤：

步骤 S61，获取待翻译的语音信息，其中，待翻译的语音信息包括：多个第一语音片段；

步骤 S62，对多个第一语音片段进行特征转换，得到多个语音特征向量；

步骤 S63，对多个语音特征向量进行特征解码，得到多个第二语音片段；

步骤 S64，基于多个第二语音片段确定目标语音翻译结果。

基于上述步骤 S61 至步骤 S64，通过获取待翻译的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音翻译结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音翻译结果的目的，从而实现了提升对于语音数据流进行处理时的翻译效率和翻译准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的翻译效率低、翻译准确性差的技术问题。

需要说明的是，本公开实施例可以但不限于应用于健康、科学、社会、语言和艺术等领域中任何涉及语音交互的场景中，尤其涉及智能家居、智能客服、智能会议等领域内的语音识别、语音翻译等应用场景下，本公开实施例中的语音处理方法可以更加准确、高效地实现语音转换成文本的功能，进一步为用户提供更加智能化的交互体验。

在本公开的上述实施例中沒有详述的部分，可以参见实施例 1 的相关描述，不予赘述。

实施例 4

根据本公开实施例，还提供了一种语音处理方法，图 7 是根据本公开实施例 4 的一种语音处理方法的流程图，如图 7 所示，该方法包括以下步骤：

步骤 S71，获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；

步骤 S72，采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果；

步骤 S73，基于解码结果确定目标语音处理结果。

基于上述步骤 S71 至步骤 S73，通过获取待处理的语音信息，进而采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果，最后基于解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音识别结果的目的，从而实现了提升对于语音数据流进行处理时的识别效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的识别效率低、识别准确性差的技术问题。

需要说明的是，本公开实施例可以但不限于应用于健康、科学、社会、语言和艺术等领域中任何涉及语音交互的场景中，尤其涉及智能家居、智能客服、智能会议等领域内的语音识别、语音翻译等应用场景下，本公开实施例中的语音处理方法可以更加准确、高效地实现语音转换成文本的功能，进一步为用户提供更加智能化的交互体验。

在一种可选的实施例中，采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果包括：

在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应；

在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

在本公开的上述实施例中沒有详述的部分，可以参见实施例 1 的相关描述，不予赘述。

实施例 5

根据本公开实施例，还提供了一种实施上述语音处理方法的语音处理装置，图 8 是根据本公开实施例 1 的一种语音处理装置的结构框图，如图 8 所示，该语音处理装置包括：

获取模块 801，被设置为获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；

转换模块 802，被设置为对多个语音片段进行特征转换，得到转换结果；

解码模块 803，被设置为对转换结果进行特征解码，得到解码结果；

确定模块 804，被设置为基于解码结果确定目标语音处理结果。

可选地，转换模块 802 还被设置为在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应。

可选地，解码模块 803 还被设置为在连续多个时刻，采用语音处理大模型依次对多个语音特征向量进行特征解码，得到多个解码片段。

可选地，解码模块 803 还被设置为在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

可选地，解码模块 803 还被设置为在连续多个时刻，采用语音处理大模型依次对当前时刻语音特征向量、当前时刻语音特征向量相邻的前一时刻语音片段以及当前时刻语音特征向量相邻的前一时刻解码片段进行特征解码以得到当前时刻解码片段，直至多个语音特征向量全部解码完毕，得到多个解码片段。

可选地，语音处理大模型包括：特征转换模型与特征解码模型。

可选地，特征解码模型为多模态语言模型。

可选地，解码模块 803 还被设置为：采用多模态语言模型对转换结果进行特征解码，得到解码结果；或者，采用多模态语言模型对转换结果进行非特征解码，得到解码结果。

可选地，语音处理装置还包括：响应模块 805，被设置为响应作用于图形用户界面的第一触控操作，从多种候选语音处理类型中选取目标语音处理类型，其中，多种候选语音处理类型包括：语音识别类型、语音翻译类型、语音交互类型；响应模块 805 还被设置为响应作用于图形用户界面的第二触控操作，确定目标语音处理类型对应的多个语音片段；响应模块 805 还被设置为响应作用于图形用户界面的第三触控操作，将多个语音片段转换为多个语音特征向量，对多个语音特征向量进行特征解码以得到多个解码片段，以及对多个解码片段进行拼接以得到目标语音处理结果；显示模块 806，被设置为在图形用户界面内显示目标语音处理结果。

在本公开实施例中，通过获取待处理的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音处理结果，达到了对

语音信息进行高效处理以获得多个语音片段对应的语音处理结果的目的，从而实现了提升对于语音数据流进行处理时的处理效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

此处需要说明的是，上述获取模块 801、转换模块 802、解码模块 803 和确定模块 804 对应于实施例 1 中的步骤 S21 至步骤 S24，四个模块与对应的步骤所实现的实例和应用场景相同，但不限于上述实施例 1 所公开的内容。需要说明的是，上述模块或单元可以是存储在存储器中并由一个或多个处理器处理的硬件组件或软件组件。

图 9 是根据本公开实施例 2 的一种语音处理装置的结构框图，如图 9 所示，该语音处理装置包括：

获取模块 901，被设置为获取待识别的语音信息，其中，待识别的语音信息包括：多个语音片段；

转换模块 902，被设置为对多个语音片段进行特征转换，得到多个语音特征向量；

解码模块 903，被设置为对多个语音特征向量进行特征解码，得到多个文本片段；

确定模块 904，被设置为基于多个文本片段确定目标语音识别结果。

在本公开实施例中，通过获取待识别的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音识别结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音识别结果的目的，从而实现了提升对于语音数据流进行处理时的识别效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的识别效率低、识别准确性差的技术问题。

