

音频编码方法、音频解码方法、装置、设备及存储介质

相关申请的交叉引用

5 本申请实施例基于申请号为 202310597138.3、申请日为 2023 年 05 月 24 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的全部内容在此引入本申请实施例作为参考。

技术领域

本申请涉及人工智能技术，尤其涉及一种音频编码方法、音频解码方法、装置、设备及存储介质。

背景技术

10 人工智能（AI，Artificial Intelligence）是计算机科学的一个综合技术，通过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，例如自然语言处理技术以及机器学习/深度学习等几大方向，随着技术的发展，人工智能技术将在更多的领域得到应用，并发挥越来越重要的价值。

15 音频编解码技术是人工智能领域的重要应用之一，音频编解码技术是包括远程音视频通话在内的通信服务中的一项核心技术。语音编码技术，简单讲，就是使用较少的网络带宽资源去尽量多的传递语音信息。从香农信息论的角度来讲，语音编码是一种信源编码，信源编码的目的是在编码端尽可能的压缩我们想要传递信息的数据量，去掉信息中的冗余，同时在解码端还能够无损（或接近无损）的恢复出来。

20 相关技术中，在编码的过程中，为了保证音频质量，音频编码的效率会大打折扣。

发明内容

本申请实施例提供一种音频编码方法、音频解码方法、装置、电子设备、计算机可读存储介质及计算机程序产品，能够在保证音频质量的情况下，提高音频编码效率。

本申请实施例的技术方案是这样实现的：

25 本申请实施例提供一种音频编码方法，应用于电子设备，包括：

对音频信号进行下采样处理，得到所述音频信号的低频信号；

对所述低频信号进行低频特征提取处理，得到所述音频信号的低频特征；

对所述音频信号进行高频分析处理，得到所述音频信号的高频特征；

其中，所述高频特征的特征维度低于所述低频特征的特征维度；

30 对所述低频特征进行编码处理，得到所述音频信号的低频码流，并对所述高频特征进行编码处理，得到所述音频信号的高频码流。

本申请实施例提供一种音频解码方法，应用于电子设备，包括：

对音频信号的低频码流进行解码处理，得到所述低频码流对应的低频特征，并对所述音频信号的高频码流进行解码处理，得到所述高频码流对应的高频特征；

35 其中，所述低频码流是通过所述音频信号经过下采样后得到的低频信号进行编码得到的，所述高频特征的特征维度低于所述低频特征的特征维度；

对所述低频特征进行低频特征重建处理，得到所述低频特征对应的低频信号；

- 对所述低频信号进行上采样处理，得到所述低频信号的上采样信号；
- 对所述高频特征以及所述上采样信号进行信号重建处理，得到所述低频码流和所述高频码流对应的合成音频信号。
- 上述技术方案中，所述上采样处理是通过多个级联的解码层实现的；
- 5 所述对所述卷积特征进行上采样处理，得到所述低频特征的上采样特征，包括：
通过所述多个级联的解码层中的第一个解码层，对所述卷积特征进行上采样处理；
将所述第一个解码层的上采样结果输出到后续级联的解码层，通过所述后续级联的解码层继续进行上采样处理和上采样结果输出，直至输出到最后一个解码层；
将所述最后一个解码层输出的上采样结果确定为所述低频特征的上采样特征。
- 10 本申请实施例提供一种音频编码装置，包括：
下采样模块，配置为对音频信号进行下采样处理，得到所述音频信号的低频信号；
低频提取模块，配置为对所述低频信号进行低频特征提取处理，得到所述音频信号的低频特征；
高频分析模块，配置为对所述音频信号进行高频分析处理，得到所述音频信号的高频特征；
- 15 其中，所述高频特征的特征维度低于所述低频特征的特征维度；
编码模块，配置为对所述低频特征进行编码处理，得到所述音频信号的低频码流，并对所述高频特征进行编码处理，得到所述音频信号的高频码流。
- 本申请实施例提供一种音频解码装置，包括：
- 20 解码模块，配置为对音频信号的低频码流进行解码处理，得到所述低频码流对应的低频特征，并对所述音频信号的高频码流进行解码处理，得到所述高频码流对应的高频特征；
其中，所述低频码流是通过对所述音频信号经过下采样后得到的低频信号进行编码得到的，所述高频特征的特征维度低于所述低频特征的特征维度；
- 25 低频重建模块，配置为对所述低频特征进行低频特征重建处理，得到所述低频特征对应的低频信号；
上采样模块，配置为对所述低频信号进行上采样处理，得到所述低频信号的上采样信号；
信号重建模块，配置为对所述高频特征以及所述上采样信号进行信号重建处理，得到所述低频码流和所述高频码流对应的合成音频信号。
- 30 本申请实施例提供一种电子设备，所述电子设备包括：
存储器，用于存储计算机程序或计算机可执行指令；
处理器，用于执行所述存储器中存储的计算机程序或计算机可执行指令时，实现本申请实施例提供的音频编码方法或音频解码方法。
- 35 本申请实施例提供一种计算机可读存储介质，存储有计算机程序或计算机可执行指令，所述计算机程序或计算机可执行指令被处理器执行时实现本申请实施例提供的音频编码方法或音频解码方法。
- 本申请实施例提供一种计算机程序产品，包括计算机程序或计算机可执行指令，所述计算机程序或计算机可执行指令被处理器执行时实现本申请实施例提供的音频编码方法或音频解码方法。
- 40 本申请实施例具有以下有益效果：
通过对音频信号进行下采样，以得到低频信号，由于低频信号相对于音频信号中的高频信号对音频编码的影响更大，因此通过差异化的信号处理，分别提取音频信号的低频特征以及高频特征，使得高频特征的特征维度低于低频特征的特征维度，对减小了特

征维度的低频特征以及高频特征分别进行编码，从而在保证音频质量的情况下，提高音频编码效率。

附图说明

- 图 1 是本申请实施例提供的不同码率下的频谱比较示意图；
- 图 2 是本申请实施例提供的音频编解码系统的架构示意图；
- 图 3A-图 3B 是本申请实施例提供的电子设备的结构示意图；
- 图 4A 是本申请实施例提供的普通卷积网络的示意图；
- 图 4B 是本申请实施例提供的空洞卷积网络的示意图；
- 图 5 是本申请实施例提供的频带扩展的示意图；
- 图 6A 是本申请实施例提供的音频编码方法的第一流程示意图；
- 图 6B 是本申请实施例提供的音频编码方法的第二流程示意图；
- 图 6C 是本申请实施例提供的音频编码方法的第三流程示意图；
- 图 6D 是本申请实施例提供的音频编码方法的第四流程示意图；
- 图 7A 是本申请实施例提供的音频解码方法的第一流程示意图；
- 图 7B 是本申请实施例提供的音频解码方法的第二流程示意图；
- 图 7C 是本申请实施例提供的音频解码方法的第三流程示意图；
- 图 7D 是本申请实施例提供的音频解码方法的第四流程示意图；
- 图 8 是本申请实施例提供的端到端的语音通信链路示意图；
- 图 9 是本申请实施例提供的低速率神经网络编解码方法的流程示意图；
- 图 10 是本申请实施例提供的第一神经网络的示意图；
- 图 11 是本申请实施例提供的第二神经网络的示意图。

具体实施方式

为了使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请作进一步地详细描述，所描述的实施例不应视为对本申请的限制，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它实施例，都属于本申请保护的范围。

在以下的描述中，所涉及的术语“第一\第二”仅仅是是区别类似的对象，不代表针对对象的特定排序，可以理解地，“第一\第二”在允许的情况下可以互换特定的顺序或先后次序，以使这里描述的本申请实施例能够以除了在这里图示或描述的以外的顺序实施。

在以下的描述中，涉及到“一些实施例”，其描述了所有可能实施例的子集，但是可以理解，“一些实施例”可以是所有可能实施例的相同子集或不同子集，并且可以在不冲突的情况下相互结合。

除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中所使用的术语只是为了描述本申请实施例的目的，不是旨在限制本申请。

对本申请实施例进行进一步详细说明之前，对本申请实施例中涉及的名词和术语进行说明，本申请实施例中涉及的名词和术语适用于如下的解释。

1) 神经网络 (NN, Neural Network): 是一种模仿动物神经网络行为特征，进行分布式并行信息处理的算法数学模型。这种网络依靠系统的复杂程度，通过调整内部大量节点之间相互连接的关系，从而达到处理信息的目的。

2) 深度学习 (DL, Deep Learning): 是机器学习 (ML, Machine Learning) 领域中

一个新的研究方向，深度学习是学习样本数据的内在规律和表示层次，这些学习过程中获得的信息对诸如文字，图像和声音等数据的解释有很大的帮助。它的最终目标是让机器能够像人一样具有分析学习能力，能够识别文字、图像和声音等数据。

5 3) 量化：是指将信号的连续取值（或者大量离散取值）近似为有限多个（或较少的）离散值的过程。其中，量化包括矢量量化（VQ, Vector Quantization）以及标量量化。

其中，矢量量化是一种有效的有损压缩技术，其理论基础是香农的速率失真理论。矢量量化的基本原理是用码书中与输入矢量最匹配的码字的索引代替输入矢量进行传输与存储，而解码时仅需要简单地查表操作。例如，将若干个标量数据组成一个矢量空间，将矢量空间划分为若干个小区域，量化时将落入小区域的矢量，使用对应的索引代替输入矢量。

标量量化是对标量进行量化，即一维的矢量量化，将动态范围分成若干个小区域，每个小区域有一个代表值（即索引）。当输入信号落入某区间时，将输入信号量化成该代表值。

15 4) 熵编码：编码过程中按熵原理不丢失任何信息的无损编码方式，也是有损编码中的一个关键模块，处于编码器的末端。熵编码包括香农（Shannon）编码、哈夫曼（Huffman）编码、指数哥伦布编码（Exp-Golomb）和算术编码（arithmetic coding）。

语音编码技术就是使用较少的网络带宽资源去尽量多的传递语音信息。语音编解码器的压缩率可以达到 10 倍以上，也就是原本 10MB 的语音数据经过编码器的压缩后，只需要 1MB 来传输，大大降低了传递信息所需消耗的带宽资源。例如，对于采样率为 16000Hz 的宽带语音信号，如果采用 16-bit 采样深度（取样中对语音强度记录的精细程度），无压缩版本的码率（单位时间内传送数据量）为 256kbps；如果使用语音编码技术，即使是有损编码，在 10-20kbps 的码率范围内，重建的语音信号的质量可以接近无压缩版本，甚至听感上认为无差别。如果需要更高采样率的服务，比如 32000Hz 的超宽带语音，码率范围至少要达到 30kbps 以上。

在通信系统中，为了保证通信的顺利，行业内部部署标准的语音编解码协议，例如来自 ITU-T、3GPP、IETF、AVS、CCSA 等国际国内标准组织的标准，G.711、G.722、AMR 系列、EVS、OPUS 等标准。图 1 给出一个不同码率下的频谱比较示意图，以示范压缩码率与质量的关系。曲线 101 为原始语音的频谱曲线，即没有压缩的信号；曲线 102 为 OPUS 编码器在 20kbps 码率下的频谱曲线；曲线 103 为 OPUS 编码在 6kbps 码率下的频谱曲线。由图 1 可知，随着编码码率的提升，压缩后的信号更为接近原始信号。

相关技术中，语音编码原理大致如下：语音编码可以直接对语音波形样本，逐样本地进行编码；或者，基于人的发声原理，提取相关低维度特征，编码端对特征进行编码，解码端基于这些参数重建语音信号。

上述编码原理均来自语音信号建模，即基于信号处理的压缩方法，无法保证音频的编码质量。为了在保证语音质量的情况下，提高编码效率，本申请实施例提供一种基于人工智能的音频编码方法、音频解码方法、装置、电子设备、计算机可读存储介质及计算机程序产品。下面说明本申请实施例提供的电子设备的示例性应用，本申请实施例提供的电子设备可以实施为终端设备，也可以实施为服务器，或者由终端设备和服务器协同实施。下面以电子设备实施为终端设备为例进行说明。

示例的，参见图 2，图 2 是本申请实施例提供的音频编解码系统 10 的架构示意图，音频编解码系统 10 包括：服务器 200、网络 300、终端设备 400（即编码端）和终端设备 500（即解码端），其中，网络 300 可以是局域网，或者是广域网，又或者是二者的组合。

