



(12) 发明专利申请

(10) 申请公布号 CN 118942471 A

(43) 申请公布日 2024. 11. 12

(21) 申请号 202411353107.4

G10L 19/032 (2013.01)

(22) 申请日 2022.06.15

(62) 分案原申请数据

202210681037.X 2022.06.15

(71) 申请人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区
科技中一路腾讯大厦35层

(72) 发明人 王蒙 肖玮 康迁勇 黄庆博

史裕鹏

(74) 专利代理机构 北京派特恩知识产权代理有

限公司 11270

专利代理师 崔晓岚 张颖玲

(51) Int. Cl.

G10L 19/16 (2013.01)

G10L 19/008 (2013.01)

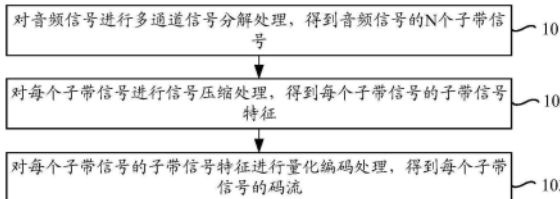
权利要求书4页 说明书25页 附图8页

(54) 发明名称

音频处理方法、装置、设备、存储介质及计算机程序产品

(57) 摘要

本申请提供了一种音频处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品；方法包括：对音频信号进行多通道信号分解处理，得到音频信号的N个子带信号，其中，N为大于2的整数，N个子带信号的频段依次递增；对每个子带信号进行信号压缩处理，得到每个子带信号子带信号特征；对每个子带信号子带信号特征进行量化编码处理，得到每个子带信号码流。通过本申请，能够提高音频编码效率。



1. 一种音频处理方法,其特征在于,所述方法包括:
 - 获取音频信号的N个子带信号,其中,所述N为大于2的整数;
 - 通过神经网络模型对前k个子带信号分别进行特征提取处理,得到所述前k个子带信号分别对应的子带信号特征;
 - 对后N-k个子带信号分别进行频带扩展处理,得到所述后N-k个子带信号分别对应的子带信号特征,其中,所述k为整数且取值范围为 $1 \leq k < N$;
 - 对每个所述子带信号的子带信号特征进行量化编码处理,得到每个所述子带信号的码流。
2. 根据权利要求1所述的方法,其特征在于,所述对后N-k个子带信号分别进行频带扩展处理,得到所述后N-k个子带信号分别对应的子带信号特征,包括:
 - 针对所述后N-k个子带信号中的每个所述子带信号执行以下处理:
 - 基于所述子带信号包括的多个样本点进行频域变换处理,得到所述多个样本点分别对应的变换系数;
 - 将所述多个样本点分别对应的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均值处理,得到每个所述子带对应的平均能量,并将所述平均能量作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子带信号对应的子带信号特征。
3. 根据权利要求2所述的方法,其特征在于,所述基于所述子带信号包括的多个样本点进行频域变换处理,得到所述多个样本点分别对应的变换系数,包括:
 - 获取参考音频信号的参考子带信号,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,所述参考子带信号与所述子带信号的频段相同;
 - 基于所述参考子带信号包括的多个样本点以及所述子带信号包括的多个样本点,对所述子带信号包括的多个样本点进行离散余弦变换处理,得到所述子带信号包括的多个样本点分别对应的变换系数。
4. 根据权利要求1所述的方法,其特征在于,所述通过神经网络模型对前k个子带信号分别进行特征提取处理,得到所述前k个子带信号分别对应的子带信号特征,包括:
 - 通过所述神经网络模型对前k个子带信号中的每个所述子带信号执行以下处理:
 - 对所述子带信号进行卷积处理,得到所述子带信号的卷积特征;
 - 对所述卷积特征进行池化处理,得到所述子带信号的池化特征;
 - 对所述池化特征进行下采样处理,得到所述子带信号的下采样特征;
 - 对所述下采样特征进行卷积处理,得到所述子带信号的特征。
5. 根据权利要求1所述的方法,其特征在于,每个所述子带信号的特征维度与每个所述子带信号的频段非正相关,第N个子带信号的特征维度小于第一个子带信号的特征维度。
6. 根据权利要求1所述的方法,其特征在于,所述获取音频信号的N个子带信号之前,所述方法还包括:
 - 对音频信号进行多通道信号分解处理,得到所述音频信号的N个子带信号。
7. 一种音频处理方法,其特征在于,所述方法包括:
 - 获取音频信号的N个子带信号,其中,所述N为大于2的整数;

通过每个所述子带信号对应的第一神经网络模型,对每个所述子带信号进行特征提取处理,得到每个所述子带信号的子带信号特征;

对每个所述子带信号的子带信号特征进行量化编码处理,得到每个所述子带信号的码流。

8. 根据权利要求7所述的方法,其特征在于,所述第一神经网络模型的结构复杂度与所述子带信号的子带信号特征的维度正相关。

9. 根据权利要求7所述的方法,其特征在于,所述通过每个所述子带信号对应的第一神经网络模型,对每个所述子带信号进行特征提取处理,得到每个所述子带信号的子带信号特征,包括:

通过每个所述子带信号对应的第一神经网络模型对每个所述子带信号执行以下处理:

对所述子带信号进行卷积处理,得到所述子带信号的卷积特征;

对所述卷积特征进行池化处理,得到所述子带信号的池化特征;

对所述池化特征进行下采样处理,得到所述子带信号的下采样特征;

对所述下采样特征进行卷积处理,得到所述子带信号的子带信号特征。

10. 一种音频处理方法,其特征在于,所述方法包括:

对N个码流进行量化解码处理,得到每个所述码流对应的子带信号特征;

其中,所述N为大于2的整数,所述N个码流是对音频信号的N个子带信号分别进行编码得到的;

通过神经网络模型对前k个子带信号特征分别进行特征重建处理,得到所述前k个子带信号特征分别对应的估计子带信号;

对后N-k个子带信号特征分别进行频带扩展的逆处理,得到所述后N-k个子带信号特征分别对应的估计子带信号,其中,所述k为整数且取值范围为 $1 \leq k < N$;

基于多个所述估计子带信号,确定所述多个码流对应的合成音频信号。

11. 根据权利要求10所述的方法,其特征在于,所述对后N-k个子带信号特征分别进行频带扩展的逆处理,得到所述后N-k个子带信号特征分别对应的估计子带信号,包括:

针对所述后N-k个子带信号特征中的每个所述子带信号特征执行以下处理:

对所述前k个估计子带信号中与所述子带信号特征关联的所述估计子带信号进行信号合成处理,得到与所述子带信号特征对应的低频子带信号;

基于所述低频子带信号包括的多个样本点进行频域变换处理,得到所述多个样本点分别对应的变换系数;

对所述多个样本点分别对应的变换系数中的后半部分的变换系数进行频谱复制处理,得到参考子带信号的参考变换系数;

基于所述子带信号特征对应的子带谱包络,对所述参考子带信号的参考变换系数进行增益处理,得到增益后的所述参考变换系数;

对增益后的所述参考变换系数进行反频域变换处理,得到所述子带信号特征对应的估计子带信号。

12. 根据权利要求11所述的方法,其特征在于,所述基于所述子带信号特征对应的子带谱包络,对所述参考子带信号的参考变换系数进行增益处理,得到增益后的所述参考变换系数,包括:

基于所述子带信号特征对应的子带谱包络,将所述参考子带信号的参考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第一平均能量,并确定所述子带对应的第二平均能量;

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变换系数相乘,得到增益后的所述参考变换系数。

13. 根据权利要求10所述的方法,其特征在于,所述通过神经网络模型对前k个子带信号特征分别进行特征重建处理,得到所述前k个子带信号特征分别对应的估计子带信号,包括:

通过所述神经网络模型对前k个子带信号特征中的每个所述子带信号特征执行以下处理:

对所述子带信号特征进行卷积处理,得到所述子带信号特征的卷积特征;

对所述卷积特征进行上采样处理,得到所述子带信号特征的上采样特征;

对所述上采样特征进行反池化处理,得到所述子带信号特征的反池化特征;

对所述反池化特征进行卷积处理,得到所述子带信号特征对应的估计子带信号。

14. 根据权利要求10所述的方法,其特征在于,所述基于多个所述估计子带信号,确定所述多个码流对应的合成音频信号,包括:

对多个所述估计子带信号分别进行上采样处理,得到多个所述估计子带信号分别对应的滤波信号;

对多个所述滤波信号进行滤波合成处理,得到所述多个码流对应的合成音频信号。

15. 一种音频处理方法,其特征在于,所述方法包括:

对N个码流进行量化解码处理,得到每个所述码流对应的子带信号特征;

其中,所述N为大于2的整数,所述N个码流是对音频信号的N个子带信号分别进行编码得到的;

通过每个所述子带信号特征对应的第二神经网络模型,对每个所述子带信号特征进行特征重建,得到每个所述子带信号特征对应的估计子带信号;

基于多个所述估计子带信号,确定所述多个码流对应的合成音频信号。

16. 根据权利要求15所述的方法,其特征在于,所述第二神经网络模型的结构复杂度与所述子带信号特征的维度正相关。

17. 根据权利要求15所述的方法,其特征在于,所述通过每个所述子带信号特征对应的第二神经网络模型,对每个所述子带信号特征进行特征重建,得到每个所述子带信号特征对应的估计子带信号,包括:

通过每个所述子带信号特征对应的第二神经网络模型对每个所述子带信号特征执行以下处理:

对所述子带信号特征进行卷积处理,得到所述子带信号特征的卷积特征;

对所述卷积特征进行上采样处理,得到所述子带信号特征的上采样特征;

对所述上采样特征进行反池化处理,得到所述子带信号特征的反池化特征;

对所述反池化特征进行卷积处理,得到所述子带信号特征对应的估计子带信号。

18.一种音频处理装置,其特征在于,所述装置包括:

分解模块,用于获取音频信号的N个子带信号,其中,所述N为大于2的整数;

压缩模块,用于通过神经网络模型对前k个子带信号分别进行特征提取处理,得到所述前k个子带信号分别对应的子带信号特征;

对后N-k个子带信号分别进行频带扩展处理,得到所述后N-k个子带信号分别对应的子带信号特征,其中,所述k为整数且取值范围为 $1 \leq k < N$;

编码模块,用于对每个所述子带信号的特征进行量化编码处理,得到每个所述子带信号的码流。

19.一种计算机可读存储介质,其特征在于,存储有可执行指令,用于被处理器执行时实现权利要求1至17任一项所述的音频处理方法。

20.一种计算机程序产品,包括计算机程序或指令,其特征在于,所述计算机程序或指令被处理器执行时实现权利要求1至17任一项所述的音频处理方法。

音频处理方法、装置、设备、存储介质及计算机程序产品

[0001] 本申请是申请号为202210681037.X,申请日为2022年06月15日,发明名称为“音频处理方法、装置、设备、存储介质及计算机程序产品”的专利申请的分案申请。

技术领域

[0002] 本申请涉及数据处理技术,尤其涉及一种音频处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品。

背景技术

[0003] 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。

[0004] 然而,对于如何在保证音频质量的情况下,有效提高音频编码的效率,相关技术尚无有效的解决方案。

发明内容

[0005] 本申请实施例提供一种音频处理方法、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。

[0006] 本申请实施例的技术方案是这样实现的:

[0007] 本申请实施例提供一种音频处理方法,包括:

[0008] 对音频信号进行多通道信号分解处理,得到所述音频信号的N个子带信号,其中,所述N为大于2的整数,所述N个子带信号的频段依次递增;

[0009] 对每个所述子带信号进行信号压缩处理,得到每个所述子带信号子带信号特征;