此处需要说明的是，上述获取模块 901、转换模块 902、解码模块 903 和确定模块 904 对应于实施例 2 中的步骤 S51 至步骤 S54，四个模块与对应的步骤所实现的实例和应用场景相同，但不限于上述实施例 2 所公开的内容。需要说明的是，上述模块或单元可以是存储在存储器中并由一个或多个处理器处理的硬件组件或软件组件。

图 10 是根据本公开实施例 3 的一种语音处理装置的结构框图，如图 10 所示，

该语音处理装置包括：

获取模块 1001，被设置为获取待翻译的语音信息，其中，待翻译的语音信息包括：多个第一语音片段；

转换模块 1002，被设置为对多个第一语音片段进行特征转换，得到多个语音特征向量；

解码模块 1003，被设置为对多个语音特征向量进行特征解码，得到多个第二语音片段；

确定模块 1004，被设置为基于多个第二语音片段确定目标语音翻译结果。

在本公开实施例中，通过获取待翻译的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音翻译结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音翻译结果的目的，从而实现了提升对于语音数据流进行处理时的翻译效率和翻译准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的翻译效率低、翻译准确性差的技术问题。

此处需要说明的是，上述获取模块 1001、转换模块 1002、解码模块 1003 和确定模块 1004 对应于实施例 3 中的步骤 S61 至步骤 S64，四个模块与对应的步骤所实现的实例和应用场景相同，但不限于上述实施例 3 所公开的内容。需要说明的是，上述模块或单元可以是存储在存储器中并由一个或多个处理器处理的硬件组件或软件组件。

图 11 是根据本公开实施例 4 的一种语音处理装置的结构框图，如图 11 所示，该语音处理装置包括：

获取模块 1101，获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；

处理模块 1102，采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果；

确定模块 1103，基于解码结果确定目标语音处理结果。

可选地，处理模块 1102 还被设置为：在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应；在连续多个时刻，采用语音处理大模型依次

对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

在本公开实施例中，通过获取待处理的语音信息，进而采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果，最后基于解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音识别结果的目的，从而实现了提升对于语音数据流进行处理时的识别效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的识别效率低、识别准确性差的技术问题。

此处需要说明的是，上述获取模块 1101、处理模块 1102 和确定模块 1103 对应于实施例 4 中的步骤 S71 至步骤 S73，三个模块与对应的步骤所实现的实例和应用场景相同，但不限于上述实施例 4 所公开的内容。需要说明的是，上述模块或单元可以是存储在存储器中并由一个或多个处理器处理的硬件组件或软件组件。

实施例 6

本公开的实施例可以提供一种计算机终端，该计算机终端可以是计算机终端群中的任意一个计算机终端设备。可选地，在本实施例中，上述计算机终端也可以替换为移动终端等终端设备。

可选地，在本实施例中，上述计算机终端可以位于计算机网络的多个网络设备中的至少一个网络设备。

在本实施例中，上述计算机终端可以执行语音处理方法中以下步骤的程序代码：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到转换结果；对转换结果进行特征解码，得到解码结果；基于解码结果确定目标语音处理结果。

在本实施例中，上述计算机终端可以执行语音处理方法中以下步骤的程序代码：获取待识别的语音信息，其中，待识别的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到多个语音特征向量；对多个语音特征向量进行特征解码，得到多个文本片段；基于多个文本片段确定目标语音识别结果。

在本实施例中，上述计算机终端可以执行语音处理方法中以下步骤的程序代码：获取待翻译的语音信息，其中，待翻译的语音信息包括：多个第一语音片段；对多个第一语音片段进行特征转换，得到多个语音特征向量；对多个语音特征向

量进行特征解码，得到多个第二语音片段；基于多个第二语音片段确定目标语音翻译结果。

可选地，图 12 是根据本公开实施例 6 的一种计算机终端的结构框图。如图 12 所示，该计算机终端可以包括：一个或多个（图中仅示出一个）处理器 122、存储器 124、存储控制器、以及外设接口，其中，外设接口与射频模块、音频模块和显示器连接。

其中，存储器可被设置为存储软件程序以及模块，如本公开实施例中的语音处理方法和装置对应的程序指令/模块，处理器通过运行存储在存储器内的软件程序以及模块，从而执行各种功能应用以及数据处理，即实现上述的语音处理方法。存储器可包括高速随机存储器，还可以包括非易失性存储器，如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中，存储器可进一步包括相对于处理器远程设置的存储器，这些远程存储器可以通过网络连接至计算机终端。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

处理器可以通过传输装置调用存储器存储的信息及应用程序，以执行下述步骤：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到转换结果；对转换结果进行特征解码，得到解码结果；基于解码结果确定目标语音处理结果。

可选地，上述处理器还可以执行如下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应。

可选地，上述处理器还可以执行如下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对多个语音特征向量进行特征解码，得到多个解码片段。

可选地，上述处理器还可以执行如下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

可选地，上述处理器还可以执行如下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对当前时刻语音特征向量、当前时刻语音特征向量相邻的前一时刻语音片段以及当前时刻语音特征向量相邻的前一时刻解码片段进行特征解码以得到当前时刻解码片段，直至多个语音特征向量全部解码完毕，得到多个

解码片段。

可选地，语音处理大模型包括：特征转换模型与特征解码模型。

可选地，特征解码模型为多模态语言模型。

可选地，上述处理器还可以执行如下步骤的程序代码：采用多模态语言模型对转换结果进行特征解码，得到解码结果；或者，采用多模态语言模型对转换结果进行非特征解码，得到解码结果。