45 在一些实施例中，在终端设备 400 上运行有客户端 410，客户端 410 可以是各种类

型的客户端，例如即时通信客户端、网络会议客户端、直播客户端、浏览器等。客户端 410 响应于发送方（例如网络会议的发起者、主播、语音通话的发起者等）触发的音频采集指令，调用终端设备 400 自带的麦克风进行音频信号的采集，并对采集得到的音频信号进行音频编码处理，得到码流（高频码流以及低频码流）。

5 例如，客户端 410 调用本申请实施例提供的音频编码方法对采集得到的音频信号进行编码，即对音频信号进行下采样处理，得到音频信号的低频信号；对低频信号进行低频特征提取处理，得到音频信号的低频特征；对音频信号进行高频分析处理，得到音频信号的高频特征；其中，高频特征的特征维度低于低频特征的特征维度；对低频特征进行编码处理，得到音频信号的低频码流，并对高频特征进行编码处理，得到音频信号的高频码流。编码端（即终端设备 400）结合信号处理技术与人工智能技术，通过对音频信号进行下采样，以得到低频信号，由于低频信号相对于音频信号中的高频信号对音频编码的影响更大，因此通过差异化的信号处理，分别提取音频信号的低频特征以及高频特征，使得高频特征的特征维度低于低频特征的特征维度，对减小了特征维度的低频特征以及高频特征分别进行编码，从而在保证音频质量的情况下，提高音频编码效率。

10 客户端 410 可以将音频信号的码流（即高频码流以及低频码流）通过网络 300 发送至服务器 200，以使服务器 200 将码流（高频码流以及低频码流）发送至接收方（例如网络会议的参会对象、观众、语音通话的接收者等）关联的终端设备 500。

在终端设备 500 上运行的客户端 510（例如即时通信客户端、网络会议客户端、直播客户端、浏览器等）在接收到服务器 200 发送的码流（高频码流以及低频码流）后，
20 可以对码流进行音频解码处理，以得到音频信号（即合成音频信号），从而实现音频通信。

例如，客户端 510 调用本申请实施例提供的音频解码方法对接收到的码流（高频码流以及低频码流）进行解码，即对音频信号的低频码流进行解码处理，得到低频码流对应的低频特征，并对音频信号的高频码流进行解码处理，得到高频码流对应的高频特征；
25 对低频特征进行低频特征重建处理，得到低频特征对应的低频信号；对低频信号进行上采样处理，得到低频信号的上采样信号；对高频特征以及上采样信号进行信号重建处理，得到低频码流和高频码流对应的合成音频信号。

在一些实施例中，本申请实施例可以借助云技术（Cloud Technology）实现，云技术是指在广域网或局域网内将硬件、软件、网络等系列资源统一起来，实现数据的计算、
30 存储、处理和共享的一种托管技术。

云技术是基于云计算商业模式应用的网络技术、信息技术、整合技术、管理平台技术、以及应用技术等的总称，可以组成资源池，按需所用，灵活便利。云计算技术将变成重要支撑。上述服务器 200 之间的服务交互功能可以通过云技术实现。

35 示例的，图 2 中示出的服务器 200 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、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。图 2 中示出的终端设备 400 和终端设备 500 可以是笔记本电脑、平板电脑、台式计算机、机顶盒、移动设备（例如，移动电话，便携式音乐播放器，个人数字助理，专用消息设备，便携式游戏设备，车载设备）、
40 智能手机、智能音箱、智能手表、智能电视、车载终端等各种类型的用户终端等，但并不局限于此。终端设备（例如终端设备 400 和终端设备 500）以及服务器 200 可以通过有线或无线通信方式进行直接或间接地连接，本申请实施例中不做限制。

在一些实施例中，终端设备或服务器 200 还可以通过运行计算机程序来实现本申请实施例提供的音频编码方法或音频解码方法。举例来说，计算机程序可以是操作系统

中的原生程序或软件模块；可以是本地（Native）应用程序（APP，Application），即需要在操作系统中安装才能运行的程序，如直播 APP、网络会议 APP、或者即时通信 APP 等；还可以是能够嵌入至任意 APP 中的小程序，即只需要下载到浏览器环境中就可以运行的程序。总而言之，上述计算机程序可以是任意形式的应用程序、模块或插件。

5 在一些实施例中，多个服务器可组成为一区块链，而服务器 200 为区块链上的节点，区块链中的每个节点之间可以存在信息连接，节点之间可以通过上述信息连接进行信息传输。其中，本申请实施例提供的音频编码方法或音频解码方法所相关的数据（例如音频编码的逻辑、音频解码的逻辑、高频码流以及低频码流）可保存于区块链上。

10 参见图 3A，图 3A 是本申请实施例提供的电子设备 500 的结构示意图，以电子设备 500 是终端设备为例说明，图 3A 所示的电子设备 500 包括：至少一个处理器 520、存储器 550、至少一个网络接口 530 和用户接口 540。电子设备 500 中的各个组件通过总线系统 550 耦合在一起。可理解，总线系统 550 用于实现这些组件之间的连接通信。总线系统 550 除包括数据总线之外，还包括电源总线、控制总线和状态信号总线。但是为了
15 清楚说明起见，在图 3A 中将各种总线都标为总线系统 550。

处理器 520 可以是一种集成电路芯片，具有信号的处理能力，例如通用处理器、数字信号处理器（DSP，Digital Signal Processor），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，其中，通用处理器可以是微处理器或者任何常规的处理器等。

20 用户接口 540 包括使得能够呈现媒体内容的一个或多个输出装置 541，包括一个或多个扬声器和/或一个或多个视觉显示屏。用户接口 540 还包括一个或多个输入装置 542，包括有助于用户输入的用户接口部件，比如键盘、鼠标、麦克风、触屏显示屏、摄像头、其他输入按钮和控件。

存储器 550 可以是可移除的，不可移除的或其组合。示例性的硬件设备包括固态存储器，硬盘驱动器，光盘驱动器等。存储器 550 可选地包括在物理位置上远离处理器 520
25 的一个或多个存储设备。

存储器 550 包括易失性存储器或非易失性存储器，也可包括易失性和非易失性存储器两者。非易失性存储器可以是只读存储器（ROM，Read Only Memory），易失性存储器可以是随机存取存储器（RAM，Random Access Memory）。本申请实施例描述的存储器 550 旨在包括任意适合类型的存储器。
30

在一些实施例中，存储器 550 能够存储数据以支持各种操作，这些数据的示例包括程序、模块和数据结构或者其子集或超集，下面示例性说明。

操作系统 551，包括用于处理各种基本系统服务和执行硬件相关任务的系统程序，例如框架层、核心库层、驱动层等，用于实现各种基础业务以及处理基于硬件的任务；

35 网络通信模块 552，用于经由一个或多个（有线或无线）网络接口 530 到达其他计算设备，示例性的网络接口 530 包括：蓝牙、无线相容性认证（WiFi）、和通用串行总线（USB，Universal Serial Bus）等；

在一些实施例中，本申请实施例提供的音频编码装置或者音频解码装置可以采用软件方式实现，图 3A 示出了存储在存储器 550 中的音频编码装置 555，其可以是程序和插件等形式的软件，包括以下软件模块：下采样块 5551、低频提取模块 5552、高频分析模块 5553、编码模块 5554，其中，下采样块 5551、低频提取模块 5552、高频分析模块 5553、编码模块 5554 用于实现音频编码功能，这些模块是逻辑上的，因此根据所实现的功能可以进行任意的组合或进一步拆分。
40

参见图 3B，图 3B 是本申请实施例提供的电子设备 600 的结构示意图，电子设备
45 600 与电子设备 500 的结构类似，图 3B 所示的电子设备 600 包括：至少一个处理器 520、

存储器 550、至少一个网络接口 530 和用户接口 540。电子设备 500 中的各个组件通过总线系统 550 耦合在一起。

用户接口 540 包括使得能够呈现媒体内容的一个或多个输出装置 541。用户接口 540 还包括一个或多个输入装置 542。

5 在一些实施例中，存储器 550 能够存储数据以支持各种操作，这些数据的示例包括程序、模块和数据结构或者其子集或超集，下面示例性说明。

操作系统 551，包括用于处理各种基本系统服务和执行硬件相关任务的系统程序，例如框架层、核心库层、驱动层等，用于实现各种基础业务以及处理基于硬件的任务；

10 网络通信模块 552，用于经由一个或多个（有线或无线）网络接口 530 到达其他计算设备，示例性的网络接口 530 包括：蓝牙、无线相容性认证（WiFi）、和通用串行总线（USB，Universal Serial Bus）等。

图 3B 示出了存储在存储器 550 中的音频解码装置 556，其可以是程序和插件等形式的软件，包括以下软件模块：解码模块 5561、低频重建模块 5562、上采样模块 5563、信号重建模块 5564，其中，解码模块 5561、低频重建模块 5562、上采样模块 5563、信号重建模块 5564 用于实现音频解码功能，这些模块是逻辑上的，因此根据所实现的功能可以进行任意的组合或进一步拆分。

下面在具体介绍本申请实施例提供的音频编码方法以及音频解码方法之前，先介绍空洞卷积网络以及频带扩展技术。

20 参见图 4A 和图 4B，图 4A 是本申请实施例提供的普通卷积网络的示意图，图 4B 是本申请实施例提供的空洞卷积网络的示意图。相对普通卷积网络，空洞卷积能够增加感受野的同时保持特征图的尺寸不变，还可以避免因为上采样、下采样引起的误差。虽然图 4A 和图 4B 中示出的卷积核大小（Kernel Size）均为 3×3 ；但是，图 4A 所示的普通卷积的感受野 401 只有 3，而图 4B 所示的空洞卷积的感受野 402 达到了 5。也就是说，对于尺寸为 3×3 的卷积核，图 4A 所示的普通卷积的感受野为 3、扩张率（Dilation Rate）（卷积核中的点的间隔数量）为 1；而图 4B 所示的空洞卷积的感受野为 5、扩张率为 2。

卷积核还可以在类似图 4A 或者图 4B 的平面上进行移动，这里是涉及移位率（Stride Rate）（步长）概念。比如，每次卷积核移位 1 格，则对应的移位率为 1。

30 此外，还有卷积通道数的概念，就是用多少个卷积核对应的参数去进行卷积分析。理论上，通道数越多，对信号的分析更为全面，精度越高；但是，通道越高，复杂度也越高。比如，一个 1×320 的张量，可以使用 24 通道卷积运算，输出就是 24×320 的张量。

需要说明的是，可以根据实际应用需要，自行定义空洞卷积核大小（例如：针对语音信号，卷积核的大小可以设置为 1×3 ）、扩张率、移位率和通道数，本申请实施例对此不作具体限定。

35 如图 5 所示的频带扩展（或者频带复制）示意图，先重建宽带信号，然后将宽带信号复制到超宽带信号上，最后基于超宽带包络进行整形，图 5 示出的频域实现方案具体包括：1）在低采样率下，实现一个核心层编码；2）选择低频部分的频谱复制到高频；3）根据提前记录的边界信息（描述高频与低频的能量相关性等），对复制的高频频谱进行增益控制。仅需 1-2kbps 码率，就可以产生采样率翻倍的效果。

40 如前所述，本申请实施例提供的音频编码方法可以由各种类型的电子设备实施，例如终端、服务器，又或者是二者的结合。因此下文中不再重复说明各个步骤的执行主体。参见图 6A，图 6A 是本申请实施例提供的音频编码方法的第一流程示意图，通过音频编码方法实现音频编码功能，下面结合图 6A 示出的步骤进行说明。

在步骤 101 中，对音频信号进行下采样处理，得到音频信号的低频信号。

45 作为获取音频信号的示例，编码端响应于发送方（例如网络会议的发起者、主播、

语音通话的发起者等)触发的音频采集指令,调用编码端的终端设备自带的麦克风进行音频信号的采集,以获取音频信号 $x(n)$ (又称输入信号)。

在获取音频信号后,对音频信号进行下采样处理,以从音频信号中提取出低频信号,由于低频信号相对于高频信号对音频编码的影响更大,以便后续对低频信号以及高频信号进行差异化的信号处理。