[0010] 对每个所述子带信号子带信号特征进行量化编码处理,得到每个所述子带信号码流。

[0011] 本申请实施例提供一种音频处理方法,包括:

[0012] 对N个码流进行量化解码处理,得到每个所述码流对应的子带信号特征;

[0013] 其中,所述N为大于2的整数,所述N个码流是对音频信号经过多通道信号分解处理后得到的N个子带信号分别进行编码得到的;

[0014] 对每个所述子带信号特征进行信号解压处理,得到每个所述子带信号特征对应的估计子带信号;

[0015] 对多个所述估计子带信号进行信号合成处理,得到所述多个码流对应的合成音频信号。

[0016] 本申请实施例提供一种音频处理装置,包括:

- [0017] 分解模块,用于对音频信号进行多通道信号分解处理,得到所述音频信号的N个子带信号,其中,所述N为大于2的整数,所述N个子带信号的频段依次递增;
- [0018] 压缩模块,用于对每个所述子带信号进行信号压缩处理,得到每个所述子带信号子带信号特征;
- [0019] 编码模块,用于对每个所述子带信号子带信号特征进行量化编码处理,得到每个所述子带信号的码流。
- [0020] 本申请实施例提供一种音频处理装置,包括:
- [0021] 解码模块,用于对N个码流进行量化解码处理,得到每个所述码流对应的子带信号特征;
- [0022] 其中,所述N为大于2的整数,所述N个码流是对音频信号经过多通道信号分解处理后得到的N个子带信号分别进行编码得到的;
- [0023] 解压模块,用于对每个所述子带信号特征进行信号解压处理,得到每个所述子带信号特征对应的估计子带信号;
- [0024] 合成模块,用于对多个所述估计子带信号进行信号合成处理,得到所述多个码流对应的合成音频信号。
- [0025] 本申请实施例提供一种用于音频处理的电子设备,所述电子设备包括:
- [0026] 存储器,用于存储可执行指令;
- [0027] 处理器,用于执行所述存储器中存储的可执行指令时,实现本申请实施例提供的音频处理方法。
- [0028] 本申请实施例提供一种计算机可读存储介质,存储有可执行指令,用于引起处理器执行时,实现本申请实施例提供的音频处理方法。
- [0029] 本申请实施例提供一种计算机程序产品,包括计算机程序或指令,所述计算机程序或指令被处理器执行时实现本申请实施例提供的音频处理方法。
- [0030] 本申请实施例具有以下有益效果:
- [0031] 通过将音频信号分解为多个子带信号,对子带信号进行差异化的信号处理,并对减小了特征维度的子带信号特征进行量化编码,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

- [0032] 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;
- [0033] 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;
- [0034] 图3是本申请实施例提供的电子设备的结构示意图;
- [0035] 图4是本申请实施例提供的音频处理方法的流程示意图;
- [0036] 图5是本申请实施例提供的音频处理方法的流程示意图;
- [0037] 图6是本申请实施例提供的端到端的语音通信链路示意图;
- [0038] 图7是本申请实施例提供的基于子带分解和神经网络的语音编解码方法的流程示意图;
- [0039] 图8A是本申请实施例提供的滤波器组示意图;
- [0040] 图8B是本申请实施例提供的基于滤波器组得到4通道的子带信号的原理示意图;

- [0041] 图8C是本申请实施例提供的基于滤波器组得到3通道的子带信号的原理示意图；
- [0042] 图9A是本申请实施例提供的普通卷积网络的示意图；
- [0043] 图9B是本申请实施例提供的空洞卷积网络的示意图；
- [0044] 图10是本申请实施例提供的频带扩展的示意图；
- [0045] 图11是本申请实施例提供的通道分析的网络结构图；
- [0046] 图12是本申请实施例提供的通道合成的网络结构。

具体实施方式

[0047] 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。

[0048] 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。

[0049] 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。

[0050] 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。

[0051] 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。

[0052] 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。

[0053] 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。

[0054] 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。

[0055] 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间,每个小区间有一个代表值(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。

[0056] 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)。

[0057] 5) 正交镜像滤波器组(QMF, Quadrature Mirror Filters):是一个包含分析-合成的滤波器对,其中,QMF分析滤波器组用于子带信号分解,以降低信号带宽,使各个子带信号可顺利通过各自通道进行处理;QMF合成滤波器组用于将解码端恢复出的各子带信号进行合成处理,例如通过零值内插和带通滤波等方式重建出原始的音频信号。

[0058] 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。

[0059] 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。

[0060] 相关技术中,语音编码根据编码原理大致如下: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。

[0061] 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音频处理方法、装置、电子设备、计算机可读存储介质及计算机程序产品,能够提高编码效率。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。

[0062] 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100的架构示意图,音频解码系统100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。

[0063] 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编码处理,得到码流。

[0064] 例如,客户端410调用本申请实施例提供的音频处理方法对采集得到的音频信号进行编码,即对音频信号进行多通道信号分解处理,得到音频信号的N个子带信号,对每个子带信号进行信号压缩处理,得到每个子带信号子带信号特征;对每个子带信号子带信号特征进行量化编码处理,得到每个子带信号的码流。

[0065] 客户端410可以将码流(即低频码流以及高频码流)通过网络300发送至服务器200,以使服务器200将码流发送至接收方(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。

[0066] 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发送的码流后,可以对码流进行解码处理,以得到音频信号,从而实现音频通信。

[0067] 例如,客户端510调用本申请实施例提供的音频处理方法对接收到的码流进行解码,即对N个码流进行量化解码处理,得到每个码流对应的子带信号特征;对每个子带信号特征进行信号解压处理,得到每个子带信号特征对应的估计子带信号;对多个估计子带信号进行信号合成处理,得到解码出的音频信号。

[0068] 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、存储、处理和共享的一种托管技术。

[0069] 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。

[0070] 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可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表、车载终端等,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。

[0071] 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处理方法。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;也可以是小程序,即只需要下载到浏览器环境中就可以运行的程序;还可以是能够嵌入至任意APP中的小程序。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。

[0072] 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处理方法所相关的数据(例如音频处理的逻辑、码流)可保存于区块链上。

[0073] 参见图3,图3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3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。但是为了清楚说明起见,在图3中将各种总线都标为总线系统550。

[0074] 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。

[0075] 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。

[0076] 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硬盘,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的一个或多个存储设备。

[0077] 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。

[0078] 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。

[0079] 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;

[0080] 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;

[0081] 呈现模块553,用于经由一个或多个与用户接口540相关联的输出装置541(例如,显示屏、扬声器等)使得能够呈现信息(例如,用于操作外围设备和显示内容和信息的用户接口);

[0082] 输入处理模块554,用于对一个或多个来自一个或多个输入装置542之一的一个或多个用户输入或互动进行检测以及翻译所检测的输入或互动。

[0083] 在一些实施例中,本申请实施例提供的音频处理装置可以采用软件方式实现,图3示出了存储在存储器550中的音频处理装置555,其可以是程序和插件等形式的软件,包括以下软件模块:分解模块5551、压缩模块5552、编码模块5553,或解码模块5554,解压模块5555、合成模块5556,其中,分解模块5551、压缩模块5552、编码模块5553用于实现音频编码功能,解码模块5554,解压模块5555、合成模块5556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。

[0084] 如前所述,本申请实施例提供的音频处理方法可以由各种类型的电子设备实施。参见图4,图4是本申请实施例提供的音频处理方法的流程示意图,通过音频处理方式实现音频编码功能,下面结合图4示出的步骤进行说明。

[0085] 在步骤101中,对音频信号进行多通道信号分解处理,得到音频信号的N个子带信号,其中,N为大于2的整数,N个子带信号的频段依次递增。

[0086] 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(又称输入信号)。

[0087] 在获取音频信号后,通过QMF分析滤波器将音频信号分解成多个子带信号,由于子带信号中偏向低频的子带信号对音频编码的影响更大,因此后续对子带信号进行差异化的信号处理。

[0088] 在一些实施例中,多通道信号分解处理是通过多层二通道子带分解实现的;对音频信号进行多通道信号分解处理,得到音频信号的N个子带信号,包括:对音频信号进行第一层的二通道子带分解处理,得到第一层的低频子带信号以及第一层的高频子带信号;对第i层的子带信号进行第i+1层的二通道子带分解处理,得到第i+1层的低频子带信号以及i+1层的高频子带信号;其中,第i层的子带信号为第i层的低频子带信号,或第i层的高频子带信号和第i层的低频子带信号,i为递增的自然数且取值范围为 $1 \leq i < N$;将最后一层的子带信号以及每一层中未进行二通道子带分解处理的高频子带信号,作为音频信号子带信号。

[0089] 其中,子带信号包括对音频信号采样得到的多个样本点。图8B所示,音频信号经过迭代两层的2通道QMF分析滤波器,即对音频信号进行迭代两层二通道子带分解,可以获得4通道的子带信号($x_k(n), n=1,2,3,4$),即对音频信号进行第1层的二通道子带分解处理,得到第1层的低频子带信号以及第1层的高频子带信号;对第1层的低频子带信号进行第2层的二通道子带分解处理,得到第1层的低频子带信号在第2层的低频子带信号 $x_1(n)$ 以及第1层的低频子带信号在2层的高频子带信号 $x_2(n)$;对第1层的高频子带信号进行第2层的二通道子带分解处理,得到第1层的高频子带信号在第2层的低频子带信号 $x_3(n)$ 以及第1层的高频子带信号在2层的高频子带信号 $x_4(n)$,从而得到4通道的子带信号 $x_k(n), n=1,2,3,4$ 。

[0090] 图8C所示,音频信号经过迭代两层的2通道QMF分析滤波器,即对音频信号进行迭代两层二通道子带分解,可以获得3通道的子带信号($x_{2,1}(n)、x_{2,2}(n)、x_{1,2}(n)$),即对音频信号进行第1层的二通道子带分解处理,得到第1层的低频子带信号以及第1层的高频子带信号 $x_{1,2}(n)$;对第1层的低频子带信号进行第2层的二通道子带分解处理,得到第1层的低频子带信号在第2层的低频子带信号 $x_{2,1}(n)$ 以及第1层的低频子带信号在2层的高频子带信号 $x_{2,2}(n)$;不对第1层的高频子带信号 $x_{1,2}(n)$ 进行二通道子带分解处理,从而得到3通道的子带信号 $x_{2,1}(n)、x_{2,2}(n)、x_{1,2}(n)$ 。

[0091] 在一些实施例中,对音频信号进行第一层的二通道子带分解处理,得到第一层的低频子带信号以及第一层的高频子带信号,包括:对音频信号进行采样处理,得到采样信号,其中,采样信号包括采样得到的多个样本点;对采样信号进行第一层的低通滤波处理,得到第一层的低通滤波信号;对第一层的低通滤波信号进行下采样处理,得到第一层的低频子带信号;对采样信号进行第一层的高通滤波处理,得到第一层的高通滤波信号;对第一层的高通滤波信号进行下采样处理,得到第一层的高频子带信号。

[0092] 需要说明的是音频信号为连续的模拟信号,采样信号为离散的数字信号,采样点为从音频信号中采样得到的采样值。

[0093] 作为示例,以音频信号为采样率 $F_s=32000\text{Hz}$ 的输入信号为例,对音频信号进行采样,得到包括640个样本点的采样信号 $x(n)$ 。调用QMF滤波器组中的分析滤波器(2通道),对采样信号进行低通滤波处理,得到低通滤波信号,对采样信号进行高通滤波处理,得到高通