可选地，上述处理器还可以执行如下步骤的程序代码：响应作用于图形用户界面的第一触控操作，从多种候选语音处理类型中选取目标语音处理类型，其中，多种候选语音处理类型包括：语音识别类型、语音翻译类型、语音交互类型；响应作用于图形用户界面的第二触控操作，确定目标语音处理类型对应的多个语音片段；响应作用于图形用户界面的第三触控操作，将多个语音片段转换为多个语音特征向量，对多个语音特征向量进行特征解码以得到多个解码片段，以及对多个解码片段进行拼接以得到目标语音处理结果；在图形用户界面内显示目标语音处理结果。

可选地，上述处理器还可以执行如下步骤的程序代码：获取待识别的语音信息，其中，待识别的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到多个语音特征向量；对多个语音特征向量进行特征解码，得到多个文本片段；基于多个文本片段确定目标语音识别结果。

可选地，上述处理器还可以执行如下步骤的程序代码：获取待翻译的语音信息，其中，待翻译的语音信息包括：多个第一语音片段；对多个第一语音片段进行特征转换，得到多个语音特征向量；对多个语音特征向量进行特征解码，得到多个第二语音片段；基于多个第二语音片段确定目标语音翻译结果。

可选地，上述处理器还可以执行如下步骤的程序代码：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果；基于解码结果确定目标语音处理结果。

可选地，上述处理器还可以执行如下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应；在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关

联的历史时刻语音片段与历史时刻解码片段确定。

采用本公开实施例，通过获取待处理的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音处理结果的目的，从而实现了提升对于语音数据流进行处理时的处理效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

本领域普通技术人员可以理解，图 12 所示的结构仅为示意，计算机终端也可以是智能手机（如 Android 手机、iOS 手机等）、平板电脑、掌上电脑以及移动互联网设备（Mobile Internet Devices, MID）等终端设备。图 12 其并不对上述电子装置的结构造成限定。例如，计算机终端还可包括比图 12 中所示更多或者更少的组件（如网络接口、显示装置等），或者具有与图 12 所示不同的配置。

本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令终端设备相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括：闪存盘、只读存储器（Read-Only Memory, ROM）、随机存取器（Random Access Memory, RAM）、磁盘或光盘等。

实施例 7

本公开的实施例还提供了一种计算机可读存储介质。可选地，在本实施例中，上述计算机可读存储介质可以保存上述实施例所提供的语音处理方法所执行的程序代码。

可选地，在本实施例中，上述存储介质可以位于计算机网络中计算机终端群中的任意一个计算机终端中，或者位于移动终端群中的任意一个移动终端中。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到转换结果；对转换结果进行特征解码，得到解码结果；基于解码结果确定目标语音处理结果。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代

码：在连续多个时刻，采用语音处理大模型依次对多个语音特征向量进行特征解码，得到多个解码片段。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对当前时刻语音特征向量、当前时刻语音特征向量相邻的前一时刻语音片段以及当前时刻语音特征向量相邻的前一时刻解码片段进行特征解码以得到当前时刻解码片段，直至多个语音特征向量全部解码完毕，得到多个解码片段。

可选地，语音处理大模型包括：特征转换模型与特征解码模型。

可选地，特征解码模型为多模态语言模型。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：采用多模态语言模型对转换结果进行特征解码，得到解码结果；或者，采用多模态语言模型对转换结果进行非特征解码，得到解码结果。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：响应作用于图形用户界面的第一触控操作，从多种候选语音处理类型中选取目标语音处理类型，其中，多种候选语音处理类型包括：语音识别类型、语音翻译类型、语音交互类型；响应作用于图形用户界面的第二触控操作，确定目标语音处理类型对应的多个语音片段；响应作用于图形用户界面的第三触控操作，将多个语音片段转换为多个语音特征向量，对多个语音特征向量进行特征解码以得到多个解码片段，以及对多个解码片段进行拼接以得到目标语音处理结果；在图形用户界面内显示目标语音处理结果。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：获取待识别的语音信息，其中，待识别的语音信息包括：多个语音片段；对多个语音片段进行特征转换，得到多个语音特征向量；对多个语音特征向量进行特征解码，得到多个文本片段；基于多个文本片段确定目标语音识别结果。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：获取待翻译的语音信息，其中，待翻译的语音信息包括：多个第一语音片段；对多个第一语音片段进行特征转换，得到多个语音特征向量；对多个语音特征向

量进行特征解码，得到多个第二语音片段；基于多个第二语音片段确定目标语音翻译结果。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：获取待处理的语音信息，其中，待处理的语音信息包括：多个语音片段；采用语音处理大模型对多个语音片段进行特征转换以得到转换结果，以及对转换结果进行特征解码以得到解码结果；基于解码结果确定目标语音处理结果。

可选地，在本实施例中，存储介质被设置为存储用于执行以下步骤的程序代码：在连续多个时刻，采用语音处理大模型依次对多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，多个时刻与多个语音片段一一对应；在连续多个时刻，采用语音处理大模型依次对多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，上下文信息由多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

采用本公开实施例，通过获取待处理的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音处理结果的目的，从而实现了提升对于语音数据流进行处理时的处理效率和识别准确性的技术效果，进而解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

实施例 8

本公开的实施例还提供了一种语音处理系统，包括：

处理器；以及存储器，与所述处理器连接，被设置为所述处理器提供处理以下处理步骤的指令：

获取待处理的语音信息，其中，所述待处理的语音信息包括：多个语音片段；对所述多个语音片段进行特征转换，得到转换结果；对所述转换结果进行特征解码，得到解码结果；基于所述解码结果确定目标语音处理结果。