其中,音频信号包括低频部分以及高频部分,低频信号为基于音频信号特点,通过滤波器从特定采样率的音频信号中,分离出音频信号中的低频部分,高频信号为从特定采样率的音频信号中,分离出音频信号中的高频部分。例如,音频信号 $x(n)$ 的有效带宽为 0-16kHz,则低频信号的有效带宽是 0-8kHz,高频信号 $x_{HB}(n)$ 的有效带宽可以是 6-16kHz。另外,本申请实施例并不限定音频信号的频带划分,如可以平均或者非平均划分音频信号,以得到均匀的低频信号以及高频信号。

需要说明的是,音频信号为离散的数字信号,音频信号包括采样得到的多个第一样本点,其中,第一样本点为从连续的模拟信号中采样得到的采样值。

在一些实施例中,音频信号包括采样得到的多个第一样本点;步骤 101 可以通过以下方式实现:通过下采样滤波器对音频信号包括的第一采样点进行下采样处理,得到音频信号的低频信号。

其中,在数字信号处理领域中,下采样滤波器用于将音频信号的采样率降低,以减少数据量、降低系统复杂度或适应特定应用需求。其中,下采样滤波器的下采样因子可以为 2 的倍数,例如 2、4、8 等。

这里,通过下采样滤波器执行以下处理:对音频信号包括的第一采样点进行基于数字信号的滤波处理,得到滤波后的音频信号;对滤波后的音频信号进行基于数字信号的下采样处理,得到音频信号的低频信号。其中,基于数字信号的滤波处理可以是低通滤波,也可以是带通滤波。

例如,在数字信号处理领域,首先通过下采样滤波器对音频信号进行滤波,以去除音频信号中的高频成分和混叠干扰,以确保降采样后的音频信号不会丢失必要的信息;然后,通过下采样滤波器在滤波后的音频信号中每隔一定的采样点保留一个采样点,从而降低音频信号的采样率。

以音频信号 $x(n)$ 包括 640 个样本点为例进行说明,下采样滤波器用于从音频信号 $x(n)$ 中提取出低频信号 $x_{LB}(n)$,下采样滤波器采用一种因子为 1/2 的下采样滤波,低频信号的有效带宽是 0-8kHz,每帧样本点数为 320。

在步骤 102 中,对低频信号进行低频特征提取处理,得到音频信号的低频特征。

这里,由于低频信号相对于高频信号对音频编码的影响更大,因此可以通过第一神经网络模型对低频信号进行低频特征提取处理,得到低频特征,以在保证低频特征的完备性的情况下,尽量减小低频特征的特征维度,从而提高音频编码效果。其中,低频特征为表征低频信号的特征,低频特征的特征维度小于低频信号的特征维度。

需要说明的是,本申请实施例并不局限于第一神经网络的模型结构,例如第一神经网络可以是深度神经网络、卷积神经网络等。

参见图 6B,图 6B 是本申请实施例提供的音频编码方法的第二流程示意图,图 6B 示出图 6A 中的步骤 102 可以通过步骤 1021-步骤 1024 实现:在步骤 1021 中,对低频信号进行卷积处理,得到低频信号的卷积特征;在步骤 1022 中,对卷积特征进行池化处理,得到低频信号的池化特征;在步骤 1023 中,对池化特征进行下采样处理,得到低频信号的下采样特征;在步骤 1024 中,对下采样特征进行卷积处理,得到音频信号的低频特征。

在音频编解码领域中,神经网络(NN)的卷积、池化和下采样等操作扮演着重要

的角色，用于处理音频信号并提取音频信号中的特征。在音频编解码中，卷积操作可以用于提取音频信号中的局部特征。通过应用卷积核（一种可学习的滤波器），可以在音频信号的时间维度上进行卷积运算，以捕获信号中的模式和共振。卷积可以提取音频信号中的时间域和频域特征，用于降噪、特征提取、信号分离等任务。池化操作用于减少音频信号的时间维度，从而降低数据的复杂性和计算量。池化操作可以取样输入信号的局部区域，并将该区域的信息进行汇总，如最大值或平均值，从而生成更紧凑的特征表示。在音频信号中，池化操作可以有助于提高网络的鲁棒性和泛化能力，减少过拟合的风险。下采样操作可以用于降低音频信号的采样率，即降低信号的频率，下采样操作有助于减少数据量，节省存储和传输开销，并降低复杂度。下采样可以在频域或时间域上进行，以减少采样率或减少信号的维度。在音频编解码领域中，卷积、池化和下采样等操作可以通过构建适当的神经网络结构来实现对音频信号的特征提取、编码和解码等任务。这些操作有助于提高音频信号处理的效率和质量，扩展了音频编解码技术在音频处理、语音识别和音乐生成等领域的应用范围。

如图 10 所示，基于低频信号 $x_{LB}(n)$ 调用第一神经网络模型，生成更低维度的特征向量 $F_{LB}(n)$ ，即低频特征。首先，步骤 1021 中，通过因果卷积 1001 对输入的低频信号 $x_{LB}(n)$ 进行卷积处理，得到 16×320 的卷积特征。然后，对步骤 1022 中， 24×320 的卷积特征进行因子为 2 的池化处理（即预处理 1002），得到 16×160 的池化特征。接下来，在步骤 1023 中，通过编码块对 24×160 的池化特征进行下采样处理，得到 256×1 的下采样特征。最后，在步骤 1024 中，对 256×1 的下采样特征，再经过因果卷积 1004 进行卷积处理，可以得到一个 56 维的特征向量 $F_{LB}(n)$ 。

其中，步骤 1023 中的下采样处理是通过多个级联的编码层实现的，因此，步骤 1023 可以通过以下方式实现：通过多个级联的编码层中的第一个编码层，对池化特征进行下采样处理；将第一个编码层的下采样结果输出到后续级联的编码层，通过后续级联的编码层继续进行下采样处理和下采样结果输出，直至输出到最后一个编码层；将最后一个编码层输出的下采样结果确定为低频信号的下采样特征。

如图 10 所示，级联 4 个不同下采样因子（Down_factor）的编码块（即编码层）对池化特征进行下采样处理，即先通过下采样因子（Down_factor = 2）的编码块 1003-1 对 16×160 的池化特征进行下采样处理，得到 32×80 的下采样结果，然后通过下采样因子（Down_factor = 4）的编码块 1003-2 对 32×80 的下采样结果进行下采样处理，得到 64×20 的下采样结果，接着通过下采样因子（Down_factor = 4）的编码块 1003-3 对 64×20 的下采样结果进行下采样处理，得到 128×5 的下采样结果，最后通过下采样因子（Down_factor = 5）的编码块 1003-4 对 128×5 的下采样结果进行下采样处理，得到 256×1 的下采样特征。以编码块（Down_factor = 4）为例，可以先执行 1 个或者多个空洞卷积，并基于 Down_factor 进行池化，以实现下采样的作用。

需要说明的是，通过一层编码层的处理后，对下采样特征的理解就加深一步，通过多层编码层的学习后，就能够逐步准确学习到低频信号的下采样特征。通过级联形式的编码层，能够获取精度渐进的低频信号的下采样特征。

在步骤 103 中，对音频信号进行高频分析处理，得到音频信号的高频特征。

其中，高频特征的特征维度低于低频特征的特征维度，例如，低频特征的特征维度为 56，高频特征的特征维度为 8。由于低频信号相对于高频信号对音频编码的影响更大，因此通过差异化的信号处理，分别提取音频信号的低频特征以及高频特征，使得高频特征的特征维度低于低频特征的特征维度。其中，高频分析处理用于对音频信号中的高频信号进行降维，实现数据压缩的功能。

参见图 6C，图 6C 是本申请实施例提供的音频编码方法的第三流程示意图，图 6C

示出图 6A 中的步骤 103 可以通过步骤 1031 实现：在步骤 1031 中，对音频信号进行频带扩展处理，得到音频信号的高频特征。

这里，通过步骤 1031 快速压缩音频信号中的高频信号，提取出高频特征。

参见图 6D，图 6D 是本申请实施例提供的音频编码方法的第四流程示意图，图 6D 示出图 6C 中的步骤 1031 可以通过步骤 10311-步骤 10314 实现：

在步骤 10311 中，对音频信号包括的多个第二样本点进行频域变换处理，得到多个第二样本点分别对应的变换系数。

需要说明的是，步骤 10311 中的频域变换方法（也就是时频变换方法）可以是改进型离散余弦变换（MDCT，Modified Discrete Cosine Transform）、离散余弦变换（DCT，Discrete Cosine Transform）、快速傅立叶变换（FFT，Fast Fourier Transform）等，本申请实施例并不局限于频域变换的方式。

在一些实施例中，步骤 10311 可以通过以下方式实现：获取参考音频信号包括的多个第三样本点，其中，参考音频信号是与音频信号相邻的音频信号；基于参考音频信号包括的多个第三样本点以及音频信号包括的多个第二样本点，对音频信号包括的多个第二样本点进行离散余弦变换处理，得到多个第二样本点分别对应的变换系数。

作为步骤 10311 的示例，对于包括 640 个第二样本点的音频信号 $x(n)$ ，调用改进型离散余弦变换（MDCT，Modified Discrete Cosine Transform）方法，生成 640 个第二样本点的 MDCT 系数（即多个第二样本点分别对应的变换系数）。例如，如果改进型离散余弦变换方法是 50% 交叠，可以将第 $n+1$ 帧音频信号（即参考音频信号）与第 n 帧音频信号（即音频信号）合并（拼接），计算出 1280 个样本点的 MDCT，将有效的前 640 个样本点的 MDCT 系数作为多个第二样本点分别对应的变换系数。

在步骤 10312 中，将多个第二样本点分别对应的变换系数中的高频变换系数划分为多个子带。

承接步骤 10311 的示例，步骤 10312 具体为：640 个第二样本点的 MDCT 系数（即多个第二样本点分别对应的变换系数）的后 320 个第二样本点的 MDCT 系数，代表高频信号，也就是高频变换系数。将高频变换系数分成 N 个子带，这里的子带就是将相邻的多个 MDCT 系数组成一组，320 个第二样本点的 MDCT 系数可以分成 8 个子带。例如，可以均匀分配 320 个第二样本点的 MDCT 系数，即每个子带包括的样本点的数量一致，那么每个子带包括 40 个第二样本点的 MDCT 系数。当然，本申请实施例不限制对 320 个第二样本点进行非均匀划分，比如偏低频的子带包括的 MDCT 系数更少（频率分辨率更高），偏高频的子带包括的 MDCT 系数更多（频率分辨率更低）。

根据奈奎斯特（Nyquist）采样定律（要从抽样信号中无失真地恢复原信号，抽样频率应大于 2 倍原信号最高频率，抽样频率小于 2 倍频谱最高频率时，信号的频谱有混叠，抽样频率大于 2 倍频谱最高频率时，信号的频谱无混叠），上述 320 个第二样本点（简称点）的 MDCT 系数表示了 8-16kHz 的频谱。但超宽带语音通信，不一定要求频谱到 16kHz，比如，如果设置频谱到 14kHz，只需要考虑前 240 个第二样本点的 MDCT 系数，对应地，子带数可以控制为 6。

在步骤 10313 中，对每个子带包括的变换系数进行平均处理，得到每个子带对应的平均能量，并将平均能量作为每个子带对应的子带谱包络。

需要说明的是，步骤 10313 中的平均处理可以是算术平均、几何平均，本申请实施例并不局限于平均处理的方式。

以几何平均为例，步骤 10313 可以通过以下方式实现：确定每个子带包括的第二样本点对应的变换系数的平方和；将平方和与子带包括的第二样本点的数量的比值，确定为每个子带对应的平均能量。

承接步骤 10312 的示例, 步骤 10313 具体为: 对于每一个子带, 计算当前子带中所有 MDCT 系数的平均能量作为子带谱包络 (子带谱包络是经过频谱各主峰值点的平滑曲线), 比如, 当前子带中包括的 MDCT 系数为 $x(n), n=1, 2, \dots, 40$, 则当前子带对应的平均能量 $Y=((x(1)^2+x(2)^2+\dots+x(40)^2)/40)$ 。

5 在步骤 10314 中, 将多个子带分别对应的子带谱包络确定为音频信号的高频特征。

承接步骤 10313 的示例, 步骤 10314 具体为: 对于 320 个点的 MDCT 系数分成 8 个子带的情况, 可以获得 8 个子带谱包络, 这 8 个子带谱包络, 就是音频信号的高频特征 $F_{HB}(n)$, 也就是高频的特征向量。