滤波信号,对低通滤波信号进行下采样处理,得到第一层的低频子带信号 $x_{LB}(n)$,对高通滤波信号进行下采样处理,得到第一层的高频子带信号。 $x_{LB}(n)$ 和 $x_{HB}(n)$ 的有效带宽分别是0-8kHz和8-16kHz, $x_{LB}(n)$ 和 $x_{HB}(n)$ 的样本点数为320。

[0094] 需要说明的是,QMF滤波器组是一个包含分析-合成的滤波器对。对于QMF分析滤波器,可以将输入的采样率为 F_s 的信号分解成两路采样率为 $F_s/2$ 的信号,分别表示QMF低通信号和QMF高通信号。解码端恢复出的低通信号和高通信号后,经过QMF合成滤波器进行合成处理,即可以恢复出输入信号对应的采样率 F_s 的重建信号。

[0095] 在一些实施例中,对音频信号进行多通道信号分解处理,得到音频信号的 N 个子带信号,包括:对音频信号进行采样处理,得到采样信号,其中,采样信号包括采样得到的多个样本点;对采样信号进行第 j 个通道的滤波处理,得到第 j 个滤波信号;对第 j 个滤波信号进行下采样处理,得到音频信号的第 j 个子带信号;其中, j 为递增的自然数且取值范围为 $1 \leq j \leq N$ 。

[0096] 例如,可以预先将QMF分析滤波器组分为多通道,通过第 j 个通道的滤波器对采样信号进行第 j 个通道的滤波处理,得到第 j 个滤波信号;对第 j 个滤波信号进行下采样处理,得到音频信号的第 j 个子带信号。

[0097] 在步骤102中,对每个子带信号进行信号压缩处理,得到每个子带信号的特征。

[0098] 其中,每个子带信号的特征的特征维度与每个子带信号的频段非正相关,第 N 个子带信号的特征的特征维度小于第一个子带信号的特征的特征维度,其中,非正相关表示子带信号的特征的特征维度随子带信号的频段的增加而递减或者保持不变,即某子带信号的特征的特征维度小于或者等于前一个子带信号的特征的特征维度。通过信号压缩(即通道分析)可以对子带信号进行数据压缩,以减小子带信号的数据量,即子带信号的特征的维度小于子带信号的维度。

[0099] 例如,由于频率越低的子带信号对音频编码的影响越大,因此对子带信号进行差异化的信号处理,使得偏向高频的子带信号的特征的特征维度越低。

[0100] 在一些实施例中,对每个子带信号进行信号压缩处理,得到每个子带信号的特征,包括:针对任意子带信号执行以下处理:调用子带信号对应的第一神经网络模型;通过第一神经网络模型对子带信号进行特征提取处理,得到子带信号的特征;其中,第一神经网络模型的结构复杂度与子带信号的特征的维度正相关。

[0101] 例如,通过第一神经网络模型对子带信号进行特征提取处理,得到子带信号特征,以在保证子带信号特征的完备性的情况下,尽量减小子带信号特征的特征维度。其中,本申请实施例并不局限于第一神经网络模型的结构。

[0102] 在一些实施例中,通过第一神经网络模型对子带信号进行特征提取处理,得到子带信号的特征,包括:通过第一神经网络模型对子带信号执行以下处理:对子带信号进行卷积处理,得到子带信号的卷积特征;对卷积特征进行池化处理,得到子带信号的池化特征;对池化特征进行下采样处理,得到子带信号的下采样特征;对下采样特征进行卷积处理,得到子带信号的特征。

[0103] 如图11所示,基于子带信号 $x_1(n)$ 调用第1通道的神经网络模型,生成更低维度的特征向量 $F_1(n)$,即子带信号特征。首先,通过因果卷积对输入的子带信号 $x_1(n)$ 进行卷积处

理,得到 24×160 的卷积特征。然后,对 24×160 的卷积特征进行因子为2的池化处理(即预处理),得到 24×80 的池化特征。接下来,对 24×80 的池化特征进行下采样处理,得到 192×1 的下采样特征。最后,对 192×1 的下采样特征,再经过因果卷积进行卷积处理,可以得到一个32维的特征向量 $F_1(n)$ 。

[0104] 在一些实施例中,下采样处理是通过多个级联的编码层实现的;对池化特征进行下采样处理,得到子带信号的下采样特征,包括: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作为子带信号的下采样特征。

[0105] 如图11所示,级联3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)的编码块对 24×80 的池化特征进行下采样处理,得到 48×40 的下采样结果,然后通过下采样因子(Down_factor=5)的编码块对 48×40 的下采样结果进行下采样处理,得到 96×8 的下采样结果,最后通过下采样因子(Down_factor=8)的编码块对 96×8 的下采样结果进行下采样处理,得到 192×1 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。

[0106] 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子带信号的下采样特征。通过级联形式的编码层,能够获取精度渐进的低频子带信号的下采样特征。

[0107] 在一些实施例中,对每个子带信号进行信号压缩处理,得到每个子带信号的子带信号特征,包括:对前k个子带信号分别进行特征提取处理,得到前k个子带信号分别对应的子带信号特征;对后N-k个子带信号分别进行频带扩展处理,得到后N-k个子带信号分别对应的子带信号特征;其中,k为整数且取值范围为 $1 < k < N$ 。

[0108] 其中,k为2的倍数,由于频率越低的子带信号对音频编码的影响更大,因此对子带信号进行差异化的信号处理,对频率越高的子带信号的压缩力度更大,即通过另一种方法压缩子带信号,即频带扩展(从频带受限的窄带语音信号中恢复宽带语音信号),以快速压缩子带信号,提取子带信号的高频特征。其中,高频扩展处理用于对子带信号进行降维,实现数据压缩的功能。

[0109] 作为示例,调用QMF分析滤波器(2通道QMF),进行下采样。如图8C所示,通过QMF分析滤波器实现3通道分解,最终可以获得3个子带信号 $x_{\text{HB}}(n)$ 、 $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 。如图8C所示, $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 分别是0-4kHz、4-8kHz频谱,是经过两次迭代的2通道QMF分析滤波生成的,等效于第一种实现方式中的 $x_1(n)$ 和 $x_2(n)$ 。这两个子带信号 $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 分别包含160个样本点。如图8C所示,对应于8-16kHz频段,无需做精细分析。因此,只需对原32kHz采样的输入信号做一次QMF高通滤波即可生成一路高频子带信号 $x_{\text{HB}}(n)$,每帧包含320样本点。

[0110] 对于 $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 对应的两种子带信号可以调用神经网络模型(图11的第一通道和第二通道)进行特征提取,结果上,生成了对应子带信号的特征向量 $F_1(n)$ 和 $F_2(n)$,维度分别是32维和16维。对于包括320点的高频子带信号 $x_{\text{HB}}(n)$ 通过频带扩展,生成了对应子带信号的特征向量 $F_{\text{HB}}(n)$ 。

[0111] 在一些实施例中,对后N-k个子带信号分别进行频带扩展处理,得到后N-k个子带

信号分别对应的子带信号特征,包括:针对后N-k个子带信号中的任意子带信号执行以下处理:基于子带信号包括的多个样本点进行频域变换处理,得到多个样本点分别对应的变换系数;将多个样本点分别对应的变换系数划分为多个子带;对每个子带包括的变换系数进行均值处理,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;将多个子带分别对应的子带谱包络确定为子带信号对应的子带信号特征。

[0112] 需要说明的是,本申请实施例的频域变换方法包括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。本申请实施例的均值处理包括算术平均、几何平均,本申请实施例并不局限于均值处理的方式。

[0113] 在一些实施例中,基于子带信号包括的多个样本点进行频域变换处理,得到多个样本点分别对应的变换系数,包括:获取参考音频信号的参考子带信号,其中,参考音频信号是与音频信号相邻的音频信号,参考子带信号与子带信号的频段相同;基于参考子带信号包括的多个样本点以及子带信号包括的多个样本点,对子带信号包括的多个样本点进行离散余弦变换处理,得到子带信号包括的多个样本点分别对应的变换系数。

[0114] 在一些实施例中,对每个子带包括的变换系数进行几何均值处理的过程如下:确定每个子带包括的样本点对应的变换系数的平方和;将平方和与子带包括的样本点的数量的比值,确定为得到每个子带对应的平均能量。

[0115] 作为示例,对于包括320点的高频子带信号 $x_{\text{HB}}(n)$,调用改进型离散余弦变换(MDCT, Modified Discrete Cosine Transform),生成320个点的MDCT系数(即高频子带信号包括的多个样本点分别对应的变换系数)。具体地,如果是50%交叠,可以将第n+1帧高频数据(即参考音频信号)与第n帧高频数据(即音频信号)合并(拼接),计算640个点的MDCT,获得320个点的MDCT系数。

[0116] 将320个点的MDCT系数分成N个子带(即将多个样本点分别对应的变换系数划分为多个子带),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,本申请实施例不可以对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)。

[0117] 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。

[0118] 对于每一个子带,计算当前子带中所有MDCT系数的平均能量(即对每个子带包括的变换系数进行均值处理)作为子带谱包络(频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 $x(n)$, $n=1, 2, \dots, 40$,则通过几何均值计算平均能量 $Y = ((x(1)^2 + x(2)^2 + \dots + x(40)^2) / 40)$ 。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子带信号的特征向量 $F_{\text{HB}}(n)$,即高频特征。

[0119] 在步骤103中,对每个子带信号的子带信号特征进行量化编码处理,得到每个子带信号的码流。

[0120] 例如,对子带信号进行差异化的信号处理,使得偏向高频的子带信号的子带信号特征的特征维度越低,对减小了特征维度的子带信号特征进行量化编码,并传输码流至解码端,通过解码端对码流进行解码,以恢复出音频信号,从而在保证音频质量的情况下,提高音频编码效率。

[0121] 在一些实施例中,对每个子带信号的子带信号特征进行量化编码处理,得到每个子带信号的码流,包括:对每个子带信号的子带信号特征进行量化处理,得到子带信号特征的索引值;对子带信号特征的索引值进行熵编码处理,得到子带信号的码流。

[0122] 例如,对于子带信号的子带信号特征,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,将编码得到的码流传输至解码端,通过解码端对码流进行解码。

[0123] 如前所述,本申请实施例提供的音频处理方法可以由各种类型的电子设备实施。参见图5,图5是本申请实施例提供的音频处理方法的流程示意图,通过音频处理方式实现音频解码功能,下面结合图5示出的步骤进行说明。

[0124] 在步骤201中,对N个码流进行量化解码处理,得到每个码流对应的子带信号特征,其中,N为大于2的整数,N个码流是对音频信号经过多通道信号分解处理后得到的N个子带信号分别进行编码得到的。

[0125] 例如,通过如图4所示的音频处理方法编码得到子带信号的码流后,将编码得到的子带信号的码流传输至解码端,解码端接收到子带信号的码流后,对子带信号的码流进行量化解码处理,得到码流对应的子带信号特征。

[0126] 需要说明的是,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,并通过查量化表(即逆量化,量化表为编码过程中量化所产生的映射表),获得子带信号特征。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子带信号特征相对于编码过程中的子带信号特征是一种估计值。

[0127] 例如,对N个码流进行量化解码处理,得到每个码流对应的子带信号特征,包括:对N个码流中的任意码流执行以下处理:对码流进行熵解码处理,得到码流对应的索引值;对码流对应的索引值进行逆量化处理,得到码流对应的子带信号特征。