采用本公开实施例，通过获取待处理的语音信息中的多个语音片段，进而对所述多个语音片段进行特征转换，得到转换结果，随后对所述转换结果进行特征解码，得到解码结果，最后基于所述解码结果确定目标语音处理结果，达到了对语音信息进行高效处理以获得多个语音片段对应的语音处理结果的目的，从而实现了提升对于语音数据流进行处理时的处理效率和识别准确性的技术效果，进而

解决了相关技术中对语音数据流进行处理时的处理效率低、识别准确性差的技术问题。

上述本公开实施例序号仅仅为了描述，不代表实施例的优劣。

在本公开的上述实施例中，对各个实施例的描述都各有侧重，某个实施例中沒有详述的部分，可以参见其他实施例的相关描述。

在本公开所提供的几个实施例中，应该理解到，所揭露的技术内容，可通过其它的方式实现。其中，以上所描述的装置实施例仅仅是示意性的，例如所述单元的划分，仅仅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。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以通过一些接口，单元或模块的间接耦合或通信连接，可以是电性或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本公开各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本公开的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可为个人计算机、服务器或者网络设备等）执行本公开各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述仅是本公开的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本公开原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本公开的保护范围。

权 利 要 求 书

1. 一种语音处理方法，包括：

获取待处理的语音信息，其中，所述待处理的语音信息包括：多个语音片段；
对所述多个语音片段进行特征转换，得到转换结果；
对所述转换结果进行特征解码，得到解码结果；
基于所述解码结果确定目标语音处理结果。
2. 根据权利要求 1 所述的语音处理方法，其中，对所述多个语音片段进行特征转换，得到所述转换结果包括：

在连续多个时刻，采用语音处理大模型依次对所述多个语音片段进行语音编码或者语音分词，得到多个语音特征向量，其中，所述多个时刻与所述多个语音片段一一对应。
3. 根据权利要求 2 所述的语音处理方法，其中，对所述转换结果进行特征解码，得到所述解码结果包括：

在所述连续多个时刻，采用所述语音处理大模型依次对所述多个语音特征向量进行特征解码，得到多个解码片段。
4. 根据权利要求 3 所述的语音处理方法，其中，在所述连续多个时刻，采用所述语音处理大模型依次对所述多个语音特征向量进行特征解码，得到所述多个解码片段包括：

在所述连续多个时刻，采用所述语音处理大模型依次对所述多个语音特征向量和上下文信息进行特征解码，得到所述多个解码片段，其中，所述上下文信息由所述多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。
5. 根据权利要求 4 所述的语音处理方法，其中，在所述连续多个时刻，采用所述语音处理大模型依次对所述多个语音特征向量和所述上下文信息进行特征解码，得到所述多个解码片段包括：

在所述连续多个时刻，采用所述语音处理大模型依次对所述当前时刻语音特征向量、所述当前时刻语音特征向量相邻的前一时刻语音片段以及所述当前时刻语音特征向量相邻的前一时刻解码片段进行特征解码以得到当前时刻解码片段，直至所述多个语音特征向量全部解码完毕，得到所述多个解码片段。
6. 根据权利要求 2 所述的语音处理方法，其中，所述语音处理大模型包括：特征转换模型与特征解码模型。
7. 根据权利要求 6 所述的语音处理方法，其中，所述特征解码模型为多模态语言模型。

8. 根据权利要求 7 所述的语音处理方法, 其中, 对所述转换结果进行特征解码, 得到所述解码结果包括:

采用所述多模态语言模型对所述转换结果进行特征解码, 得到所述解码结果;
或者,

采用所述多模态语言模型对所述转换结果进行非特征解码, 得到所述解码结果。

9. 根据权利要求 1 所述的语音处理方法, 其中, 通过终端设备提供一图形用户界面, 所述图形用户界面所显示的内容至少部分地包含一语音处理场景, 所述语音处理方法还包括:

响应作用于所述图形用户界面的第一触控操作, 从多种候选语音处理类型中选取目标语音处理类型, 其中, 所述多种候选语音处理类型包括: 语音识别类型、语音翻译类型、语音交互类型;

响应作用于所述图形用户界面的第二触控操作, 确定所述目标语音处理类型对应的所述多个语音片段;

响应作用于所述图形用户界面的第三触控操作, 将所述多个语音片段转换为所述多个语音特征向量, 对所述多个语音特征向量进行特征解码以得到多个解码片段, 以及对所述多个解码片段进行拼接以得到所述目标语音处理结果;

在所述图形用户界面内显示所述目标语音处理结果。

10. 一种语音处理方法, 包括:

获取待识别的语音信息, 其中, 所述待识别的语音信息包括: 多个语音片段;

对所述多个语音片段进行特征转换, 得到多个语音特征向量;

对所述多个语音特征向量进行特征解码, 得到多个文本片段;

基于所述多个文本片段确定目标语音识别结果。

11. 一种语音处理方法, 包括:

获取待处理的语音信息, 其中, 所述待处理的语音信息包括: 多个语音片段;

采用语音处理大模型对所述多个语音片段进行特征转换以得到转换结果, 以及对所述转换结果进行特征解码以得到解码结果;

基于所述解码结果确定目标语音处理结果。

12. 根据权利要求 11 所述的语音处理方法, 其中, 采用所述语音处理大模型对所述多个语音片段进行特征转换以得到所述转换结果, 以及对所述转换结果进行特征解码以得到所述解码结果包括:

在连续多个时刻, 采用语音处理大模型依次对所述多个语音片段进行语音编

码或者语音分词，得到多个语音特征向量，其中，所述多个时刻与所述多个语音片段一一对应；

在所述连续多个时刻，采用所述语音处理大模型依次对所述多个语音特征向量和上下文信息进行特征解码，得到多个解码片段，其中，所述上下文信息由所述多个语音特征向量中当前时刻语音特征向量相关联的历史时刻语音片段与历史时刻解码片段确定。