10 在步骤 104 中, 对低频特征进行编码处理, 得到音频信号的低频码流, 并对高频特征进行编码处理, 得到音频信号的高频码流。

这里, 在数字信号处理领域中, 对低频特征进行基于数字信号的编码处理, 得到音频信号的低频码流, 并对高频特征进行基于数字信号的编码处理, 得到音频信号的高频码流。

15 在一些实施例中, 步骤 104 可以通过以下方式实现: 对低频特征进行量化处理, 得到低频特征的索引值; 对低频特征的索引值进行熵编码处理, 得到音频信号的低频码流; 对高频特征进行量化处理, 得到高频特征的索引值; 对高频特征的索引值进行熵编码处理, 得到音频信号的高频码流。

20 作为步骤 104 的示例, 对于低频特征 $F_{LB}(n)$ 和高频特征 $F_{HB}(n)$, 均可以进行标量量化 (各分量单独量化) 和熵编码的方法。另外, 本申请实施例也不限制矢量量化 (相邻多个分量组合成一个矢量进行联合量化) 和熵编码的技术组合, 将编码得到的高频码流以及低频码流合并传输至解码端, 通过解码端对高频码流以及低频码流进行解码。

25 综上, 本申请实施例提供的音频编码方法, 通过对音频信号进行下采样, 以得到低频信号, 由于低频信号相对于音频信号中的高频信号对音频编码的影响更大, 因此通过差异化的信号处理, 分别提取音频信号的低频特征以及高频特征, 使得高频特征的特征维度低于低频特征的特征维度, 对减小了特征维度的低频特征以及高频特征分别进行编码, 从而在保证音频质量的情况下, 提高音频编码效率。

30 如前所述, 本申请实施例提供的音频解码方法可以由各种类型的电子设备实施。参见图 7A, 图 7A 是本申请实施例提供的音频解码方法的第一流程示意图, 通过音频解码方法实现音频解码功能, 下面结合图 7A 示出的步骤进行说明。

在步骤 201 中, 对音频信号的低频码流进行解码处理, 得到低频码流对应的低频特征, 并对音频信号的高频码流进行解码处理, 得到高频码流对应的高频特征。

其中, 低频码流是通过音频信号经过下采样后得到的低频信号进行编码得到的, 高频特征的特征维度低于低频特征的特征维度。

35 例如, 通过如图 6A 所示的音频编码方法编码得到高频码流以及低频码流后, 将编码得到的高频码流以及低频码流合并传输至解码端, 解码端接收到高频码流以及低频码流后, 对低频码流进行解码处理, 得到低频码流对应的低频特征, 并对高频码流进行解码处理, 得到高频码流对应的高频特征。

40 其中, 步骤 201 中的解码是步骤 104 中的编码的逆过程。在数字信号处理领域中, 对音频信号的低频码流进行基于数字信号的解码处理, 得到低频码流对应的低频特征, 并对音频信号的高频码流进行基于数字信号的解码处理, 得到高频码流对应的高频特征。

45 在一些实施例中, 步骤 201 可以通过以下方式实现: 对低频码流进行熵解码处理, 得到低频码流对应的索引值; 对低频码流对应的索引值进行逆量化处理, 得到低频码流对应的低频特征; 对高频码流进行熵解码处理, 得到高频码流对应的索引值; 对高频码流对应的索引值进行逆量化处理, 得到高频码流对应的高频特征。其中, 逆量化处理是

通过查询量化表实现的, 量化表为编码过程中量化所产生的映射表。

作为示例, 对于收到码流 (高频码流以及低频码流), 先进行熵解码, 得到码流对应的索引值, 并通过码流对应的索引值查询量化表, 将查询到的特征向量作为码流对应的特征向量, 也即获得低频的特征向量的估计值 $F'_{LB}(n)$ (即低频码流对应的低频特征) 以及高频的特征向量的估计值 $F'_{HB}(n)$ (即高频码流对应的高频特征)。需要说明的是, 解码端对接收到的码流进行解码的过程是编码端进行编码的过程的逆过程, 因此, 解码过程中产生的值是相对于编码过程中的值的估计值, 例如解码过程中产生的高频特征相对于编码过程中的高频特征是一种估计值。

在步骤 202 中, 对低频特征进行低频特征重建处理, 得到低频特征对应的低频信号。例如, 步骤 202 中的低频特征重建是步骤 102 中的低频特征提取的逆过程, 通过第二神经网络模型对低频特征进行低频特征重建处理, 得到低频特征对应的低频信号 (对应编码过程中所得到的低频信号的一种估计值)。

需要说明的是, 本申请实施例并不局限于第二神经网络的模型结构, 例如第二神经网络可以是深度神经网络、卷积神经网络等。

参见图 7B, 图 7B 是本申请实施例提供的音频解码方法的第二流程示意图, 图 7B 示出图 7A 中的步骤 202 可以通过步骤 2021-步骤 2024 实现: 在步骤 2021 中, 对低频特征进行卷积处理, 得到低频特征的卷积特征; 在步骤 2022 中, 对卷积特征进行上采样处理, 得到低频特征的上采样特征; 在步骤 2023 中, 对上采样特征进行池化处理, 得到低频特征的池化特征; 在步骤 2024 中, 对池化特征进行卷积处理, 得到低频特征对应的低频信号。

在音频编解码领域中, 上采样操作用于增加特征图的分辨率, 以更准确地重建音频信号。上采样涉及插值或其他形式的上采样技术, 以生成更高精度的特征图, 这有助于在解码过程中更好地恢复音频信号的原始细节和特性。通过在音频解码中使用卷积、池化和上采样等神经网络技术, 可以有效地提取有用特征、减少计算复杂度, 并更准确地重建音频信号的原始内容。这些技术对于提高音频解码的性能和效率具有重要意义, 有助于推动音频编解码技术的发展和应。

如图 11 所示, 基于低频特征 $F'_{LB}(n)$, 调用如图 11 所示的神经网络模型, 生成低频信号 $x'_{LB}(n)$ 。其中, 如图 11 所示的神经网络模型与如图 10 所示的神经网络模型类似, 比如因果卷积、后处理结构与预处理类似。解码块的结构与编码端的编码块是对称地: 编码端的编码块是先做空洞卷积再池化完成下采样, 解码端的解码块是先进行池化完成上采样, 再做空洞卷积。

首先, 步骤 2021 中, 通过因果卷积 1101 对输入的低频特征 $F'_{LB}(n)$ 进行卷积处理, 得到 256×1 的卷积特征。然后, 步骤 2022 中, 对 256×1 的卷积特征进行上采样处理, 得到 16×160 的上采样特征。接下来, 步骤 2023 中, 对 16×160 的上采样特征进行池化处理 (即后处理 1103), 得到 16×320 的池化特征。最后, 步骤 2024 中, 对 16×320 的池化特征, 再经过因果卷积 1104 进行卷积处理, 可以得到一个 320 维的低频信号 $x'_{LB}(n)$ 。

其中, 步骤 2022 中的上采样处理是通过多个级联的解码层实现的, 因此, 步骤 2022 可以通过以下方式实现: 通过多个级联的解码层中的第一个解码层, 对卷积特征进行上采样处理; 将第一个解码层的上采样结果输出到后续级联的解码层, 通过后续级联的解码层继续进行上采样处理和上采样结果输出, 直至输出到最后一个解码层; 将最后一个解码层输出的上采样结果确定为低频特征的上采样特征。

作为步骤 2022 的示例, 如图 11 所示, 级联 4 个不同上采样因子 (Up_factor) 的解码块 (即解码层) 对卷积特征进行上采样处理, 即先通过上采样因子 (Up_factor = 5)

的解码块 1102-1 对 256×1 的卷积特征进行上采样处理, 得到 128×5 的上采样结果, 然后通过上采样因子 ($Up_factor = 4$) 的解码块 1102-2 对 128×5 的上采样结果进行上采样处理, 得到 64×20 的上采样结果, 接着通过上采样因子 ($Up_factor = 4$) 的解码块 1102-3 对 64×20 的上采样结果进行上采样处理, 得到 32×80 的上采样结果, 最后通过上采样因子 ($Up_factor = 2$) 的解码块 1102-4 对 32×80 的上采样结果进行上采样处理, 得到 16×160 的上采样特征。以解码块 ($Up_factor = 4$) 为例, 可以先并基于 Up_factor 进行池化, 然后执行 1 个或者多个空洞卷积, 以实现上采样的作用。

需要说明的是, 通过一层解码层的处理后, 对上采样特征的理解就加深一步, 通过多层解码层的学习后, 就能够逐步准确学习到低频特征的上采样特征。通过级联形式的解码层, 能够获取精度渐进的低频特征的上采样特征。

在步骤 203 中, 对低频信号进行上采样处理, 得到低频信号的上采样信号。

例如, 以编码端的采样率为 $1/2$ 为例, 在解码端所得到的低频信号 $x'_{LB}(n)$ 的采样率仅为编码端音频信号 $x(n)$ 的采样率的 $1/2$ 。因此, 需要对低频信号 $x'_{LB}(n)$ 进行上采样, 将采样率扩大 1 倍, 生成上采样信号 $x_{up}(n)$, 从而, 保证上采样信号的采样率与音频信号的采样率相同。

在一些实施例中, 步骤 203 可以通过以下方式实现: 通过上采样滤波器对低频信号包括的第四采样点进行上采样处理, 得到低频信号的上采样信号。

其中, 在数字信号处理领域中, 上采样滤波器用于将输入上采样滤波器中的信号的采样率提高, 以便更好地捕获信号中的细节信息或扩展频谱范围。其中, 上采样滤波器的上采样因子可以为 2 的倍数, 例如 2、4、8 等。

这里, 通过上采样滤波器执行以下处理: 对低频信号包括的第四采样点进行基于数字信号的上采样处理, 得到上采样后的低频信号; 对上采样后的低频信号进行基于数字信号的滤波处理, 得到低频信号的上采样信号。其中, 基于数字信号的滤波处理可以是低通滤波, 也可以是带通滤波。

例如, 在数字信号处理领域, 首先通过上采样滤波器对低频信号包括的第四采样点进行基于数字信号的上采样处理, 其中, 上采样是指在低频信号中插入零值来增加采样率; 对上采样后的低频信号进行基于数字信号的滤波处理, 以去除插值引入的混叠干扰。

以低频信号 $x'_{LB}(n)$ 包括 320 个样本点为例进行说明, 上采样滤波器用于将低频信号 $x'_{LB}(n)$ 的采样率扩大一倍, 上采样滤波器采用一种因子为 2 的上采样滤波, 将低频信号 $x'_{LB}(n)$ 的采样率扩大 1 倍, 生成上采样信号 $x_{up}(n)$, 其中, 每帧 $x_{up}(n)$ 包含 640 个样本点 (简称点)。

在步骤 204 中, 对高频特征以及上采样信号进行信号重建处理, 得到低频码流和高频码流对应的合成音频信号。

对于信号重建, 基于高频特征以及上采样信号频带扩展出高频信号, 并基于上采样信号和重建出的高频信号合并出合并音频信号 $x'(n)$ 。

参见图 7C, 图 7C 是本申请实施例提供的音频解码方法的第三流程示意图, 图 7C 示出图 7A 中的步骤 204 可以通过步骤 2041-步骤 2043 实现:

在步骤 2041 中, 对上采样信号包括的多个第五样本点进行频域变换处理, 得到多个第五样本点分别对应的变换系数。

需要说明的是, 步骤 2041 中的频域变换方法 (也就是时频变换方法) 可以是改进型离散余弦变换 (MDCT, Modified Discrete Cosine Transform)、离散余弦变换 (DCT, Discrete Cosine Transform)、快速傅立叶变换 (FFT, Fast Fourier Transform) 等, 本申请实施例并不局限于频域变换的方式。

在一些实施例中, 步骤 2041 可以通过以下方式实现: 获取参考上采样信号包括的

多个样本点, 其中, 参考上采样信号是与上采样信号相邻的上采样信号; 基于参考上采样信号包括的多个样本点以及上采样信号包括的多个第五样本点, 对上采样信号包括的多个第五样本点进行离散余弦变换处理, 得到多个第五样本点分别对应的变换系数。