[0128] 在步骤202中,对每个子带信号特征进行信号解压处理,得到每个子带信号特征对应的估计子带信号。

[0129] 例如,信号解压(又称通道合成)是信号压缩的逆过程,对子带信号特征进行信号解压处理,得到每个子带信号特征对应的估计子带信号(一种估计值)。

[0130] 在一些实施例中,对每个子带信号特征进行信号解压处理,得到每个子带信号特征对应的估计子带信号,包括:针对任意子带信号特征执行以下处理:调用子带信号特征对应的第二神经网络模型;通过第二神经网络模型对子带信号特征进行特征重建,得到子带信号特征对应的估计子带信号;其中,第二神经网络模型的结构复杂度与子带信号特征的维度正相关。

[0131] 例如,当编码端是对所有子带信号进行特征提取,获得子带信号特征时,则解码端对子带信号特征进行特征重建处理,得到高频特征对应的高频子带信号。

[0132] 作为示例,当解码端接收到4个码流,对4个码流进行量化解码处理,得到4个通道的子带信号向量(即特征向量)的估计值 $F'_k(n)$, $k=1,2,3,4$,基于特征向量的估计值 $F'_k(n)$, $k=1,2,3,4$,调用深度神经网络(如图12所示),生成子带信号的估计值 $x'_k(n)$, $k=1,2,3,4$,即估计子带信号。

[0133] 如图12所示,信号压缩的网络结构的流程图与信号解压的网络结构类似,比如因果卷积、信号解压的网络结构中的后处理结构与信号压缩的网络结构中的预处理类似。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。

[0134] 在一些实施例中,通过第二神经网络模型对子带信号特征进行特征重建,得到子带信号特征对应的估计子带信号,包括:通过第二神经网络模型对子带信号特征执行以下处理:对子带信号特征进行卷积处理,得到子带信号特征的卷积特征;对卷积特征进行上采样处理,得到子带信号特征的上采样特征;对上采样特征进行池化处理,得到子带信号特征的池化特征;对池化特征进行卷积处理,得到子带信号特征对应的估计子带信号。

[0135] 如图12所示,基于子带信号特征 $F'_1(n)$,调用如图12所示的神经网络模型(第1通道),生成低频子带信号 $x'_1(n)$ 。首先,通过因果卷积对输入的低频特征向量 $F'_1(n)$ 进行卷积处理,得到 192×1 的卷积特征。然后,对 192×1 的卷积特征进行上采样处理,得到 24×80 的上采样特征。接下来,对 24×80 的上采样特征进行池化处理(即后处理),得到 24×160 的池化特征。最后,对池化特征,再经过因果卷积进行卷积处理,可以得到一个160维的子带信号 $x'_1(n)$ 。

[0136] 在一些实施例中,上采样处理是通过多个级联的解码层实现的;对卷积特征进行上采样处理,得到低子带信号特征的上采样特征,包括: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作为子带信号特征的上采样特征。

[0137] 如图12所示,级联3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($Up_factor=8$)的解码块对 192×1 的卷积特征进行上采样处理,得到 96×8 的上采样结果,然后通过上采样因子($Up_factor=5$)的解码块对 96×8 的上采样结果进行上采样处理,得到 48×40 的上采样结果,最后通过上采样因子($Up_factor=4$)的解码块对 48×40 的上采样结果进行上采样处理,得到 24×80 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。

[0138] 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上采样特征。通过级联形式的解码层,能够获取精度渐进的上采样特征。

[0139] 在一些实施例中,对每个子带信号特征进行信号解压处理,得到每个子带信号特征对应的估计子带信号,包括:对前 k 个子带信号特征分别进行特征重建处理,得到前 k 个子带信号特征分别对应的估计子带信号;对后 $N-k$ 个子带信号特征分别进行频带扩展的逆处

理,得到后 $N-k$ 个子带信号特征分别对应的估计子带信号;其中, k 为整数且取值范围为 $1 < k < N$ 。

[0140] 例如,当编码端是对前 k 个子带信号进行特征提取,获得子带信号特征,对后 $N-k$ 个子带信号进行频带扩展处理时,则解码端对前 k 个子带信号特征分别进行特征重建处理,得到前 k 个子带信号特征分别对应的估计子带信号;对后 $N-k$ 个子带信号特征分别进行频带扩展的逆处理,得到后 $N-k$ 个子带信号特征分别对应的估计子带信号。

[0141] 在一些实施例中,对后 $N-k$ 个子带信号特征分别进行频带扩展的逆处理,得到后 $N-k$ 个子带信号特征分别对应的估计子带信号,包括:针对后 $N-k$ 个子带信号特征中的任意子带信号特征执行以下处理:对前 k 个估计子带信号中与子带信号特征关联的估计子带信号进行信号合成处理,得到与子带信号特征对应的低频子带信号;基于低频子带信号包括的多个样本点进行频域变换处理,得到多个样本点分别对应的变换系数;对多个样本点分别对应的变换系数中的后半部分的变换系数进行频谱复制处理,得到参考子带信号的参考变换系数;基于子带信号特征对应的子带谱包络,对参考子带信号的参考变换系数进行增益处理,得到增益后的参考变换系数;对增益后的参考变换系数进行反频域变换处理,得到子带信号特征对应的估计子带信号。

[0142] 例如,当前 k 个估计子带信号中与子带信号特征关联的估计子带信号为子带信号特征对应的下一层估计子带信号时,对前 k 个估计子带信号中与子带信号特征关联的估计子带信号进行信号合成处理,得到与子带信号特征对应的低频子带信号,即当编码端对音频信号进行多层二通道子带分解,生成对应层次的子带信号,并对对应层次的子带信号进行信号压缩,得到对应的子带信号特征时,则当与子带信号特征关联的估计子带信号,需要对前 k 个估计子带信号中与子带信号特征关联的估计子带信号进行信号合成处理,得到与子带信号特征对应的低频子带信号,以使低频子带信号、与子带信号特征处于同一层次。

[0143] 当低频子带信号、与子带信号特征处于同一层次后,基于低频子带信号包括的多个样本点进行频域变换处理,得到多个样本点分别对应的变换系数;对多个样本点分别对应的变换系数中的后半部分的变换系数进行频谱复制处理,得到参考子带信号的参考变换系数;基于子带信号特征对应的子带谱包络,对参考子带信号的参考变换系数进行增益处理,得到增益后的参考变换系数;对增益后的参考变换系数进行反频域变换处理,得到子带信号特征对应的估计子带信号。

[0144] 需要说明的是,本申请实施例的频域变换方法包括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。

[0145] 在一些实施例中,基于子带信号特征对应的子带谱包络,对参考子带信号的参考变换系数进行增益处理,得到增益后的参考变换系数,包括:基于子带信号特征对应的子带谱包络,将参考子带信号的参考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第一平均能量,并确定子带对应的第二平均能量;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变换系数相乘,得到增益后的参考变换系数。

[0146] 作为示例,对于收到码流,先进行熵解码,并通过查量化表,获得3个通道的特征向

量 $F'_k(n)$, $k=1, 2$ 和 $F'_{HB}(n)$, 即子带信号特征, $F'_k(n)$, $k=1, 2$ 和 $F'_{HB}(n)$ 是基于图8C所示的二叉树形式排列的, 其中, $F'_k(n)$, $k=1, 2$ 为 $F'_{HB}(n)$ 处于下一层次。基于解码获得特征向量 $F'_k(n)$, $k=1, 2$ 参考图12中的第一通道和第二通道, 获得两个子带信号的估计值 $x'_{2,1}(n)$ 和 $x'_{2,2}(n)$, $x'_{2,1}(n)$ 和 $x'_{2,2}(n)$ 的维度均为160, $x'_{2,1}(n)$ 和 $x'_{2,2}(n)$ 处于 $F'_{HB}(n)$ 的下一层次。

[0147] 基于 $x'_{2,1}(n)$ 和 $x'_{2,2}(n)$, 调用一次2通道QMF合成滤波, 可以生成对应于0-8kHz的低频子带信号的估计值 $x'_{LB}(n)$, 简称低频子带信号, 维度为320维, 低频子带信号 $x'_{LB}(n)$ 、与子带信号特征 $F'_{HB}(n)$ 处于同一层次。 $x'_{LB}(n)$ 是为了进行后续8-16kHz的频带扩展。

[0148] 对于8-16kHz的频带扩展过程, 是基于码流中解码出来的8个子带谱包络(即 $F'_{HB}(n)$)和解码端本地生成0-8kHz的低频子带信号的估计值 $x'_{LB}(n)$ 实现的。具体频带扩展过程的逆过程如下:

[0149] 先将解码端生成的低频子带信号 $x'_{LB}(n)$, 也进行编码端类似的640点的MDCT变换, 生成320个点的MDCT系数(即低频部分的MDCT系数), 即基于低频子带信号包括的多个样本点进行频域变换处理, 得到多个样本点分别对应的变换系数。

[0150] 然后, 将由 $x'_{LB}(n)$ 生成的320个点的MDCT系数进行复制, 生成高频部分的MDCT系数(即参考子带信号的参考变换系数)。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部分的MDCT谱包含过多谐波, 可以将低频子带依赖的320个点的MDCT系数中的后160点作为母版, 频谱拷贝2次, 生成320个点的参考子带信号的MDCT系数的参考值(即参考子带信号的参考变换系数), 即对多个样本点分别对应的变换系数中的后半部分的变换系数进行频谱复制处理, 得到参考子带信号的参考变换系数。

[0151] 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8个子带谱包络, 即子带信号特征对应的子带谱包络 $F'_{HB}(n)$), 这8个子带谱包络对应8个高频子带, 并将生成的320个点的参考子带信号的MDCT系数的参考值分为8个参考子带(即基于子带信号特征对应的子带谱包络, 将参考子带信号的参考变换系数划分为多个子带), 分带地, 基于一个高频子带与对应的参考子带, 对生成的320点的参考子带信号的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(即第一平均能量)与对应的参考子带信号的平均能量(第二平均能量)计算增益因子, 将该对应的参考子带信号中的每个点对应的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。

[0152] 例如, 假定复制过去生成的参考子带信号的参考子带的平均能量是 Y_L , 当前要做增益控制的高频子带(即基于码流解码出来的子带谱包络对应的高频子带)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, 直接将复制生成的参考子带信号中的每个点均乘以 a 。

[0153] 最后, 调用MDCT反变换, 生成子带信号的估计值 $x'_{HB}(n)$ (即子带信号特征 $F'_{HB}(n)$ 对应的估计子带信号)。对增益后的320点的MDCT系数进行MDCT反变换, 生成640个点的估计值, 通过交叠, 取有效的前320个点的估计值作为 $x'_{HB}(n)$ 。

[0154] 需要说明的是, 当前 k 个子带信号特征和后 $N-k$ 个子带信号特征处于同一层次时, 可基于前 k 个子带信号特征分别对应的估计子带信号直接对后 $N-k$ 个子带信号特征进行频带扩展。

[0155] 作为示例,对于收到码流,先进行熵解码,并通过查量化表,获得4个通道的特征向量 $F'_k(n)$, $k=1,2,3,4$,即子带信号特征, $F'_k(n)$, $k=1,2,3,4$ 是基于图8B所示的二叉树形式排列的,其中, $F'_k(n)$, $k=1,2,3,4$ 处于同一层次。基于解码获得特征向量 $F'_k(n)$, $k=1,2$ 参考图12中的第一通道和第二通道,获得两个子带信号 $x'_1(n)$ 和 $x'_2(n)$ 。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部分的MDCT谱包含过多谐波,可以选择 $x'_2(n)$ 对 $F'_k(n)$, $k=3,4$ 进行频带扩展。对于 $F'_k(n)$, $k=3,4$ 的频带扩展的逆过程,是基于码流中解码出来的8个子带谱包络(即 $F'_k(n)$, $k=3,4$)和 $x'_2(n)$ 实现的,具体频带扩展的逆过程与上述频带扩展的逆过程类似。