13. 一种计算机可读存储介质，所述计算机可读存储介质包括存储的可执行程序，其中，在所述可执行程序运行时控制所述计算机可读存储介质所在设备执行权利要求 1 至 12 中任意一项所述的语音处理方法。

14. 一种语音处理系统，包括：

处理器；以及

存储器，与所述处理器连接，被设置为所述处理器提供处理以下处理步骤的指令：

获取待处理的语音信息，其中，所述待处理的语音信息包括：多个语音片段；

对所述多个语音片段进行特征转换，得到转换结果；

对所述转换结果进行特征解码，得到解码结果；

基于所述解码结果确定目标语音处理结果。

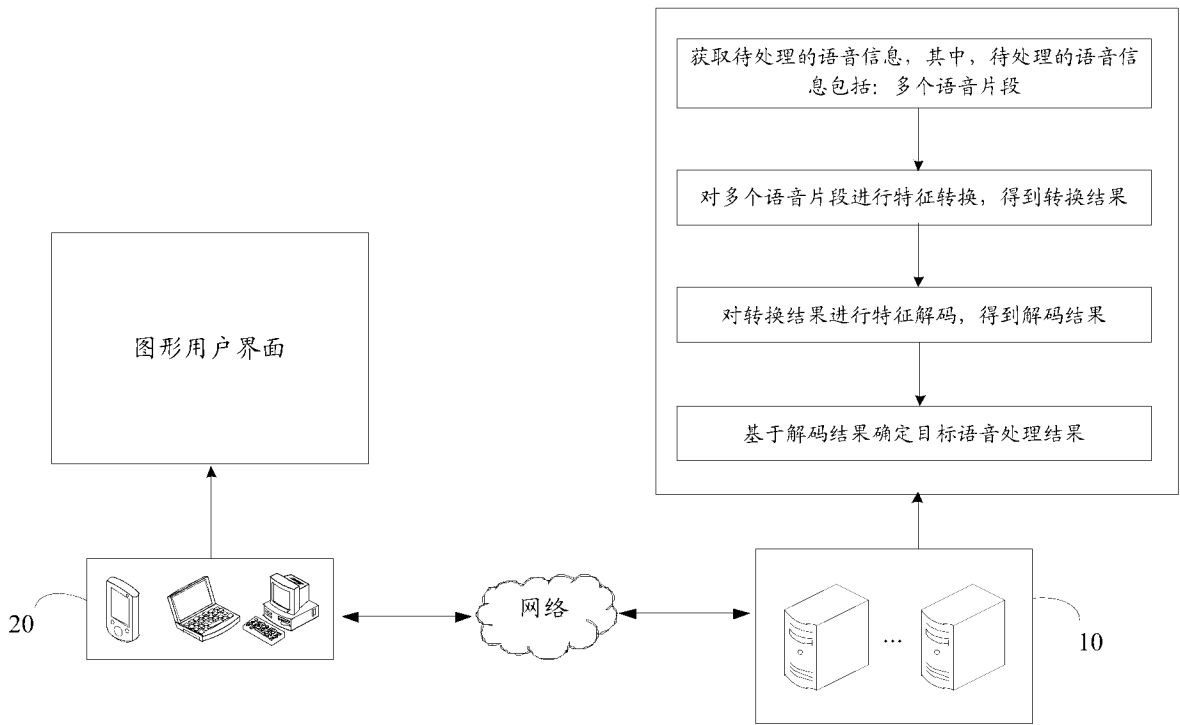


图 1

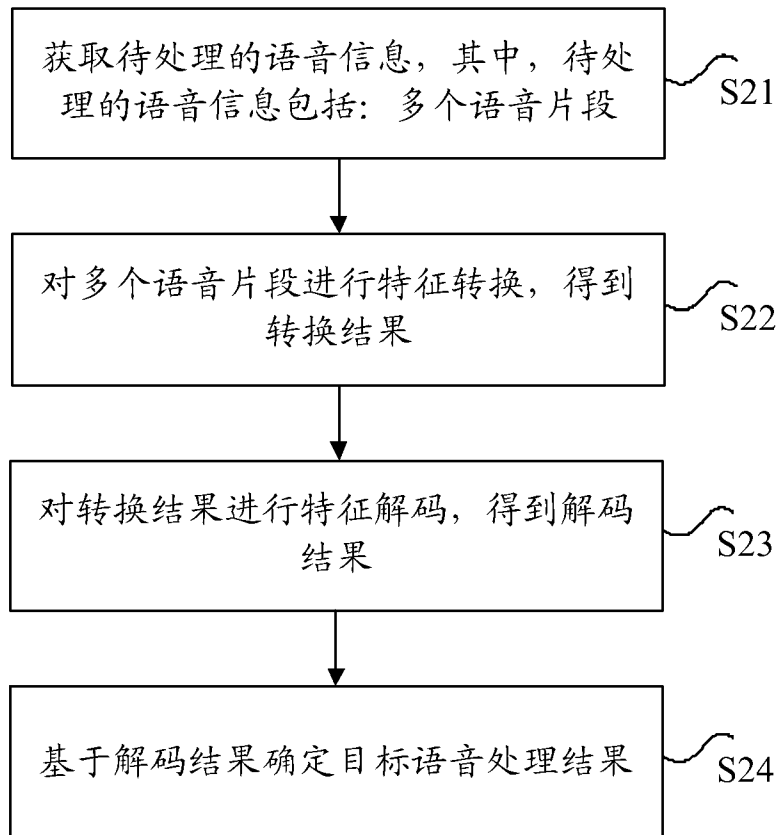


图 2

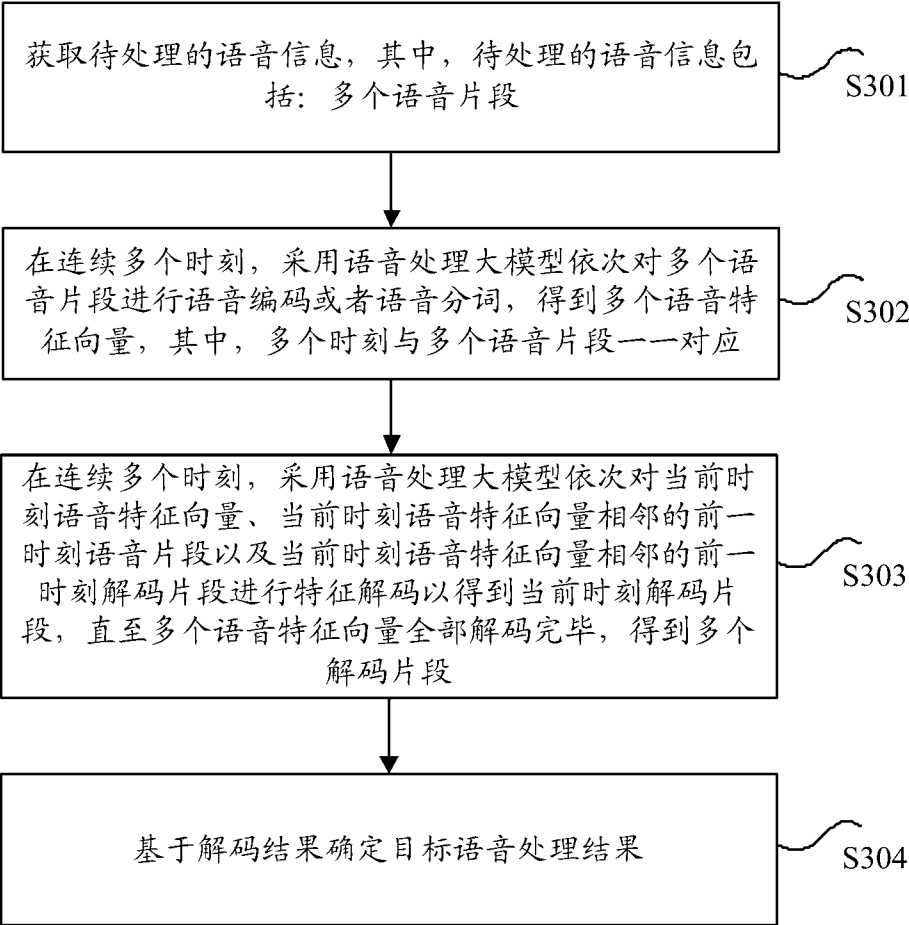


图 3

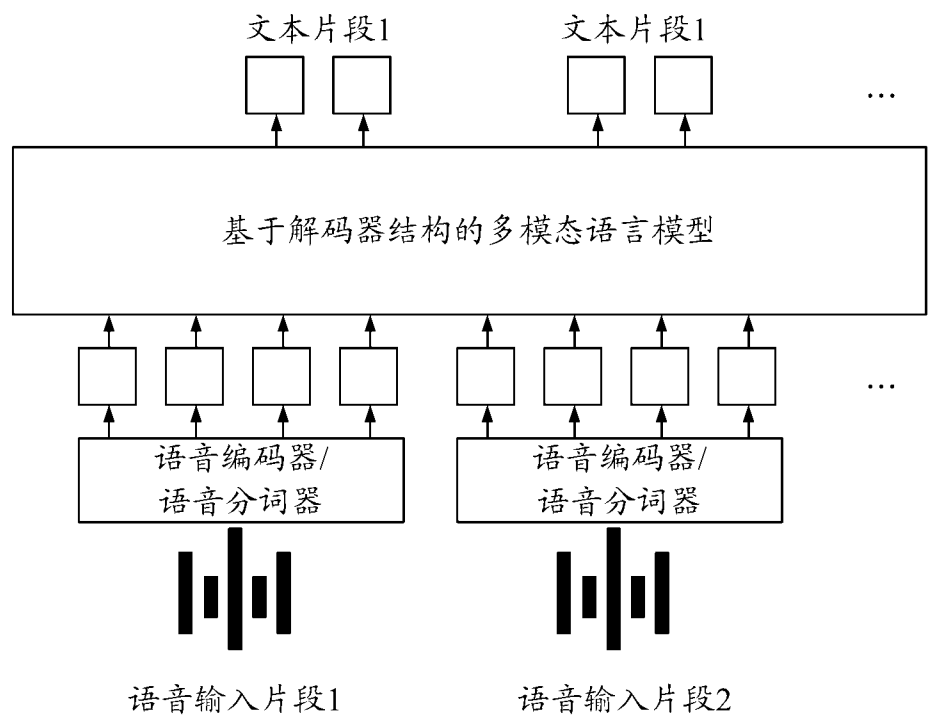


图 4

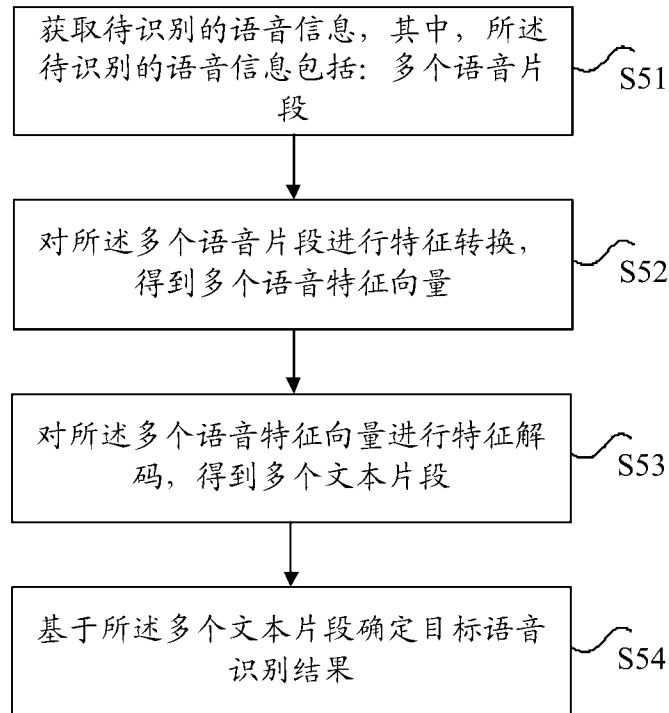


图 5

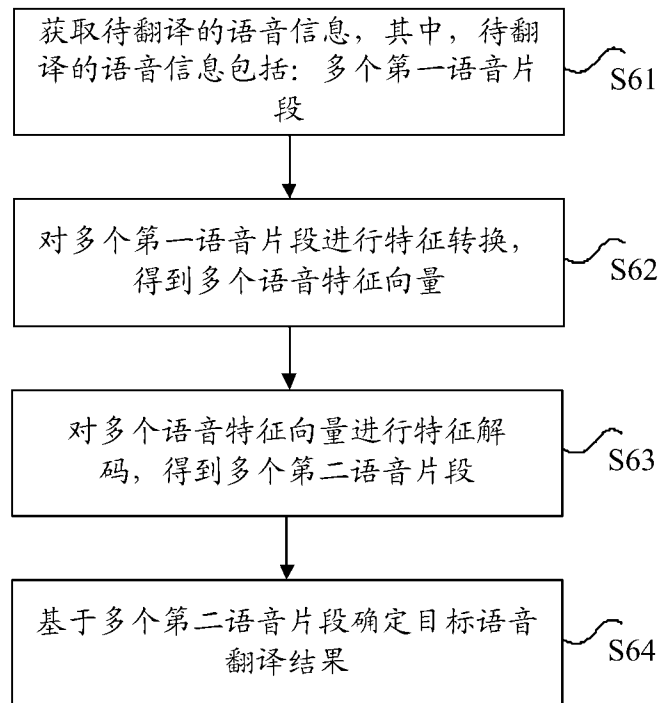


图 6

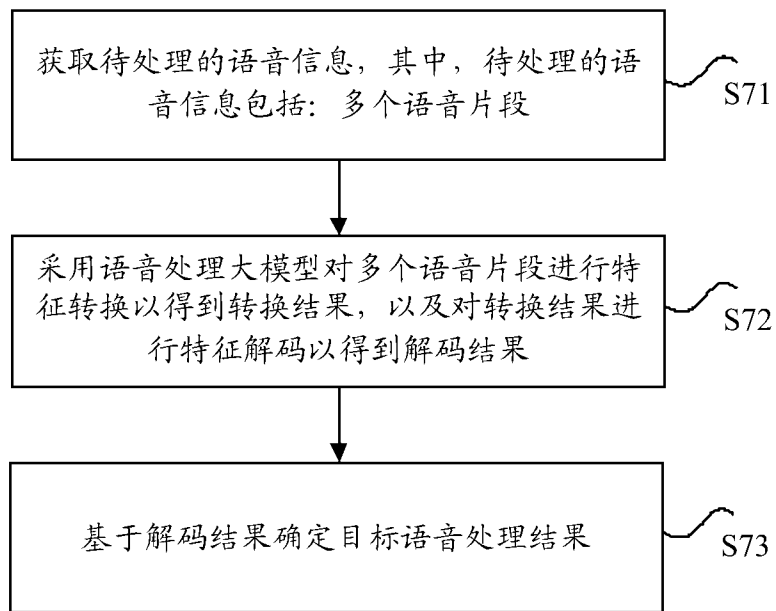


图 7



图 8



图 9



图 10



图 11

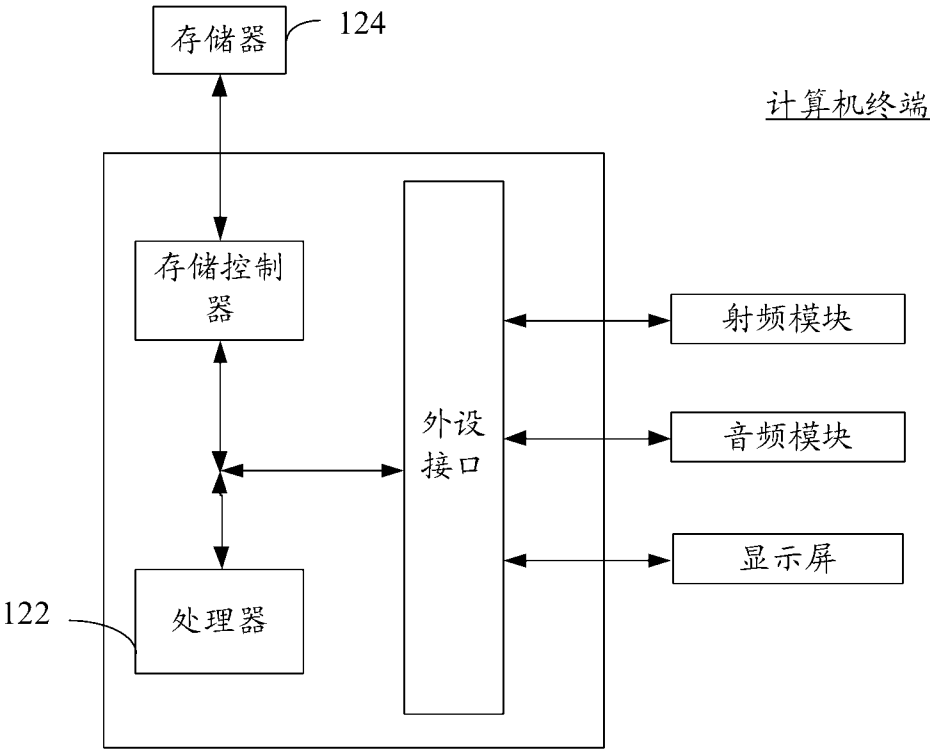


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/105264

A. CLASSIFICATION OF SUBJECT MATTER G10L15/02(2006.01)i; G10L15/16(2006.01)i; G10L25/30(2013.01)i; G10L25/03(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) IPC: G10L, G06F 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) CNABS, CNTXT, ENTXTC, WPABSC, ENTXT: 语音, 数据流, 流式, 语音流, 语音片段, 特征变换, 特征转换, 特征向量, 编码, 解码, 译码, 多模态, 上下文, 相邻, 历史, speech, voice, flow, continuous, segment, feature, cod+, decod+, multi?mode, context, history, prior		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116798409 A (HANGZHOU ALIBABA FEITIAN INFORMATION TECHNOLOGY CO., LTD.) 22 September 2023 (2023-09-22) claims 1-14	1-14