5 作为步骤 2041 的示例, 对于包括 640 个样本点的上采样信号 $x_{up}(n)$, 调用改进型离散余弦变换 (MDCT, Modified Discrete Cosine Transform) 方法, 生成 640 个样本点的 MDCT 系数 (即多个第五样本点分别对应的变换系数)。具体地, 如果改进型离散余弦变换方法是 50% 交叠, 可以将第 $n+1$ 帧上采样信号 (即参考上采样信号) 与第 n 帧上采样信号 (即上采样信号) 合并 (拼接), 计算出 1280 个样本点 (包括参考上采样信号包括的多个样本点以及上采样信号包括的多个第五样本点) 的 MDCT。由于计算出的 10 1280 个样本点的 MDCT 是一种上采样信号形式, 高频内容是缺失的, 因此, 保留计算出的 1280 个样本点的 MDCT 的低频部分内容即可, 也就是将有效的前 640 个样本点的 MDCT 系数作为多个第五样本点分别对应的变换系数 $X_{up}(n)$ 。

在步骤 2042 中, 对多个第五样本点分别对应的变换系数以及高频特征进行频带扩展处理, 得到高频信号的高频变换系数。

15 这里, 通过频带扩展技术, 基于多个第五样本点分别对应的变换系数以及高频特征重建出高频信号的高频变换系数。

参见图 7D, 图 7D 是本申请实施例提供的音频解码方法的第四流程示意图, 图 7D 示出图 7C 中的步骤 2042 可以通过以下步骤 20421-步骤 20422 实现:

20 在步骤 20421 中, 对多个第五样本点分别对应的变换系数中的前半部分的至少部分变换系数进行频谱复制处理, 得到参考高频信号的参考高频变换系数。

在一些实施例中, 步骤 20421 可以通过以下方式实现: 对多个第五样本点分别对应的变换系数中的前半部分的后半部分变换系数进行频谱复制处理, 得到参考高频信号的参考高频变换系数。

25 作为步骤 20421 的示例, 多个第五样本点分别对应的变换系数为 $X_{up}(n)$ (包括 640 个点的 MDCT 系数), 将 $X_{up}(n)$ 中的低频 MDCT 系数 (即前 320 个点的 MDCT 系数) 进行复制, 生成高频部分的 MDCT 系数的参考值, 即参考高频信号的参考高频变换系数。参考语音信号的基本特征, 低频部分谐波更多, 高频部分谐波更少。因此, 为了避免简单复制造成生成的高频 MDCT 谱包含过多谐波, 可以将 $X_{up}(n)$ 中的低频 MDCT 系数的后 160 个点 (即多个第五样本点分别对应的变换系数中的前半部分的后半部分变换系数) 作为母版, 在频谱的高频部分复制 2 次, 生成高频部分信号的 320 个点的 MDCT 系数的参考值 (即参考高频信号的参考高频变换系数)。因此, 频谱复制后的 $X_{up}(n)$ 中的 30 的后 320 个点是而非零系数。

当然, 本申请实施例不局限于将 $X_{up}(n)$ 中的低频 MDCT 系数的后 160 个点作为母版, 还可以将 $X_{up}(n)$ 中的低频 MDCT 系数作为母版, 频谱复制 1 次, 生成高频部分信号的 35 320 个点的 MDCT 系数的参考值; 还可以将 $X_{up}(n)$ 中的低频 MDCT 系数的最后的 80 个点作为母版, 频谱复制 4 次, 生成高频部分信号的 320 个点的 MDCT 系数的参考值。

在步骤 20422 中, 基于高频特征对应的子带谱包络, 对参考高频信号的参考高频变换系数进行增益处理, 得到高频信号的高频变换系数。

40 在一些实施例中, 步骤 20422 可以通过以下方式实现: 基于高频特征对应的子带谱包络, 将参考高频信号的参考高频变换系数划分为多个子带; 针对多个子带中的任意子带执行以下处理: 确定子带谱包络中与子带对应的高频子带的第一平均能量, 并确定子带的第二平均能量; 基于第一平均能量与第二平均能量的比值, 确定增益因子; 将增益因子与子带包括的每个参考高频变换系数相乘, 得到子带对应的高频变换系数; 将多个

子带分别对应的高频变换系数确定为高频信号的高频变换系数。

承接步骤 20421 的示例，步骤 20422 具体为：高频特征对应的子带谱包络（例如 8 个子带谱包络）为通过解码后得到的高频特征对应的估计值 $F'_{HB}(n)$ ，这 8 个子带谱包络对应 8 个高频子带，并将生成的高频部分信号的 320 个点的 MDCT 系数的参考值（即参考高频信号的参考高频变换系数）分为 8 个子带，基于一个高频子带与对应的子带，对高频部分信号的 320 个点的 MDCT 系数的参考值进行增益（频域上就是做乘法），例如根据高频子带的平均能量与对应的子带的平均能量计算增益因子，将该对应的子带中的每个点对应的 MDCT 系数乘以增益因子，确保解码生成的高频的 MDCT 系数的能量，与编码端原始的系数能量接近。

例如，假定子带（即由生成的高频部分信号的 320 个点的 MDCT 系数的参考值所划分得到的子带）的平均能量是 Y_L ，当前高频子带（即基于码流解码出来的子带谱包络对应的子带）的平均能量是 Y_H ，则计算一个增益因子 $a = \sqrt{Y_H/Y_L}$ 。有了增益因子 a 后，直接将子带中的每个点的 MDCT 系数均乘以 a 。经过增益控制后的 MDCT 系数（即高频信号的高频变换系数）的平均能量，就跟编码端原始的平均能量非常接近了，以尽可能还原出原始的高频信号。

在步骤 2043 中，基于高频变换系数以及多个第五样本点中的部分第五样本点分别对应的变换系数，确定低频码流和高频码流对应的合成音频信号。

在一些实施例中，步骤 2043 可以通过以下方式实现：对高频变换系数以及多个第五样本点中的前半部分第五样本点分别对应的变换系数进行合并处理，得到完整变换系数；对完整变换系数进行频谱变换的逆变换处理，得到低频码流和高频码流对应的合成音频信号。

承接步骤 20422 的示例，步骤 2043 具体为：对 $X_{up}(n)$ 中的低频 MDCT 系数（即前 320 个点的 MDCT 系数，也就是多个第五样本点中的部分第五样本点分别对应的变换系数）以及经过增益控制后的 MDCT 系数（即 $X_{up}(n)$ 中增益后的后 320 个点的 MDCT 系数的参考值，也就是高频变换系数）进行合并，得到完整变换系数（即 640 个点的 MDCT 系数，包括低频的 MDCT 系数以及高频的 MDCT 系数），对完整变换系数进行 MDCT 反变换（即频谱变换的逆变换处理），生成 1280 个点的 MDCT 的估计值，通过交叠，取有效的前 640 个点的 MDCT 的估计值作为合成音频信号。

综上，本申请实施例提供的音频解码方法，对接收的码流（即低频码流以及高频码流）进行解码，获得低频特征以及高频特征，对低频的特征向量调用对应编码端的逆过程，完成低频部分的重建。然后，将重建的低频部分（即低频信号），通过上采样滤波，恢复到与原始输入的音频信号一样的采样率；通过结合上采样滤波后的低频部分、高频特征，重建出高频信号的高频变换系数，将重建后的低频部分和高频变换系数合并，保证重建出的高频信号的能量与编码端的高频信号的能量接近。这样，低频至高频的频域系数获得完整重建。

下面，将说明本申请实施例在一个实际的应用场景中的示例性应用。

本申请实施例可以应用于各种音频场景，例如语音通话、即时通信等。下面以语音通话为例进行说明：

相关技术中，语音编码原理大致如下：语音编码可以直接对语音波形样本，逐样本地进行编码；或者，基于人的发声原理，提取相关低维度特征，编码端对特征进行编码，解码端基于这些参数重建语音信号。

上述编码原理均来自语音信号建模，即基于信号处理的压缩方法。为了相对于基于信号处理的压缩方法，在保证语音质量的情况下，提高编码效率。本申请实施例提供一种低速率神经网络编解码方法（即音频处理方法），基于语音信号特点，将特定采样率

的语音信号,通过下采样滤波,分离出最重要的低频部分(即低频信号)。对于重要的低频部分(即低频信号),基于神经网络(NN, Neural Network)技术进行处理后,会得到比输入的低频信号更低维度的特征向量。对于语音信号中相对不重要的部分(高频部分,即高频信号),使用更少的比特进行编码。对于编码端,分别对语音信号中的低频部分和高频部分进行处理后,得到比输入的语音信号更低维度的特征向量(低频的特征向量以及高频的特征向量),对特征向量进行压缩编码。对于解码端,对接收的码流(即低频码流以及高频码流)进行解码,获得低频的特征向量以及高频的特征向量,对低频的特征向量调用对应编码端的逆过程,完成低频部分的重建。然后,将重建的低频部分,通过上采样滤波,恢复到与原始输入的语音信号一样的采样率;通过结合上采样滤波后的低频部分、高频的特征向量,重建出高频部分,将重建后的低频部分和高频部分合并,以完成解码。

本申请实施例可应用于如图8所示的语音通信链路,以基于网际互连协议的语音传输(VoIP, Voice over Internet Protocol)会议系统为例,将本申请实施例涉及的语音编解码技术部署在编码和解码部分,以解决语音压缩的基本功能。编码器部署在上行客户端801,解码器部署在下行客户端802,通过上行客户端采集语音,并进行前处理增强、编码等处理,将编码得到的码流通过网络传输至下行客户端802,通过下行客户端802进行解码、增强等处理,以下行客户端802回放解码出的语音。

考虑前向兼容(即新的编码器与已有的编码器兼容),需要在系统的后台(即服务器)部署转码器,以解决新的编码器与已有的编码器互联互通问题。例如,如果发送端(上行客户端)是新的NN编码器,接收端(下行客户端)是公用电话交换网(PSTN, Public Switched Telephone Network)(G.722)。在后台,需要执行NN解码器生成语音信号,然后调用G.722编码器生成特定码流,以实现转码功能,才能让接收端基于特定码流进行正确解码。

下面结合图9说明本申请实施例提供的低速率神经网络编解码方法(由音频编码方法以及音频解码方法实现):

对于编码端执行如下处理:

对第 n 帧的输入语音信号 $x(n)$,使用下采样滤波器,提取出语音信号 $x(n)$ 的低频部分信号 $x_{LB}(n)$,简称低频信号。

对于低频部分信号 $x_{LB}(n)$,调用第一NN,获得低维度的特征向量 $F_{LB}(n)$ (简称低频特征),特征向量 $F_{LB}(n)$ 的维度小于低频子带信号的维度,以减少数据量。例如,对于每帧 $x_{LB}(n)$,调用空洞卷积网络(Dilated CNN),生成更低维的特征向量 $F_{LB}(n)$ 。本申请实施例不限制其它的NN结构,例如自编码器(Autoencoder),全连接(FC, Full-Connection)网络,长短期记忆(LSTM, Long Short-Term Memory)网络、卷积神经网络(CNN, Convolutional Neural Network)+LSTM等等。

同时,对语音信号 $x(n)$ 进行高频分析,提取出高频特征向量 $F_{HB}(n)$ (简称高频特征)。考虑到高频对质量重要性没有低频那么重要,本申请实施例基于语音信号分析的频带扩展技术,可以只用1-2kbps实现高频信号的生成。其中,基于频带扩展的高频分析的实现方式包括:将语音信号 $x(n)$ 进行时频变换,获得频域系数;将频域系数的相邻系数合并成多个子带;计算各子带的平均能量作为子带的谱包络,这里,可以只对高频相关的频域系数进行操作;将提取的与高频相关的谱包络提取作为 $F_{HB}(n)$ 。

对子带对应的特征向量(即 $F_{LB}(n)$ 以及 $F_{HB}(n)$)进行矢量量化或者标量量化,并将量化后的索引值进行熵编码,并编码后得到的码流传输到解码端。

对于解码端执行如下处理:

对解码端接收到码流进行解码,分别获得低频的特征向量(简称低频特征)的估计值 $F'_{LB}(n)$ 和高频的特征向量(简称高频特征)的估计值 $F'_{HB}(n)$ 。

对于低频部分,基于低频的特征向量的估计值 $F'_{LB}(n)$ 调用第二 NN,生成低频信号的估计值 $x'_{LB}(n)$ 。

5 将生成的低频信号的估计值 $x'_{LB}(n)$,进行上采样滤波,以生成 $x_{up}(n)$,其中, $x_{up}(n)$ 的采样率与编码端输入的语音信号 $x(n)$ 采样率一致。