[0156] 在步骤203中,对多个估计子带信号进行信号合成处理,得到多个码流对应的合成音频信号。

[0157] 例如,信号合成处理是信号分解处理的逆过程,解码端对多个估计子带信号进行子带合成处理,以恢复出音频信号,其中合成音频信号即为恢复出的音频信号。

[0158] 在一些实施例中,对多个估计子带信号进行信号合成处理,得到多个码流对应的合成音频信号,包括:对多个估计子带信号分别进行上采样处理,得到多个估计子带信号分别对应的滤波信号;对多个滤波信号进行滤波合成处理,得到多个码流对应的合成音频信号。

[0159] 例如,在获取多个估计子带信号后,通过QMF合成滤波器对多个估计子带信号进行子带合成以恢复出音频信号。

[0160] 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。

[0161] 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:

[0162] 相关技术中,语音编码根据编码原理大致如下: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。

[0163] 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。本申请实施例提供基于多通道信号分解和神经网络的语音编码方法(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,分解成多个子带信号,例如多个子带信号中包括相对低采样率的子带信号和相对高采样率的子带信号,不同的子带信号可以采用不同的数据压缩机制进行压缩。对于重要的部分(相对低采样率的子带信号),基于神经网络(NN,Neural Network)技术进行处理后,会得到比输入的子带信号更低维度的特征向量。对于相对不重要的部分(相对高采样率的子带信号),使用更少的比特进行编码。

[0164] 本申请实施例可应用于如图6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601,解码器部署在下行客户端602,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602,通过下行客户端602进行解码、增强等处理,以在下行客户端602回放解码出的语音。

[0165] 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。

[0166] 下面结合图7说明本申请实施例提供的基于多通道信号分解和神经网络的语音编码方法:

[0167] 对于编码端执行如下处理:

[0168] 对第n帧的输入语音信号 $x(n)$,使用多通道分析滤波器分解为低为N个子带信号,例如输入信号后,经过多通道QMF分解后,可以获得N个子带信号 $x_k(n)$, $k=1,2,\dots,N$ 。

[0169] 对于第k个子带信号 $x_k(n)$,调用第k个通道分析,获得低维度的特征向量 $F_k(n)$,特征向量 $F_k(n)$ 的维度小于子带信号 $x_k(n)$ 的维度,以减少数据量。例如,对于每帧 $x_k(n)$,调用空洞卷积网络(Dilated CNN),生成更低维的特征向量 $F_k(n)$ 。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。

[0170] 对于相对高采样率的子带信号,考虑到高采样率的子带信号对质量的重要性没有低频那么重要,高采样率的子带信号还可以采用其它方案提取特征向量。例如,基于语音信号分析的频带扩展技术,可以只用1-2kbps码率实现高频子带信号的生成。

[0171] 对子带对应的特征向量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。

[0172] 对于解码端执行如下处理:

[0173] 对解码端接收到码流进行解码,分别获得各通道的特征向量的估计值 $F'_k(n)$, $k=1,2,\dots,N$ 。

[0174] 对于通道k(即 $F'_k(n)$)进行第k通道合成,生成子带信号的估计值 $x'_k(n)$ 。

[0175] 最后,调用QMF合成滤波,生成重建的语音信号 $x'(n)$ 。

[0176] 下面在具体介绍本申请实施例提供的基于多通道信号分解和神经网络的语音编码方法之前,先介绍QMF滤波器组、空洞卷积网络以及频带扩展技术。

[0177] QMF滤波器组是一个包含分析-合成的滤波器对。对于QMF分析滤波器,可以将输入的采样率为 F_s 的信号分解成N路采样率为 F_s/N 的子带信号。如图8A所示的2通道QMF滤波器的低通部分 $H_{Low}(z)$ (相当于 $H_1(z)$)和高通部分 $H_{High}(z)$ (相当于 $H_2(z)$)的频谱响应。基于QMF分析滤波器组的相关理论知识,可以容易地描述上述低通滤波和高通滤波的系数之间的相关性,如公式(1)所示:

$$[0178] \quad h_{High}(k) = -1^k h_{Low}(k) \quad (1)$$

[0179] 其中, $h_{Low}(k)$ 表示低通滤波的系数, $h_{High}(k)$ 表示高通滤波的系数。

[0180] 类似地,根据QMF相关理论,可以基于QMF分析滤波器组 $H_{Low}(z)$ 和 $H_{High}(z)$,描述QMF合成滤波器组,如公式(2)所示。

$$[0181] \quad G_{Low}(z) = H_{Low}(z)$$

$$[0182] \quad G_{High}(z) = (-1) * H_{High}(z) \quad (2)$$

[0183] 其中, $G_{Low}(z)$ 表示恢复出的低通信号, $G_{High}(z)$ 表示恢复出的高通信号。

[0184] 解码端恢复出的低通和高通信号,经过QMF合成滤波器组进行合成处理,即可以恢复出输入信号对应的采样率 F_s 的重建信号。

[0185] 此外,基于上述2通道QMF方案,还可以扩展为N通道QMF的方案,例如可以使用二叉数的方式,迭代地对当前的子带信号做2通道QMF分析,以获得更低分辨率的子带信号。图8B表示经过迭代两层的2通道QMF分析滤波器,可以获得4通道的子带信号。图8C是另一种实现方式,考虑到高频部分信号对质量影响小,无需那么高精度的分析,因此,只需要对原始信号做一次高通滤波即可。类似地,可以实现更多通道的方式,比如,8、16、32通道,在此不进一步展开。

[0186] 参见图9A和图9B,图9A是本申请实施例提供的普通卷积网络的示意图,图9B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9A和图9B中示出的卷积核大小(Kernel Size)均为 3×3 ;但是,图9A所示的普通卷积的感受野 901 只有3,而图9B所示的空洞卷积的感受野 902 达到了5。也就是说,对于尺寸为 3×3 的卷积核,图9A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9B所示的空洞卷积的感受野为5、扩张率为2。

[0187] 卷积核还可以在类似图9A或者图9B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。

[0188] 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。

[0189] 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。

[0190] 如图10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10示出的频域实现方案具体包括:

[0191] 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。

[0192] 下面对本申请实施例提供的基于子带分解和神经网络的语音编解码方法进行具体说明。

[0193] 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、32000Hz、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。

[0194] 下面参考图7所示的流程图,通过两种实现方式对编码端和解码端进行详细描述。

[0195] 第一种实现方式如下:

[0196] 关于编码端的流程如下:

[0197] 1、生成输入信号。

[0198] 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,记为输入信号 $x(n)$ 。

[0199] 2、QMF信号分解。

[0200] 调用QMF分析滤波器(2通道QMF),进行下采样。如图8B所示,通过QMF分析滤波器实现4通道分解,最终可以获得4个子带信号 $x_k(n)$, $n=1,2,3,4$ 。子带信号的有效带宽分别是0-4kHz、4-8kHz、8-12kHz和12-16kHz,每帧子带信号的样本点数为160。

[0201] 3、对子带信号 $x_k(n)$ 进行第 k 通道分析。

[0202] 对于任意一个通道分析,就是通过调用深度网络(即神经网络)分析子带信号 $x_k(n)$,生成更低维度的特征向量 $F_k(n)$ 。在本申请实施例中,对于4通道QMF滤波器, $x_k(n)$ 的维度为160,输出的子带信号的特征向量可以根据所属通道分别设置,从数据量看,通道分析起到了“降维”的作用,实现数据压缩的功能。

[0203] 参见图11所示的通道分析的网络结构图,下面以第1通道为例,具体说明通道分析的流程:

[0204] 首先,调用一个24通道的因果卷积,可以将输入的张量(即向量),扩展为 24×320 的张量。

[0205] 然后,对 24×320 的张量进行预处理。例如,对 24×320 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 24 乘以 160 的张量。

[0206] 接下来,级联3个不同下采样因子(Down_factor)的编码块。以编码块(Down_factor=4)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展率。然后,3个编码块的Down_factor分别设置为4、5、8,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,3个编码块通道数分别设置为48、96、192。因此,经过3个编码块, 24×160 的张量依次转换成 48×40 、 96×8 和 192×1 的张量。

[0207] 最后,对 192×1 的张量,再经过类似预处理的因果卷积,可以输出一个32维的特征向量。

[0208] 如图11所示,越是高频对应的通道的网络结果,相对更简单。这体现在编码块的通道数和输出的特征向量的维度。这主要是考虑了高频部分相对不影响质量,无需向第1通道那样,提取高精度和高维度的子带信号的特征向量。

[0209] 如图11所示,经过4通道分析,获得了4个通道对应的子带信号的特征向量 $\{F_1(n), F_2(n), F_3(n), F_4(n)\}$,维度分别是32维、16维、16维和8维。由此可见,原输入信号为640维,输出的所有特征向量的维度和只有72维,数据量显著减少。

[0210] 4、量化编码。

[0211] 对于4个通道的特征向量,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。

[0212] 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。

[0213] 关于解码端的流程如下:

[0214] 1、量化解码。

[0215] 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,并通过查量化表,获得4个通道的特征向量的估计值 $F'_k(n)$, $k=1,2,3,4$ 。

[0216] 2、对特征向量的估计值 $F'_k(n)$ 进行第k通道合成。

[0217] 第k通道合成的目的是基于特征向量的估计值 $F'_k(n)$, $k=1,2,3,4$,调用深度神经网络(如图12所示),生成子带信号的估计值 $x'_k(n)$, $k=1,2,3,4$ 。

[0218] 如图12所示,通道合成的网络结构的流程图与分析网络的网络结构类似,比如因果卷积、通道合成的网络结构中的后处理结构与分析网络的网络结构中的预处理类似。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。

[0219] 3、合成滤波器

[0220] 在解码端获得了4个通道的子带信号的估计值 $x'_k(n)$, $k=1,2,3,4$ 之后,只需要调用4通道QMF合成滤波器(如图8B所示),就可以生成640个点的重建信号 $x'(n)$ 。

[0221] 第二种实现方式如下:

[0222] 需要说明的是,第二种实现方式主要是为了简化高频相关的两个子带信号的压缩流程。如上所述,第一种实现方式的第3通道和第4通道对应于8-16kHz(8-12kHz和12-16kHz),包含了24维特征向量(16维特征向量和8维特征向量)。根据语音信号的基本特征,8-16kHz完全可以使用更为精简的频带扩展等技术,编码端提取更少维度的特征向量。既节省比特,也节省复杂度。下面,对第二种实现方式的编码端和解码端进行详细描述。

[0223] 关于编码端的流程如下:

[0224] 1、生成输入信号。

[0225] 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,记为输入信号 $x(n)$ 。

[0226] 2、QMF信号分解。

[0227] 调用QMF分析滤波器(2通道QMF),进行下采样。如图8C所示,通过QMF分析滤波器实现3通道分解,最终可以获得3个子带信号 $x_{\text{HB}}(n)$ 、 $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 。

[0228] 如图8C所示, $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 分别是0-4kHz、4-8kHz频谱,是经过两次迭代的2通道QMF分析滤波生成的,等效于第一种实现方式中的 $x_1(n)$ 和 $x_2(n)$ 。这两个子带信号 $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 分别包含160个样本点。

[0229] 如图8C所示,对应于8-16kHz频段,无需做精细分析。因此,只需对原32kHz采样的输入信号做一次QMF高通滤波即可生成一路高频子带信号 $x_{\text{HB}}(n)$,每帧包含320样本点。