X	CN 112382278 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 19 February 2021 (2021-02-19) description, paragraphs 0035-0104, and figures 1-9	1-6, 9-14
Y	CN 112382278 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 19 February 2021 (2021-02-19) description, paragraphs 0035-0104, and figures 1-9	7-8
Y	CN 113270086 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 17 August 2021 (2021-08-17) description, paragraph 0023, and figure 1	7-8
X	CN 114974237 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 30 August 2022 (2022-08-30) description, paragraphs 0036-0196, and figures 1-8	1-6, 9-14
☒ 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 “D” document cited by the applicant in the international application “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 04 November 2024		Date of mailing of the international search report 05 November 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/ CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/105264

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 114974237 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 30 August 2022 (2022-08-30) description, paragraphs 0036-0196, and figures 1-8	7-8
A	CN 113470620 A (QINGDAO DONGTING INTELLIGENT TECHNOLOGY CO., LTD.) 01 October 2021 (2021-10-01) entire document	1-14
A	JP 2021015189 A (CHUBU ELECTRIC POWER et al.) 12 February 2021 (2021-02-12) entire document	1-14

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2024/105264

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116798409	A	22 September 2023	None			
CN	112382278	A	19 February 2021	None			
CN	113270086	A	17 August 2021	None			
CN	114974237	A	30 August 2022	None			
CN	113470620	A	01 October 2021	None			
JP	2021015189	A	12 February 2021	JP	7414231	B2	16 January 2024

A. 主题的分类

G10L15/02(2006.01)i; G10L15/16(2006.01)i; G10L25/30(2013.01)i; G10L25/03(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)

IPC: G10L, G06F

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))

CNABS, CNTXT, ENTXTC, WPABSC, ENTXT: 语音, 数据流, 流式, 语音流, 语音片段, 特征变换, 特征转换, 特征向量, 编码, 解码, 译码, 多模态, 上下文, 相邻, 历史, speech, voice, flow, continuous, segment, feature, cod+, decod+, multi?mode, context, history, prior

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 116798409 A (杭州阿里巴巴飞天信息技术有限公司) 2023年9月22日 (2023 - 09 - 22) 权利要求1-14	1-14