对于信号重建,信号重建主要包含基于频带扩展的高频部分的重建、以及低频部分信号和重建出的高频信号合并出最终输出信号 $x'(n)$ 。第一,对 $x_{up}(n)$ 进行时频变换,获得频域系数,如前所述,此频域系数的分辨率与输入的语音信号 $x(n)$ 的分辨率一致。第二,由于频域系数缺失高频成份,因此基于频带扩展技术,生成高频成份(即高频部分的频域系数),并将高频成份与低频部分的频域系数合并,生成完整的频域系数。第三,对完整的频域系数进行时频逆变换,获得最终输出信号 $x'(n)$ 。其中,频带扩展技术包括:将低频部分的频域系数复制到高频部分;基于获得的高频特征向量 $F'_{HB}(n)$,对高频部分的频域系数,按子带进行能量调整,保证重建出的高频部分的能量与编码端的高频能量接近。这样,低频至高频的频域系数获得完整重建。

15 下面对本申请实施例提供的低速率神经网络编码方法进行具体说明。

在一些实施例中,以采样率 $F_s=32000\text{Hz}$ 的语音信号为例(需要说明的是,本申请实施例提供的方法也适用于其他采样率的场景,包括但不限于:8000Hz、16000Hz、48000Hz)。同时,假设帧长设置为20ms,因此,对于 $F_s=32000\text{Hz}$,相当于每帧包含640个样本点。

下面参考图9所示的流程图,将编码端和解码端进行详细描述。

关于编码端的流程如下:

针对采样率 $F_s=32000\text{Hz}$ 的语音信号,对于第 n 帧的输入信号包括640个样本点(简称点),记为输入信号 $x(n)$ 。

25 步骤11、下采样滤波。

本申请实施例中的下采样滤波是从原始的输入信号 $x(n)$ 中,分离出低频信号 $x_{LB}(n)$ 。其中,本申请实施例采用的是一种因子为1/2的下采样滤波,则低频信号的有效带宽是0-8kHz,每帧样本点数为320。

步骤12、将低频信号 $x_{LB}(n)$ 输入第一 NN 进行数据压缩。

30 基于低频信号 $x_{LB}(n)$,调用第一 NN,生成更低维度的特征向量 $F_{LB}(n)$ 。需要说明的是, $x_{LB}(n)$ 的维度为320, $F_{LB}(n)$ 的维度为56,从数据量看,第一 NN 起到了“降维”的作用,实现数据压缩的功能。

参见图10所示的第一 NN 的网络结构图,下面具体说明第一 NN 进行数据压缩的流程:

35 首先,调用一个16通道的因果卷积1001,可以将输入的张量(即低频信号 $x_{LB}(n)$),扩展为 16×320 的张量。

然后,对 16×320 的张量进行预处理1002。例如,对 16×320 的张量做因子为2的池化(Pooling)操作、且激活函数可以为ReLU,以生成 16×160 的张量。

40 接下来,级联4个不同下采样因子(Down_factor)的编码块(即编码块1003-1、编码块1003-2、编码块1003-3、编码块1003-4)。以编码块(Down_factor=2)为例,可以先执行1个或者多个空洞卷积,每个卷积核大小均固定为 1×3 、移位率(Stride Rate)为1。此外,1个或者多个空洞卷积的扩张率(Dilation Rate)可根据实际需求设置,比如3,当然,本申请实施例也不限制不同空洞卷积设置的扩张率。然后,4个编码块的Down_factor分别设置为2、4、4、5,等效于设置了不同大小的池化因子,起到下采样

的作用。最后，4个编码块通道数分别设置为32、64、128、256。因此，经过4个编码块， 16×160 的张量依次转换成 32×80 、 64×20 、 128×5 和 256×1 的张量。需要说明的是，本申请实施例并不限制编码块的其它实现方式，例如空洞卷积的数量（包括扩展率）、每一个编码块中下采样因子的值和每个编码块的输出通道数等，只要确保上述配置，匹配第一NN的输入为 1×320 、输出为 56×1 的维度即可。

最后，对 256×1 的张量，再经过类似预处理的因果卷积1004，可以输出一个56维的特征向量 $F_{LB}(n)$ 。

步骤13、对输入信号 $x(n)$ 进行高频分析。

高频分析的目的是对输入信号 $x(n)$ 进行分析，提取出与高频信号相关的关键信息，生成更低维度的特征向量 $F_{HB}(n)$ 。

本申请实施例提出基于频带扩展（从频带受限的窄带语音信号中恢复宽带语音信号）对输入信号进行高频分析。下面具体介绍本申请实施例中频带扩展的应用：

对于包括640点的输入信号 $x(n)$ ，调用改进型离散余弦变换（MDCT, Modified Discrete Cosine Transform），生成640个点的MDCT系数。例如，如果改进型离散余弦变换方法是50%交叠，可以将第 $n+1$ 帧输入信号数据与第 n 帧输入信号合并（拼接），计算1280个点的MDCT，获得640个点的MDCT系数。

对于高频分析，本申请实施例将640个点的MDCT系数的后320个点的MDCT系数，代表高频信号。将320个点的MDCT系数（即高频信号）分成 N 个子带，这里的子带就是将相邻的多个MDCT系数组成一组，320个点的MDCT系数可以分成8个子带。例如，可以均匀分配320个点，即每个子带包括的点的数量一致。当然，本申请实施例不限制对320个点进行非均匀划分，比如偏低频的子带包括的MDCT系数更少（频率分辨率更高），偏高频的子带包括的MDCT系数更多（频率分辨率更低）。

根据奈奎斯特（Nyquist）采样定律（要从抽样信号中无失真地恢复原信号，抽样频率应大于2倍原信号最高频率，抽样频率小于2倍频谱最高频率时，信号的频谱有混叠，抽样频率大于2倍频谱最高频率时，信号的频谱无混叠），上述320个点的MDCT系数表示了8-16kHz的频谱。但超宽带语音通信，不一定要求频谱到16kHz，比如，如果设置频谱到14kHz，只需要考虑前240个点的MDCT系数，对应地，子带数可以控制为6。

对于每一个子带，计算当前子带中所有MDCT系数的平均能量作为子带谱包络（子带谱包络是经过频谱各主峰值点的平滑曲线），比如，当前子带中包括的MDCT系数为 $x(n), n=1, 2, \dots, 40$ ，则平均能量 $Y = ((x(1)^2 + x(2)^2 + \dots + x(40)^2) / 40)$ 。对于320个点的MDCT系数分成8个子带的情况，可以获得8个子带谱包络，这8个子带谱包络，就是生成的高频信号的特征向量 $F_{HB}(n)$ 。

总之，在高频分析部分，输出一个8维的特征向量，表征输入信号的高频部分的关键特征。因此，只需少量数据量，即可表示高频信息，编码效率显著提升。

步骤14、量化编码。

对于低频信号的特征向量 $F_{LB}(n)$ 和低频信号的特征向量 $F_{HB}(n)$ ，均可以进行标量量化（各分量单独量化）和熵编码的方法。另外，本申请实施例也不限制矢量量化（相邻多个分量组合成一个矢量进行联合量化）和熵编码的技术组合。

对特征向量进行量化编码后，可以生成对应的码流。根据实验，通过6-10kbps码率就可以对32kHz超宽带信号实现高质量压缩。

因此，本申请实施例通过对输入信号进行下采样滤波，提取到低频信号，由于低频信号相对于音频信号中的高频信号对音频编码的影响更大，因此通过差异化的信号处理，分别提取输入信号的低频特征以及高频特征，使得高频特征的特征维度低于低频特征的特征维度，对减小了特征维度的低频特征以及高频特征分别进行编码，从而在保证音频

质量的情况下,提高音频编码效率。

下面对本申请实施例提供的低速率神经网络编码方法进行具体说明。

承接上述示例,关于解码端的流程如下:

步骤 21、量化解码。

- 5 量化解码是量化编码的逆过程。对于收到码流,先进行熵解码,得到码流对应的索引值,并通过码流对应的索引值查询量化表,以获得低频的特征向量的估计值 $F'_{LB}(n)$ 和高频的特征向量的估计值 $F'_{HB}(n)$ 。

步骤 22、将低频的特征向量的估计值 $F'_{LB}(n)$ 输入第二 NN。

- 10 基于低频的特征向量的估计值 $F'_{LB}(n)$,调用如图 11 所示的第二 NN,生成低频信号的估计值 $x'_{LB}(n)$ 。

其中,第二 NN 与第一 NN 类似,比如因果卷积、第二 NN 中的后处理结构与第一 NN 中的预处理类似。解码块的结构与编码端的编码块是对称地:编码端的编码块是先做空洞卷积再池化完成下采样,解码端的解码块是先进行池化完成上采样,再做空洞卷积。

- 15 步骤 23、对低频信号的估计值 $x'_{LB}(n)$ 进行上采样滤波。

需要说明的是,生成的低频信号的估计值 $x'_{LB}(n)$ 的采样率仅为编码端输入信号 $x(n)$ 的采样率的 1/2。因此,需要对低频信号的估计值 $x'_{LB}(n)$ 进行上采样滤波,将采样率扩大 1 倍,生成上采样信号 $x_{up}(n)$,其中,每帧 $x_{up}(n)$ 包含 640 个点。本申请实施例中的上采样滤波相当于一种因子为 2 的上采样操作。

- 20 步骤 24、通过上采样信号以及高频的特征向量的估计值 $F'_{HB}(n)$ 进行信号重建。

- 信号重建的第一步,将获得的上采样信号 $x_{up}(n)$ 进行修改型离散余弦变换(MDCT, Modified Discrete Cosine Transform),类似编码端配置,即计算 1280 个点的 MDCT,获得 640 个点的 MDCT 系数,即 $X_{up}(n)$ 。例如,对于包括 640 个点的上采样信号 $x_{up}(n)$,调用修改型离散余弦变换(MDCT, Modified Discrete Cosine Transform),将上一帧上采样信号与这一帧上采样信号 $x_{up}(n)$ 合并,计算 1280 个点的 MDCT,获得 640 个点的 MDCT 系数。由于 $x_{up}(n)$ 是一种上采样信号形式,因此, $X_{up}(n)$ 只保留了 0-8kHz 部分内容,8-16kHz 内容是缺失的。

- 信号重建的第二步,基于获得的高频的特征向量的估计值 $F'_{HB}(n)$,调用频带扩展技术,生成缺失的高频成份。其中, $F'_{HB}(n)$ 为码流中解码出来的 8 个高频部分的子带谱包络。信号重建的第二步的具体操作如下:

- 30 将 $X_{up}(n)$ 中的低频 MDCT 系数(即前 320 个点的 MDCT 系数)进行复制,生成高频部分的 MDCT 系数的参考值。参考语音信号基本特征,低频部分谐波更多,高频部分谐波更少。因此,为了避免简单复制,造成人工生成的高频 MDCT 谱包含过多谐波,可以将 $X_{up}(n)$ 中的低频 MDCT 系数的后 160 个点作为母版,频谱复制 2 次,生成高频部分信号的 320 个点的 MDCT 系数的参考值(简称参考高频信号的 320 个点的高频 MDCT 系数的参考值);还可以将 $X_{up}(n)$ 中的低频 MDCT 系数作为母版,频谱复制 1 次,生成高频部分信号的 320 个点的 MDCT 系数的参考值;还可以将 $X_{up}(n)$ 中的低频 MDCT 系数的最后的 80 个点作为母版,频谱复制 4 次,生成高频部分信号的 320 个点的 MDCT 系数的参考值。因此,频谱复制后的 $X_{up}(n)$ 中的后 320 个点是而非零系数。

- 40 接下来,调用前面获得的 8 个子带谱包络(即通过查询量化表后得到的高频的特征向量的估计值 $F'_{HB}(n)$),这 8 个子带谱包络对应 8 个高频子带,并将生成的高频部分信号的 320 个点的 MDCT 系数的参考值分为 8 个参考高频子带,基于一个高频子带与对应的参考高频子带,对高频部分信号的 320 个点的 MDCT 系数的参考值进行增益控制(频域上就是做乘法),例如根据高频子带的平均能量与对应的参考高频子带的平均能