[0230] 3、对子带信号进行第k通道分析。

[0231] 根据上面介绍的等效性, $x_{2,1}(n)$ 和 $x_{2,2}(n)$ 对应的两种子带信号的分析,可以参考第一种实现方式中前两个通道(图11的第一通道和第二通道)的过程。结果上,生成了对应子带信号的特征向量 $F_1(n)$ 和 $F_2(n)$,维度分别是32维和16维。

[0232] 而对于与8-16kHz相关的子带信号(包含320个样本点/帧),则采用频带扩展方法(从频带受限的窄带语音信号中恢复宽带语音信号)。下面具体介绍本申请实施例中频带扩展的应用:

[0233] 对于包括320点的高频子带信号 $x_{\text{HB}}(n)$,调用改进型离散余弦变换(MDCT,Modified

Discrete Cosine Transform),生成320个点的MDCT系数。具体地,如果是50%交叠,可以将第n+1帧高频数据与第n帧高频数据合并(拼接),计算640个点的MDCT,获得320个点的MDCT系数。

[0234] 将320个点的MDCT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,本申请实施例不可以对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)。

[0235] 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。

[0236] 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(频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子带信号 $x_{HB}(n)$ 的特征向量 $F_{HB}(n)$ 。

[0237] 总之,使用上述两种方法(NN结构和频带扩展)中任意一种方法,均可以将320维的子带信号,输出为一个8维的特征向量。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。

[0238] 4、量化编码。

[0239] 对于上述的3个子带信号的特征向量(分别是32、16、8维),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。

[0240] 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。

[0241] 关于解码端的流程如下:

[0242] 1、量化解码。

[0243] 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,并通过查量化表,获得3个通道的特征向量的估计值 $F'_k(n)$, $k=1,2$ 和 $F'_{HB}(n)$ 。

[0244] 2、对特征向量的估计值进行通道合成。

[0245] 对于0-8kHz相关的两个通道,可以基于解码获得特征向量的估计值 $F'_k(n)$, $k=1,2$ 参考第一种实现方式中的相关步骤(参见图12中的第一通道和第二通道),获得两个子带信号的估计值 $x'_{2,1}(n)$ 和 $x'_{2,2}(n)$, $x'_{2,1}(n)$ 和 $x'_{2,2}(n)$ 的维度均为160。

[0246] 此外,基于 $x'_{2,1}(n)$ 和 $x'_{2,2}(n)$,调用一次2通道QMF合成滤波,可以生成对应于0-8kHz的子带信号的估计值 $x'_{LB}(n)$,维度为320维。 $x'_{LB}(n)$ 是为了进行后续8-16kHz的频带扩展。

[0247] 如图12所示,通道合成的网络结构的流程图与分析网络的网络结构类似,比如因

果卷积、通道合成的网络结构中的后处理结构与分析网络的网络结构中的预处理类似。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。

[0248] 对于8-16kHz的通道合成过程，是基于码流中解码出来的8个子带谱包络(即 $F'_{HB}(n)$)和解码端本地生成0-8kHz的子带信号的估计值 $x'_{LB}(n)$ 实现的。具体的通道合成过程如下：

[0249] 先将解码端生成的低频子带信号的估计值 $x'_{LB}(n)$ ，也进行编码端类似的640点的MDCT变换，生成320个点的MDCT系数(即低频部分的MDCT系数)。

[0250] 然后，将由 $x'_{LB}(n)$ 生成的320个点的MDCT系数进行复制，生成高频部分的MDCT系数。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部分的MDCT谱包含过多谐波，可以将低频子带依赖的320个点的MDCT系数中的后160点作为母版，频谱拷贝2次，生成320个点的高频子带信号的MDCT系数的参考值。

[0251] 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8个子带谱包络)，这8个子带谱包络对应8个高频子带，并将生成的320个点的高频子带信号的MDCT系数的参考值分为8个参考高频子带，分带地，基于一个高频子带与对应的参考高频子带，对生成的320点的高频子带信号的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。

[0252] 例如，假定复制过去生成的高频子带信号的高频子带的平均能量是 Y_L ，当前要做增益控制的高频子带(即基于码流解码出来的子带谱包络对应的高频子带)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，直接将复制生成的高频子带中的每个点均乘以 a 。

[0253] 最后，调用MDCT反变换，生成高频子带信号的估计值 $x'_{HB}(n)$ 。对增益后的320点的MDCT系数进行MDCT反变换，生成640个点的估计值，通过交叠，取有效的前320个点的估计值作为 $x'_{HB}(n)$ 。

[0254] 3、合成滤波器

[0255] 在解码端获得了低频子带信号的估计值 $x'_{LB}(n)$ 和高频子带信号的估计值 $x'_{HB}(n)$ 之后，只需要上采样并调用一次2通道QMF合成滤波器，就可以生成640个点的重建信号 $x'(n)$ 。

[0256] 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。

[0257] 综上，本申请实施例提供的基于多通道信号分解和神经网络的语音编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。

[0258] 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处理方法。本申请实施例还提供音频处理装置，实际应用中，音频处理装置

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3示出了存储在存储器550中的音频处理装置555,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。

[0259] 其中,音频处理装置555包括一系列的模块,包括分解模块5551、压缩模块5552、编码模块5553。下面继续说明本申请实施例提供的音频处理装置555中各个模块配合实现音频编码的方案。

[0260] 分解模块,用于对音频信号进行多通道信号分解处理,得到所述音频信号的N个子带信号,其中,所述N为大于2的整数,所述N个子带信号的频段依次递增;压缩模块,用于对每个所述子带信号进行信号压缩处理,得到每个所述子带信号的子带信号特征;编码模块,用于对每个所述子带信号的特征进行量化编码处理,得到每个所述子带信号的码流。

[0261] 在一些实施例中,所述多通道信号分解处理是通过多层二通道子带分解实现的;所述分解模块还用于对所述音频信号进行第一层的所述二通道子带分解处理,得到所述第一层的低频子带信号以及所述第一层的高频子带信号;对第i层的子带信号进行第i+1层的所述二通道子带分解处理,得到所述第i+1层的低频子带信号以及所述i+1层的高频子带信号;其中,所述第i层的子带信号为所述第i层的低频子带信号,或所述第i层的高频子带信号和所述第i层的低频子带信号,所述i为递增的自然数且取值范围为 $1 \leq i < N$;将最后一层的子带信号以及每一层中未进行所述二通道子带分解处理的高频子带信号,作为所述音频信号的特征子带信号。

[0262] 在一些实施例中,所述分解模块还用于对所述音频信号进行采样处理,得到采样信号,其中,所述采样信号包括采样得到的多个样本点;对所述采样信号进行所述第一层的低通滤波处理,得到所述第一层的低通滤波信号;对所述第一层的低通滤波信号进行下采样处理,得到所述第一层的低频子带信号;对所述采样信号进行所述第一层的高通滤波处理,得到所述第一层的高通滤波信号;对所述第一层的高通滤波信号进行下采样处理,得到所述第一层的高频子带信号。

[0263] 在一些实施例中,所述压缩模块还用于针对任意所述子带信号执行以下处理:调用所述子带信号对应的第一神经网络模型;通过所述第一神经网络模型对所述子带信号进行特征提取处理,得到所述子带信号的特征子带信号特征;其中,所述第一神经网络模型的结构复杂度与所述子带信号的特征子带信号特征的维度正相关。

[0264] 在一些实施例中,所述压缩模块还用于通过所述第一神经网络模型对所述子带信号执行以下处理:对所述子带信号进行卷积处理,得到所述子带信号的卷积特征;对所述卷积特征进行池化处理,得到所述子带信号的池化特征;对所述池化特征进行下采样处理,得到所述子带信号的下采样特征;对所述下采样特征进行卷积处理,得到所述子带信号的特征子带信号特征。

[0265] 在一些实施例中,所述压缩模块还用于对前k个子带信号分别进行特征提取处理,得到所述前k个子带信号分别对应的子带信号特征;对后N-k个子带信号分别进行频带扩展

处理,得到所述后 $N-k$ 个子带信号分别对应的子带信号特征;其中,所述 k 为整数且取值范围为 $1 < k < N$ 。

[0266] 在一些实施例中,所述压缩模块还用于针对所述后 $N-k$ 个子带信号中的任意所述子带信号执行以下处理:基于所述子带信号包括的多个样本点进行频域变换处理,得到所述多个样本点分别对应的变换系数;将所述多个样本点分别对应的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均值处理,得到每个所述子带对应的平均能量,并将所述平均能量作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子带信号对应的子带信号特征。

[0267] 在一些实施例中,所述压缩模块还用于获取参考音频信号的参考子带信号,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,所述参考子带信号与所述子带信号的频段相同;基于所述参考子带信号包括的多个样本点以及所述子带信号包括的多个样本点,对所述子带信号包括的多个样本点进行离散余弦变换处理,得到所述子带信号包括的多个样本点分别对应的变换系数。

[0268] 在一些实施例中,所述编码模块还用于对每个所述子带信号的特征进行量化处理,得到所述子带信号特征的索引值;对所述子带信号特征的索引值进行熵编码处理,得到所述子带信号的特征码流。

[0269] 其中,音频处理装置555包括一系列的模块,包括解码模块5554,解压模块5555、合成模块5556。下面继续说明本申请实施例提供的音频处理装置555中各个模块配合实现音频解码的方案。

[0270] 解码模块,用于对 N 个码流进行量化解码处理,得到每个所述码流对应的子带信号特征;其中,所述 N 为大于2的整数,所述 N 个码流是对音频信号经过多通道信号分解处理后得到的 N 个子带信号分别进行编码得到的;解压模块,用于对每个所述子带信号特征进行信号解压处理,得到每个所述子带信号特征对应的估计子带信号;合成模块,用于对多个所述估计子带信号进行信号合成处理,得到所述多个码流对应的合成音频信号。

[0271] 在一些实施例中,所述解压模块还用于针对任意所述子带信号特征执行以下处理:调用所述子带信号特征对应的第二神经网络模型;通过所述第二神经网络模型对所述子带信号特征进行特征重建,得到所述子带信号特征对应的估计子带信号;其中,所述第二神经网络模型的结构复杂度与所述子带信号特征的维度正相关。

[0272] 在一些实施例中,所述解压模块还用于通过所述第二神经网络模型对所述子带信号特征执行以下处理:对所述子带信号特征进行卷积处理,得到所述子带信号特征的卷积特征;对所述卷积特征进行上采样处理,得到所述子带信号特征的上采样特征;对所述上采样特征进行池化处理,得到所述子带信号特征的池化特征;对所述池化特征进行卷积处理,得到所述子带信号特征对应的估计子带信号。

[0273] 在一些实施例中,所述解压模块还用于对前 k 个子带信号特征分别进行特征重建处理,得到所述前 k 个子带信号特征分别对应的估计子带信号;对后 $N-k$ 个子带信号特征分别进行频带扩展的逆处理,得到所述后 $N-k$ 个子带信号特征分别对应的估计子带信号;其中,所述 k 为整数且取值范围为 $1 < k < N$ 。

[0274] 在一些实施例中,所述解压模块还用于针对所述后 $N-k$ 个子带信号特征中的任意所述子带信号特征执行以下处理:对所述前 k 个估计子带信号中与所述子带信号特征关联

的所述估计子带信号进行信号合成处理,得到与所述子带信号特征对应的低频子带信号;基于所述低频子带信号包括的多个样本点进行频域变换处理,得到所述多个样本点分别对应的变换系数;对所述多个样本点分别对应的变换系数中的后半部分的变换系数进行频谱复制处理,得到参考子带信号的参考变换系数;基于所述子带信号特征对应的子带谱包络,对所述参考子带信号的参考变换系数进行增益处理,得到增益后的所述参考变换系数;对增益后的所述参考变换系数进行反频域变换处理,得到所述子带信号特征对应的估计子带信号。