X	CN 112382278 A (北京百度网讯科技有限公司) 2021年2月19日 (2021 - 02 - 19) 说明书第0035-0104段及图1-9	1-6、9-14
Y	CN 112382278 A (北京百度网讯科技有限公司) 2021年2月19日 (2021 - 02 - 19) 说明书第0035-0104段及图1-9	7-8
Y	CN 113270086 A (中国科学院自动化研究所) 2021年8月17日 (2021 - 08 - 17) 说明书第0023段及图1	7-8
X	CN 114974237 A (北京百度网讯科技有限公司) 2022年8月30日 (2022 - 08 - 30) 说明书第0036-0196段及图1-8	1-6、9-14
Y	CN 114974237 A (北京百度网讯科技有限公司) 2022年8月30日 (2022 - 08 - 30) 说明书第0036-0196段及图1-8	7-8
A	CN 113470620 A (青岛洞听智能科技有限公司) 2021年10月1日 (2021 - 10 - 01) 全文	1-14

☒ 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

“D” 申请人在国际申请中引证的文件

“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

2024年11月4日

国际检索报告邮寄日期

2024年11月5日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

喻新

电话号码 (+86) 010-62089308

PCT/ISA/210 表(第2页) (2022年7月)

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	JP 2021015189 A (CHUBU ELECTRIC POWER 等) 2021年2月12日 (2021 - 02 - 12) 全文	1-14

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/105264

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	116798409	A	2023年9月22日	无			
CN	112382278	A	2021年2月19日	无			
CN	113270086	A	2021年8月17日	无			
CN	114974237	A	2022年8月30日	无			
CN	113470620	A	2021年10月1日	无			
JP	2021015189	A	2021年2月12日	JP	7414231	B2	2024年1月16日