量计算增益因子，将该对应的参考高频子带中的每个点对应的 MDCT 系数乘以增益因子，确保解码虚拟生成的高频 MDCT 系数的能量，与编码端原始的系数能量接近。

例如，假定参考高频子带（即由生成的高频部分信号的 320 个点的 MDCT 系数的参考值所划分得到的子带）的平均能量是 Y_L ，当前高频子带（即基于码流解码出来的子带谱包络对应的子带）的平均能量是 Y_H ，则计算一个增益因子 $a = \sqrt{Y_H/Y_L}$ ，其中， $\sqrt{}$ 表示平方根计算函数，用于计算 (Y_H/Y_L) 的平方根。有了增益因子 a 后，直接将参考高频子带中的每个点的 MDCT 系数均乘以 a 。经过增益控制后的 MDCT 系数（虚拟生成的）的平均能量，就跟编码端原始的非常接近的。

最后，对 $X_{up}(n)$ 中的低频 MDCT 系数（即前 320 个点的 MDCT 系数）以及经过增益控制后的 MDCT 系数（即 $X_{up}(n)$ 中增益后的后 320 个点的 MDCT 系数的参考值）进行 MDCT 反变换，生成 1280 个点的 MDCT 的估计值，通过交叠，取有效的前 640 个点的 MDCT 的估计值作为原始的输入信号的估计值，即输出信号 $x'(n)$ 。

因此，本申请实施例对接收的码流（即低频码流以及高频码流）进行解码，获得低频的特征向量以及高频的特征向量，对低频的特征向量调用对应编码端的逆过程，完成低频部分的重建。然后，将重建的低频部分，通过上采样滤波，恢复到与原始输入的语音信号一样的采样率；通过结合上采样滤波后的低频部分、高频的特征向量，重建出高频部分，将重建后的低频部分和高频部分合并，保证重建出的高频部分的能量与编码端的高频能量接近。这样，低频至高频的频域系数获得完整重建。

本申请实施例可以通过采集数据，对编码端和解码端的相关网络进行联合训练，获得最优参数。用户仅需准备好数据和设置相应的网络结构，在后台完成训练后，即可将训练好的模型投入使用。

综上，本申请实施例提供的低速率神经网络编解码方法，通过信号分解、信号处理技术与深度神经网络的有机结合，在保证音频质量、且复杂度可接受的情况下，较信号处理方案显著提升提高编码效率。

至此已经结合本申请实施例提供的终端设备的示例性应用和实施，说明本申请实施例提供的音频编码方法或音频解码方法。本申请实施例还提供音频编码装置以及音频解码装置，实际应用中，音频编码装置以及音频解码装置中的各功能模块可以由电子设备（如终端设备、服务器或服务器集群）的硬件资源，如处理器等计算资源、通信资源（如用于支持实现光缆、蜂窝等各种方式通信）、存储器协同实现。图 3A 示出了存储在存储器 550 中的音频编码装置 555 以及图 3B 示出了存储在存储器 550 中的音频解码装置 556，其可以是程序和插件等形式的软件，例如，软件 C/C++、Java 等编程语言设计的软件模块、C/C++、Java 等编程语言设计的应用软件或大型软件系统中的专用软件模块、应用程序接口、插件、云服务等实现方式，下面对不同的实现方式举例说明。

其中，音频编码装置 555 包括一系列的模块，包括下采样块 5551、低频提取模块 5552、高频分析模块 5553、编码模块 5554。下面继续说明本申请实施例提供的音频编码装置 555 中各个模块配合实现音频编码的方案。

下采样模块 5551，配置为对音频信号进行下采样处理，得到所述音频信号的低频信号；低频提取模块 5552，配置为对所述低频信号进行低频特征提取处理，得到所述音频信号的低频特征；高频分析模块 5553，配置为对所述音频信号进行高频分析处理，得到所述音频信号的高频特征；其中，所述高频特征的特征维度低于所述低频特征的特征维度；编码模块 5554，配置为对所述低频特征进行编码处理，得到所述音频信号的低频码流，并对所述高频特征进行编码处理，得到所述音频信号的高频码流。

在一些实施例中，所述音频信号包括采样得到的多个第一样本点；下采样模块 5551 还配置为通过下采样滤波器对所述音频信号包括的第一采样点进行下采样处理，得到所

述音频信号的低频信号。

在一些实施例中，所述下采样模块 5551 还配置为通过所述下采样滤波器执行以下处理：对所述音频信号包括的采样点进行基于数字信号的滤波处理，得到滤波后的所述音频信号；对滤波后的所述音频信号进行基于数字信号的下采样处理，得到所述音频信号的低频信号。

在一些实施例中，所述低频提取模块 5552 还配置为对所述低频信号进行卷积处理，得到所述低频信号的卷积特征；对所述卷积特征进行池化处理，得到所述低频信号的池化特征；对所述池化特征进行下采样处理，得到所述低频信号的下采样特征；对所述下采样特征进行卷积处理，得到所述音频信号的低频特征。

在一些实施例中，所述下采样处理是通过多个级联的编码层实现的；所述低频提取模块 5552 还配置为通过所述多个级联的编码层中的第一个编码层，对所述池化特征进行下采样处理；将所述第一个编码层的下采样结果输出到后续级联的编码层，通过所述后续级联的编码层继续进行下采样处理和下采样结果输出，直至输出到最后一个编码层；将所述最后一个编码层输出的下采样结果确定为所述低频信号的下采样特征。

在一些实施例中，所述高频分析模块 5553 还配置为对所述音频信号进行频带扩展处理，得到所述音频信号的高频特征。

在一些实施例中，所述高频分析模块 5553 还配置为对所述音频信号包括的多个第二样本点进行频域变换处理，得到所述多个第二样本点分别对应的变换系数；将所述多个第二样本点分别对应的变换系数中的高频变换系数划分为多个子带；对每个所述子带包括的变换系数进行平均处理，得到每个所述子带对应的平均能量，并将所述平均能量确定为每个所述子带对应的子带谱包络；将所述多个子带分别对应的子带谱包络确定为所述音频信号的高频特征。

在一些实施例中，所述高频分析模块 5553 还配置为获取参考音频信号包括的多个第三样本点，其中，所述参考音频信号是与所述音频信号相邻的音频信号；基于所述参考音频信号包括的多个第三样本点以及所述音频信号包括的多个第二样本点，对所述音频信号包括的多个第二样本点进行离散余弦变换处理，得到所述多个第二样本点分别对应的变换系数。

在一些实施例中，所述高频分析模块 5553 还配置为确定每个所述子带包括的第二样本点对应的变换系数的平方和；将所述平方和与所述子带包括的第二样本点的数量的比值，确定为每个所述子带对应的平均能量。

其中，音频解码装置 556 包括一系列的模块，包括解码模块 5561、低频重建模块 5562、上采样模块 5563、信号重建模块 5564。下面继续说明本申请实施例提供的音频解码装置 555 中各个模块配合实现音频解码的方案。

解码模块 5561，配置为对音频信号的低频码流进行解码处理，得到所述低频码流对应的低频特征，并对所述音频信号的高频码流进行解码处理，得到所述高频码流对应的高频特征；其中，所述低频码流是通过所述音频信号经过下采样后得到的低频信号进行编码得到的，所述高频特征的特征维度低于所述低频特征的特征维度；低频重建模块 5562，配置为对所述低频特征进行低频特征重建处理，得到所述低频特征对应的低频信号；上采样模块 5563，配置为对所述低频信号进行上采样处理，得到所述低频信号的上采样信号；信号重建模块 5564，配置为对所述高频特征以及所述上采样信号进行信号重建处理，得到所述低频码流和所述高频码流对应的合成音频信号。

在一些实施例中，所述音频信号包括采样得到的多个第四样本点；所述上采样模块 5563 还配置为通过上采样滤波器对所述低频信号包括的第四采样点进行上采样处理，得到所述低频信号的上采样信号。

在一些实施例中,所述通上采样模块 5563 还配置为通过所述上采样滤波器执行以下处理:对所述低频信号包括的第四采样点进行基于数字信号的上采样处理,得到上采样后的所述低频信号;对上采样后的所述低频信号进行基于数字信号的滤波处理,得到所述低频信号的上采样信号。

5 在一些实施例中,所述低频重建模块 5562 还配置为对所述低频特征进行卷积处理,得到所述低频特征的卷积特征;对所述卷积特征进行上采样处理,得到所述低频特征的上采样特征;对所述上采样特征进行池化处理,得到所述低频特征的池化特征;对所述池化特征进行卷积处理,得到所述低频特征对应的低频信号。

10 在一些实施例中,所述上采样处理是通过多个级联的解码层实现的;所述低频重建模块 5562 还配置为通过所述多个级联的解码层中的第一个解码层,对所述卷积特征进行上采样处理;将所述第一个解码层的上采样结果输出到后续级联的解码层,通过所述后续级联的解码层继续进行上采样处理和上采样结果输出,直至输出到最后一个解码层;将所述最后一个解码层输出的上采样结果确定为所述低频特征的上采样特征。

15 在一些实施例中,所述信号重建模块 5564 还配置为对所述上采样信号包括的多个第五样本点进行频域变换处理,得到所述多个第五样本点分别对应的变换系数;对所述多个第五样本点分别对应的变换系数以及所述高频特征进行频带扩展处理,得到高频信号的高频变换系数;基于所述高频变换系数以及所述多个第五样本点中的部分第五样本点分别对应的变换系数,确定所述低频码流和所述高频码流对应的合成音频信号。

20 在一些实施例中,所述信号重建模块 5564 还配置为对所述多个第五样本点分别对应的变换系数中的前半部分的至少部分变换系数进行频谱复制处理,得到参考高频信号的参考高频变换系数;基于所述高频特征对应的子带谱包络,对所述参考高频信号的参考高频变换系数进行增益处理,得到所述高频信号的高频变换系数。

25 在一些实施例中,所述信号重建模块 5564 还配置为基于所述高频特征对应的子带谱包络,将所述参考高频信号的参考高频变换系数划分为多个子带;针对所述多个子带中的任意所述子带执行以下处理:确定所述子带谱包络中与所述子带对应的高频子带的第一平均能量,并确定所述子带的第二平均能量;基于所述第一平均能量与所述第二平均能量的比值,确定增益因子;将所述增益因子与所述子带包括的每个所述参考高频变换系数相乘,得到所述子带对应的高频变换系数;将所述多个子带分别对应的高频变换系数确定为所述高频信号的高频变换系数。

30 在一些实施例中,所述信号重建模块 5564 还配置为对所述多个第五样本点分别对应的变换系数中的前半部分的后半部分变换系数进行频谱复制处理,得到参考高频信号的参考高频变换系数。

35 在一些实施例中,所述信号重建模块 5564 还配置为对所述高频变换系数以及所述多个第五样本点中的前半部分第五样本点分别对应的变换系数进行合并处理,得到完整变换系数;对所述完整变换系数进行频谱变换的逆变换处理,得到所述低频码流和所述高频码流对应的合成音频信号。

40 本申请实施例提供了一种计算机程序产品,该计算机程序产品包括计算机程序或计算机可执行指令,该计算机程序或计算机可执行指令存储在计算机可读存储介质中。电子设备的处理器从计算机可读存储介质读取该计算机程序或计算机可执行指令,处理器执行该计算机程序或计算机可执行指令,使得该电子设备执行本申请实施例上述的音频编码方法或音频解码方法。

45 本申请实施例提供一种存储有计算机可执行指令的计算机可读存储介质,其中存储有计算机可执行指令或者计算机程序,当计算机可执行指令或者计算机程序被处理器执行时,将引起处理器执行本申请实施例提供的音频编码方法或音频解码方法,例如,如

图 6A-图 6D 示出的音频编码方法，图 7A-图 7D 示出的音频编码方法。

在一些实施例中，计算机可读存储介质可以是 FRAM、ROM、PROM、EPROM、EEPROM、闪存、磁表面存储器、光盘、或 CD-ROM 等存储器；也可以是包括上述存储器之一或任意组合的各种设备。

5 在一些实施例中，计算机可执行指令可以采用程序、软件、软件模块、脚本或代码的形式，按任意形式的编程语言（包括编译或解释语言，或者声明性或过程性语言）来编写，并且其可按任意形式部署，包括被部署为独立的程序或者被部署为模块、组件、子例程或者适合在计算环境中使用的其它单元。