[0275] 在一些实施例中,所述解压模块还用于基于所述子带信号特征对应的子带谱包络,将所述参考子带信号的参考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第一平均能量,并确定所述子带对应的第二平均能量;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变换系数相乘,得到增益后的所述参考变换系数。

[0276] 在一些实施例中,所述解码模块还用于对所述N个码流中的任意所述码流执行以下处理:对所述码流进行熵解码处理,得到所述码流对应的索引值;对所述码流对应的索引值进行逆量化处理,得到所述码流对应的子带信号特征。

[0277] 在一些实施例中,所述合成模块还用于对多个所述估计子带信号分别进行上采样处理,得到多个所述估计子带信号分别对应的滤波信号;对多个所述滤波信号进行滤波合成处理,得到所述多个码流对应的合成音频信号。

[0278] 本申请实施例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。电子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,使得该电子设备执行本申请实施例上述的音频处理方法。

[0279] 本申请实施例提供一种存储有可执行指令的计算机可读存储介质,其中存储有可执行指令,当可执行指令被处理器执行时,将引起处理器执行本申请实施例提供的音频处理方法,例如,如图4-5示出的音频处理方法。

[0280] 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。

[0281] 在一些实施例中,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。

[0282] 作为示例,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。

[0283] 作为示例,可执行指令可被部署为在一个计算设备上执行,或者在位于一个地点的多个计算设备上执行,又或者,在分布在多个地点且通过通信网络互连的多个计算设备