10 作为示例，计算机可执行指令可以但不一定对应于文件系统中的文件，可以可被存储在保存其它程序或数据的文件的一部分，例如，存储在超文本标记语言（HTML, Hyper Text Markup Language）文档中的一个或多个脚本中，存储在专用于所讨论的程序的单个文件中，或者，存储在多个协同文件（例如，存储一个或多个模块、子程序或代码部分的文件）中。

15 作为示例，计算机可执行指令可被部署为在一个电子设备上执行，或者在位于一个地点的多个电子设备上执行，又或者，在分布在多个地点且通过通信网络互连的多个电子设备上执行。

可以理解的是，在本申请实施例中，涉及到用户信息等相关的数据，当本申请实施例运用到具体产品或技术中时，需要获得用户许可或者同意，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。

20 以上所述，仅为本申请的实施例而已，并非用于限定本申请的保护范围。凡在本申请的精神和范围之内所作的任何修改、等同替换和改进等，均包含在本申请的保护范围之内。

权利要求书

- 1、一种音频编码方法，应用于电子设备，所述方法包括：
对音频信号进行下采样处理，得到所述音频信号的低频信号；
对所述低频信号进行低频特征提取处理，得到所述音频信号的低频特征；
5 对所述音频信号进行高频分析处理，得到所述音频信号的高频特征；
其中，所述高频特征的特征维度低于所述低频特征的特征维度；
对所述低频特征进行编码处理，得到所述音频信号的低频码流，并对所述高频特征进行编码处理，得到所述音频信号的高频码流。
- 2、根据权利要求1所述的方法，其中，
10 所述音频信号包括采样得到的多个第一样本点；
所述对音频信号进行下采样处理，得到所述音频信号的低频信号，包括：
通过下采样滤波器对所述音频信号包括的第一采样点进行下采样处理，得到所述音频信号的低频信号。
- 3、根据权利要求2所述的方法，其中，所述通过下采样滤波器对所述音频信号包括的第一采样点进行下采样处理，得到所述音频信号的低频信号，包括：
15 通过所述下采样滤波器执行以下处理：
对所述音频信号包括的第一采样点进行基于数字信号的滤波处理，得到滤波后的所述音频信号；
对滤波后的所述音频信号进行基于数字信号的下采样处理，得到所述音频信号的低频信号。
20 频信号。
- 4、根据权利要求1-3任一项所述的方法，其中，所述对所述低频信号进行低频特征提取处理，得到所述音频信号的低频特征，包括：
对所述低频信号进行卷积处理，得到所述低频信号的卷积特征；
对所述卷积特征进行池化处理，得到所述低频信号的池化特征；
25 对所述池化特征进行下采样处理，得到所述低频信号的下采样特征；
对所述下采样特征进行卷积处理，得到所述音频信号的低频特征。
- 5、根据权利要求4所述的方法，其中，
所述下采样处理是通过多个级联的编码层实现的；
所述对所述池化特征进行下采样处理，得到所述低频信号的下采样特征，包括：
30 通过所述多个级联的编码层中的第一个编码层，对所述池化特征进行下采样处理；
将所述第一个编码层的下采样结果输出到后续级联的编码层，通过所述后续级联的编码层继续进行下采样处理和下采样结果输出，直至输出到最后一个编码层；
将所述最后一个编码层输出的下采样结果确定为所述低频信号的下采样特征。
- 6、根据权利要求1-5任一项所述的方法，其中，所述对所述音频信号进行高频分析处理，得到所述音频信号的高频特征，包括：
35 对所述音频信号进行频带扩展处理，得到所述音频信号的高频特征。
- 7、根据权利要求6所述的方法，其中，所述对所述音频信号进行频带扩展处理，得到所述音频信号的高频特征，包括：
对所述音频信号包括的多个第二样本点进行频域变换处理，得到所述多个第二样本点分别对应的变换系数；
40 将所述多个第二样本点分别对应的变换系数中的高频变换系数划分为多个子带；
对每个所述子带包括的变换系数进行平均处理，得到每个所述子带对应的平均能量，并将所述平均能量确定为每个所述子带对应的子带谱包络；
将所述多个子带分别对应的子带谱包络确定为所述音频信号的高频特征。

8、根据权利要求 7 所述的方法，其中，所述对所述音频信号包括的多个第二样本点进行频域变换处理，得到所述多个第二样本点分别对应的变换系数，包括：

获取参考音频信号包括的多个第三样本点，其中，所述参考音频信号是与所述音频信号相邻的音频信号；

5 基于所述参考音频信号包括的多个第三样本点以及所述音频信号包括的多个第二样本点，对所述音频信号包括的多个第二样本点进行离散余弦变换处理，得到所述多个第二样本点分别对应的变换系数。

9、根据权利要求 7 所述的方法，其中，所述对每个所述子带包括的变换系数进行平均处理，得到每个所述子带对应的平均能量，包括：

10 确定每个所述子带包括的第二样本点对应的变换系数的平方和；

将所述平方和与所述子带包括的第二样本点的数量的比值，确定为每个所述子带对应的平均能量。

10、一种音频解码方法，应用于电子设备，所述方法包括：

15 对音频信号的低频码流进行解码处理，得到所述低频码流对应的低频特征，并对所述音频信号的高频码流进行解码处理，得到所述高频码流对应的高频特征；

其中，所述低频码流是通过所述音频信号经过下采样后得到的低频信号进行编码得到的，所述高频特征的特征维度低于所述低频特征的特征维度；

对所述低频特征进行低频特征重建处理，得到所述低频特征对应的低频信号；

对所述低频信号进行上采样处理，得到所述低频信号的上采样信号；

20 对所述高频特征以及所述上采样信号进行信号重建处理，得到所述低频码流和所述高频码流对应的合成音频信号。

11、根据权利要求 10 所述的方法，其中，所述对所述低频信号进行上采样处理，得到所述低频信号的上采样信号，包括：

25 通过上采样滤波器对所述低频信号包括的第四采样点进行上采样处理，得到所述低频信号的上采样信号。

12、根据权利要求 11 所述的方法，其中，所述通过上采样滤波器对所述低频信号包括的第四采样点进行上采样处理，得到所述低频信号的上采样信号，包括：

通过所述上采样滤波器执行以下处理：

30 对所述低频信号包括的第四采样点进行基于数字信号的上采样处理，得到上采样后的所述低频信号；

对上采样后的所述低频信号进行基于数字信号的滤波处理，得到所述低频信号的上采样信号。

13、根据权利要求 10-12 任一项所述的方法，其中，所述对所述低频特征进行低频特征重建处理，得到所述低频特征对应的低频信号，包括：

35 对所述低频特征进行卷积处理，得到所述低频特征的卷积特征；

对所述卷积特征进行上采样处理，得到所述低频特征的上采样特征；

对所述上采样特征进行池化处理，得到所述低频特征的池化特征；

对所述池化特征进行卷积处理，得到所述低频特征对应的低频信号。

14、根据权利要求 13 所述的方法，其中，

40 所述上采样处理是通过多个级联的解码层实现的；

所述对所述卷积特征进行上采样处理，得到所述低频特征的上采样特征，包括：

通过所述多个级联的解码层中的第一个解码层，对所述卷积特征进行上采样处理；

将所述第一个解码层的上采样结果输出到后续级联的解码层，通过所述后续级联的解码层继续进行上采样处理和上采样结果输出，直至输出到最后一个解码层；

将所述最后一个解码层输出的上采样结果确定为所述低频特征的上采样特征。

15、根据权利要求 10-12 任一项所述的方法，其中，所述对所述高频特征以及所述上采样信号进行信号重建处理，得到所述低频码流和所述高频码流对应的合成音频信号，包括：

5 对所述上采样信号包括的多个第五样本点进行频域变换处理，得到所述多个第五样本点分别对应的变换系数；

对所述多个第五样本点分别对应的变换系数以及所述高频特征进行频带扩展处理，得到高频信号的高频变换系数；

10 基于所述高频变换系数以及所述多个第五样本点中的部分第五样本点分别对应的变换系数，确定所述低频码流和所述高频码流对应的合成音频信号。

16、根据权利要求 15 所述的方法，其中，所述对所述多个第五样本点分别对应的变换系数以及所述高频特征进行频带扩展处理，得到高频信号的高频变换系数，包括：

对所述多个第五样本点分别对应的变换系数中的前半部分的至少部分变换系数进行频谱复制处理，得到参考高频信号的参考高频变换系数；

15 基于所述高频特征对应的子带谱包络，对所述参考高频信号的参考高频变换系数进行增益处理，得到所述高频信号的高频变换系数。

17、根据权利要求 16 所述的方法，其中，所述基于所述高频特征对应的子带谱包络，对所述参考高频信号的参考高频变换系数进行增益处理，得到所述高频信号的高频变换系数，包括：

20 基于所述高频特征对应的子带谱包络，将所述参考高频信号的参考高频变换系数划分为多个子带；

针对所述多个子带中的每个所述子带执行以下处理：

确定所述子带谱包络中与所述子带对应的高频子带的第一平均能量，并确定所述子带的第二平均能量；

25 基于所述第一平均能量与所述第二平均能量的比值，确定增益因子；

将所述增益因子与所述子带包括的每个所述参考高频变换系数相乘，得到所述子带对应的高频变换系数；

将所述多个子带分别对应的高频变换系数确定为所述高频信号的高频变换系数。

30 18、根据权利要求 16 所述的方法，其中，所述对所述多个第五样本点分别对应的变换系数中的前半部分的至少部分变换系数进行频谱复制处理，得到参考高频信号的参考高频变换系数，包括：

对所述多个第五样本点分别对应的变换系数中的前半部分的后半部分变换系数进行频谱复制处理，得到参考高频信号的参考高频变换系数。

35 19、根据权利要求 15 所述的方法，其中，所述基于所述高频变换系数以及所述多个第五样本点中的部分第五样本点分别对应的变换系数，确定所述低频码流和所述高频码流对应的合成音频信号，包括：

对所述高频变换系数以及所述多个第五样本点中的前半部分第五样本点分别对应的变换系数进行合并处理，得到完整变换系数；

40 对所述完整变换系数进行频谱变换的逆变换处理，得到所述低频码流和所述高频码流对应的合成音频信号。

20、一种音频编码装置，所述装置包括：

下采样模块，配置为对音频信号进行下采样处理，得到所述音频信号的低频信号；

低频提取模块，配置为对所述低频信号进行低频特征提取处理，得到所述音频信号的低频特征；

高频分析模块，配置为对所述音频信号进行高频分析处理，得到所述音频信号的高频特征；

其中，所述高频特征的特征维度低于所述低频特征的特征维度；

5 编码模块，配置为对所述低频特征进行编码处理，得到所述音频信号的低频码流，并对所述高频特征进行编码处理，得到所述音频信号的高频码流。

21、一种音频解码装置，所述装置包括：

解码模块，配置为对音频信号的低频码流进行解码处理，得到所述低频码流对应的低频特征，并对所述音频信号的高频码流进行解码处理，得到所述高频码流对应的高频特征；

10 其中，所述低频码流是通过对所述音频信号经过下采样后得到的低频信号进行编码得到的，所述高频特征的特征维度低于所述低频特征的特征维度；

低频重建模块，配置为对所述低频特征进行低频特征重建处理，得到所述低频特征对应的低频信号；

15 上采样模块，配置为对所述低频信号进行上采样处理，得到所述低频信号的上采样信号；

信号重建模块，配置为对所述高频特征以及所述上采样信号进行信号重建处理，得到所述低频码流和所述高频码流对应的合成音频信号。

22、一种电子设备，所述电子设备包括：

存储器，用于存储计算机程序或计算机可执行指令；

20 处理器，用于执行所述存储器中存储的计算机程序或计算机可执行指令时，实现权利要求 1 至 9 任一项所述的音频编码方法，或权利要求 10 至 19 任一项所述的音频解码方法。

23、一种计算机可读存储介质，存储有计算机程序或者计算机可执行指令，所述计算机程序或计算机可执行指令被处理器执行时实现权利要求 1 至 9 任一项所述的音频编码方法，或权利要求 10 至 19 任一项所述的音频解码方法。

25 24、一种计算机程序产品，包括计算机程序或计算机可执行指令，所述计算机程序或计算机可执行指令被处理器执行时实现权利要求 1 至 9 任一项所述的音频编码方法，或权利要求 10 至 19 任一项所述的音频解码方法。