上执行。

[0284] 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。

[0285] 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。

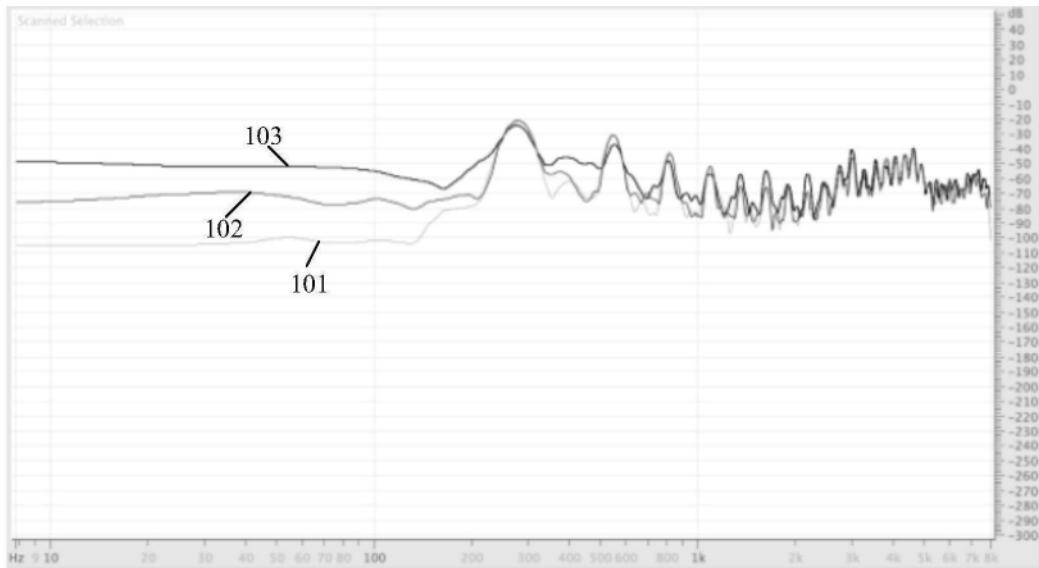


图1

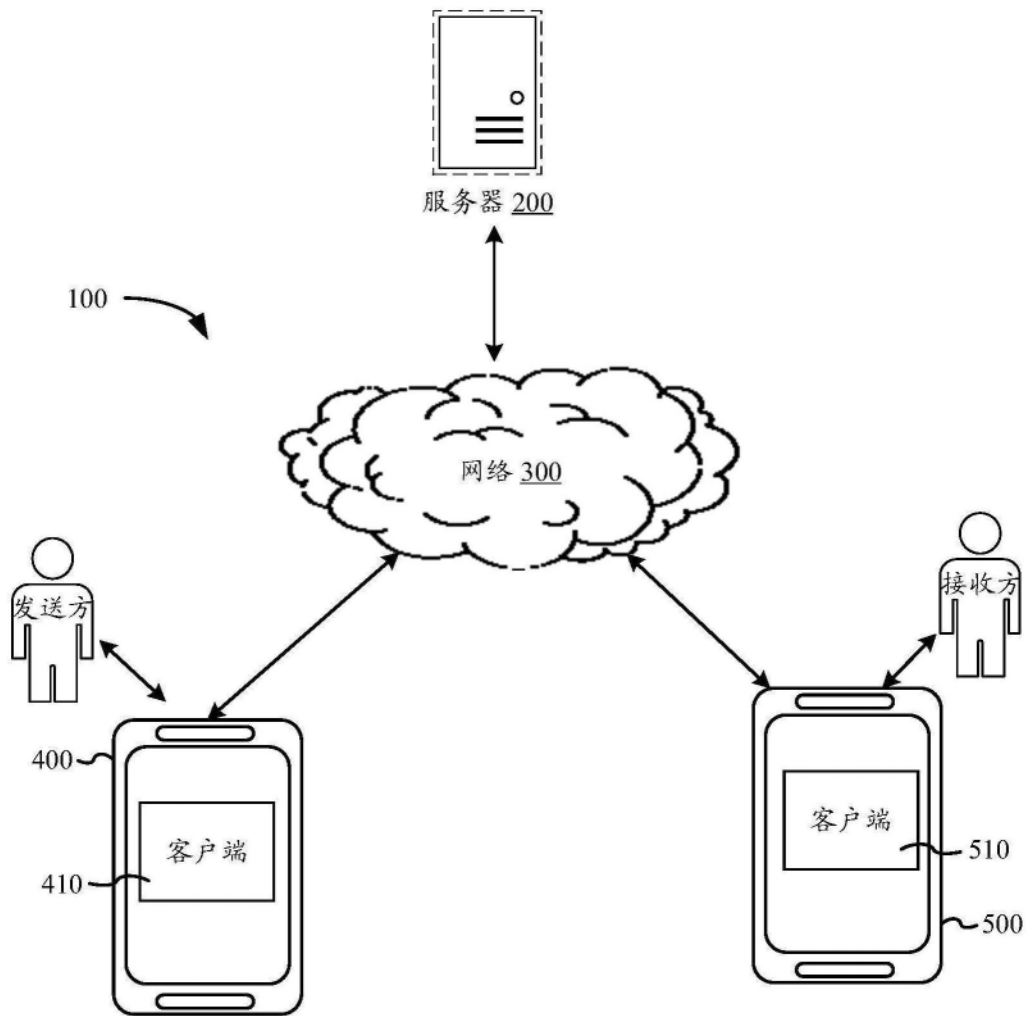


图2

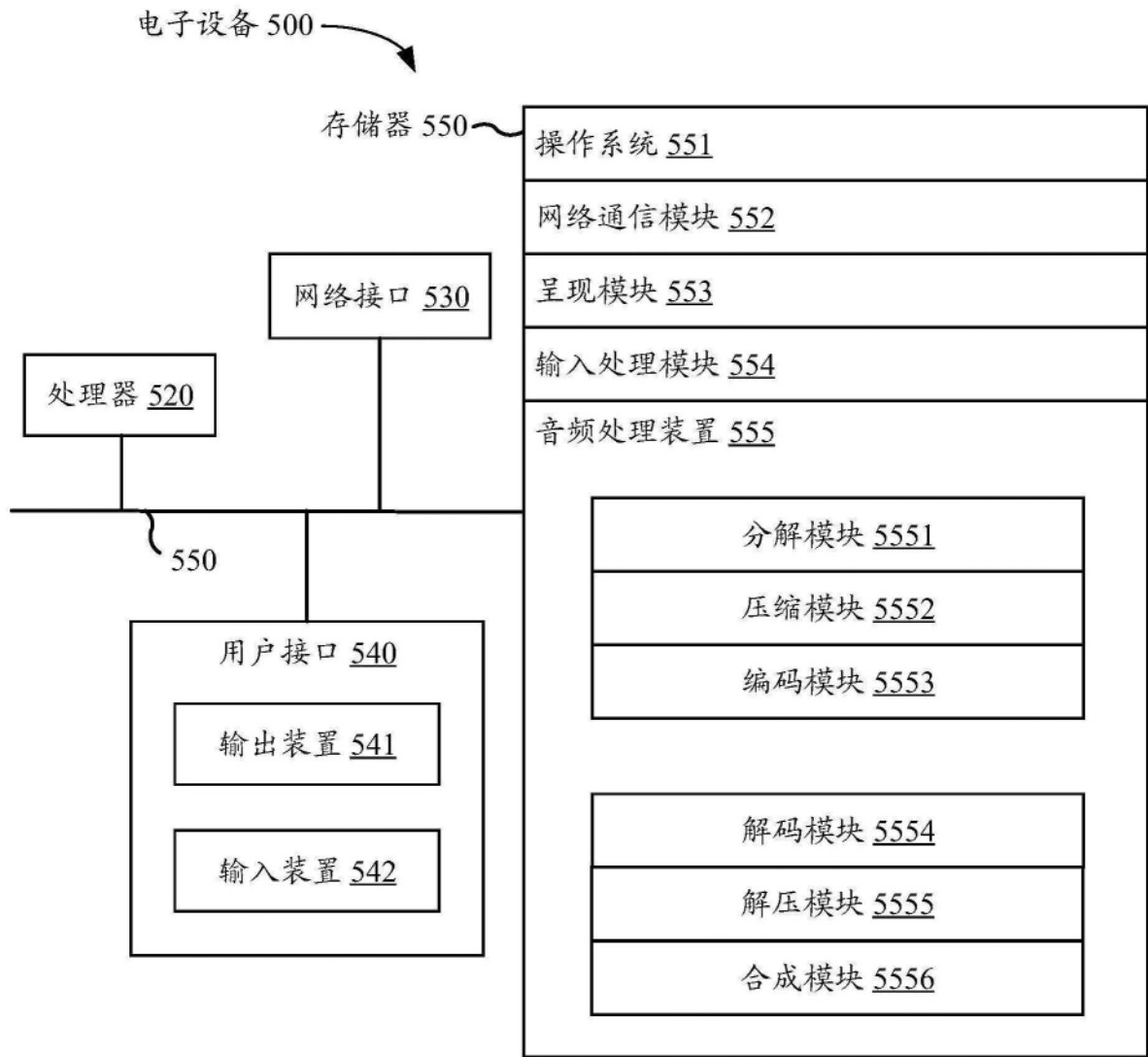


图3

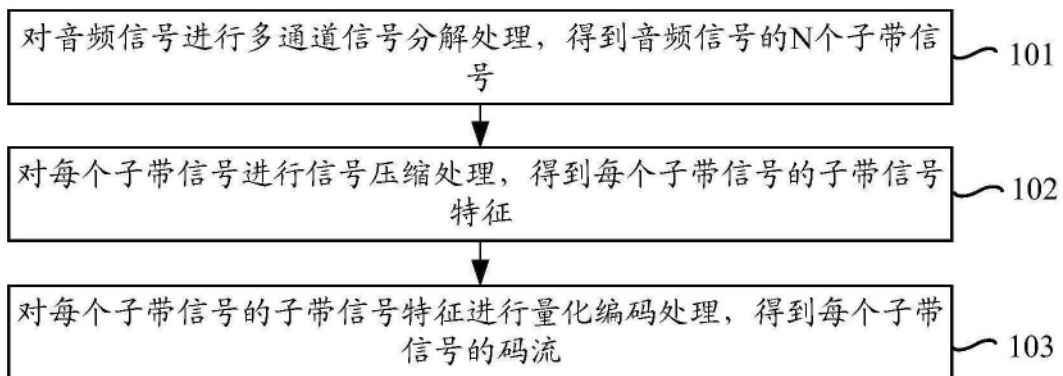


图4

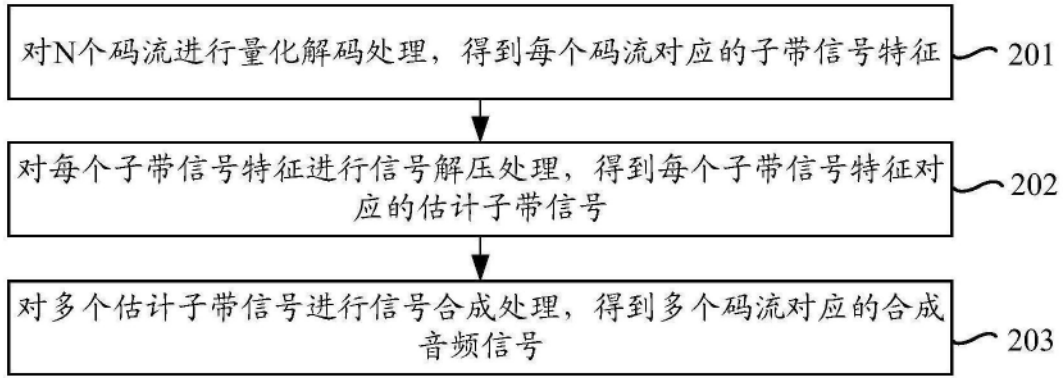


图5

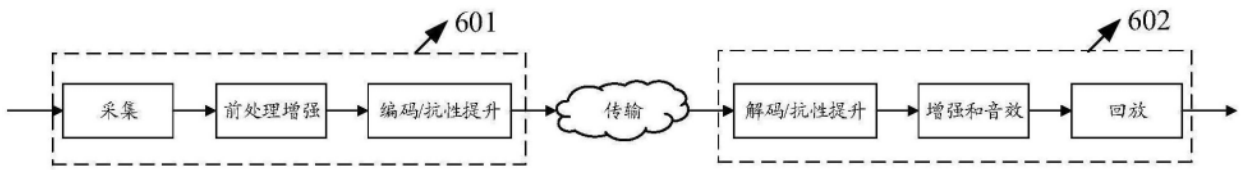


图6

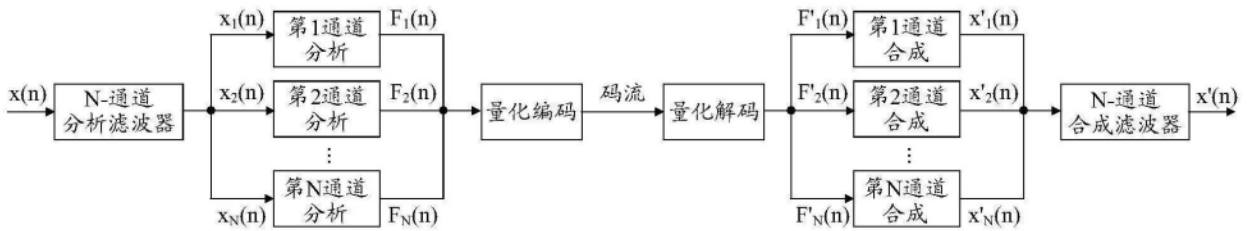


图7

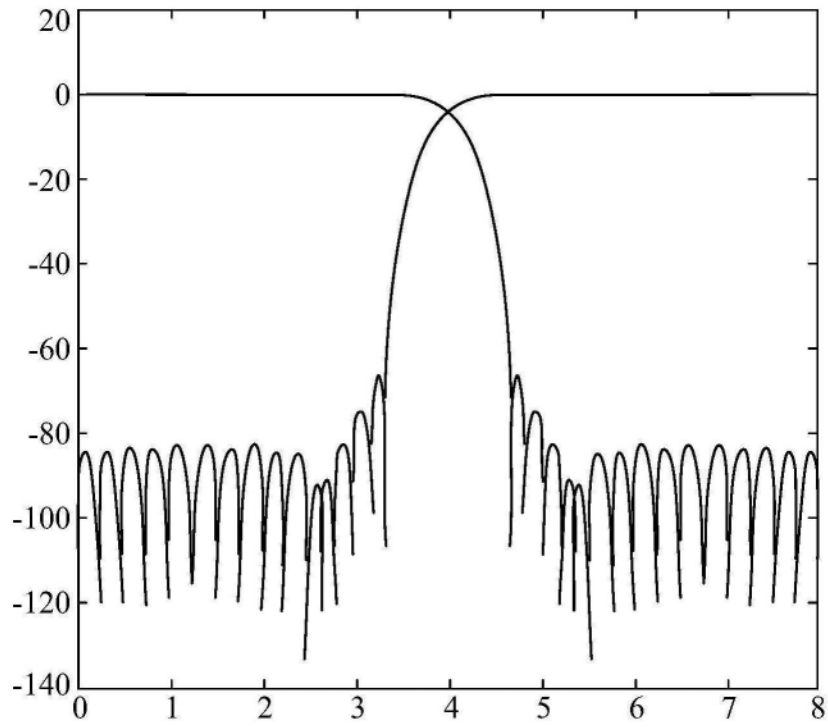


图8A

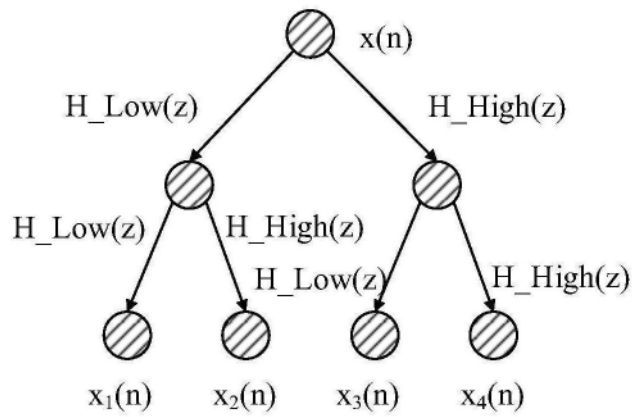


图8B

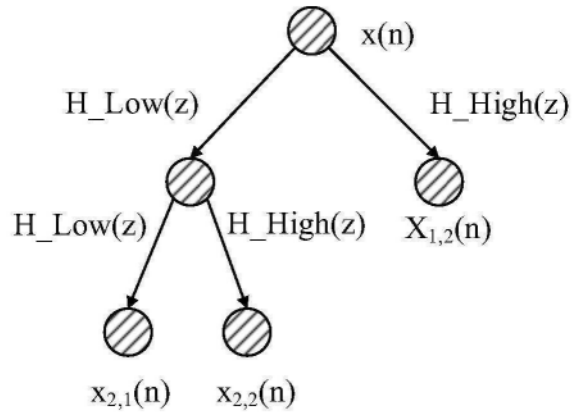


图8C

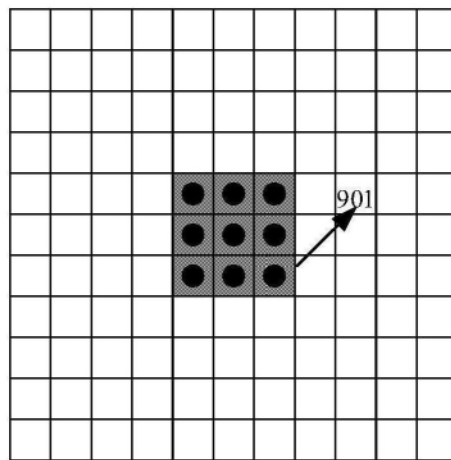


图9A

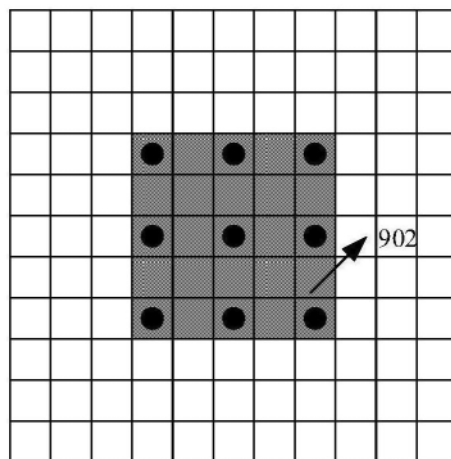


图9B

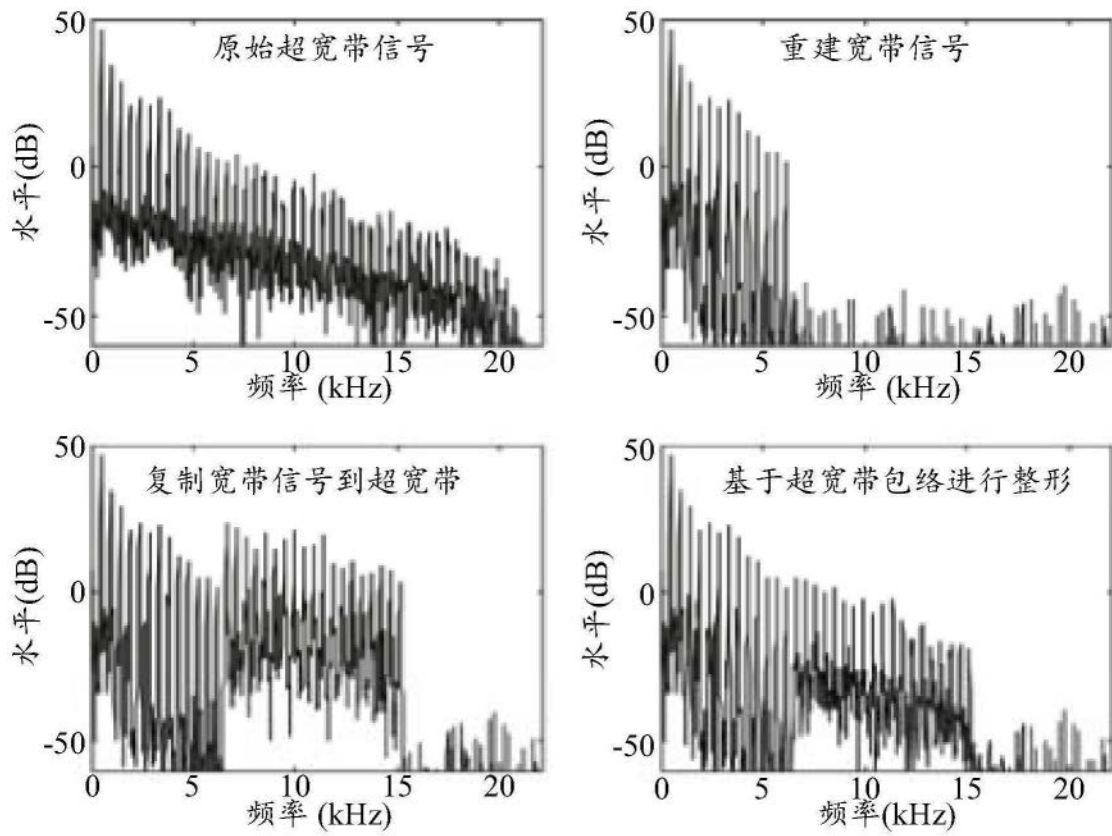


图10

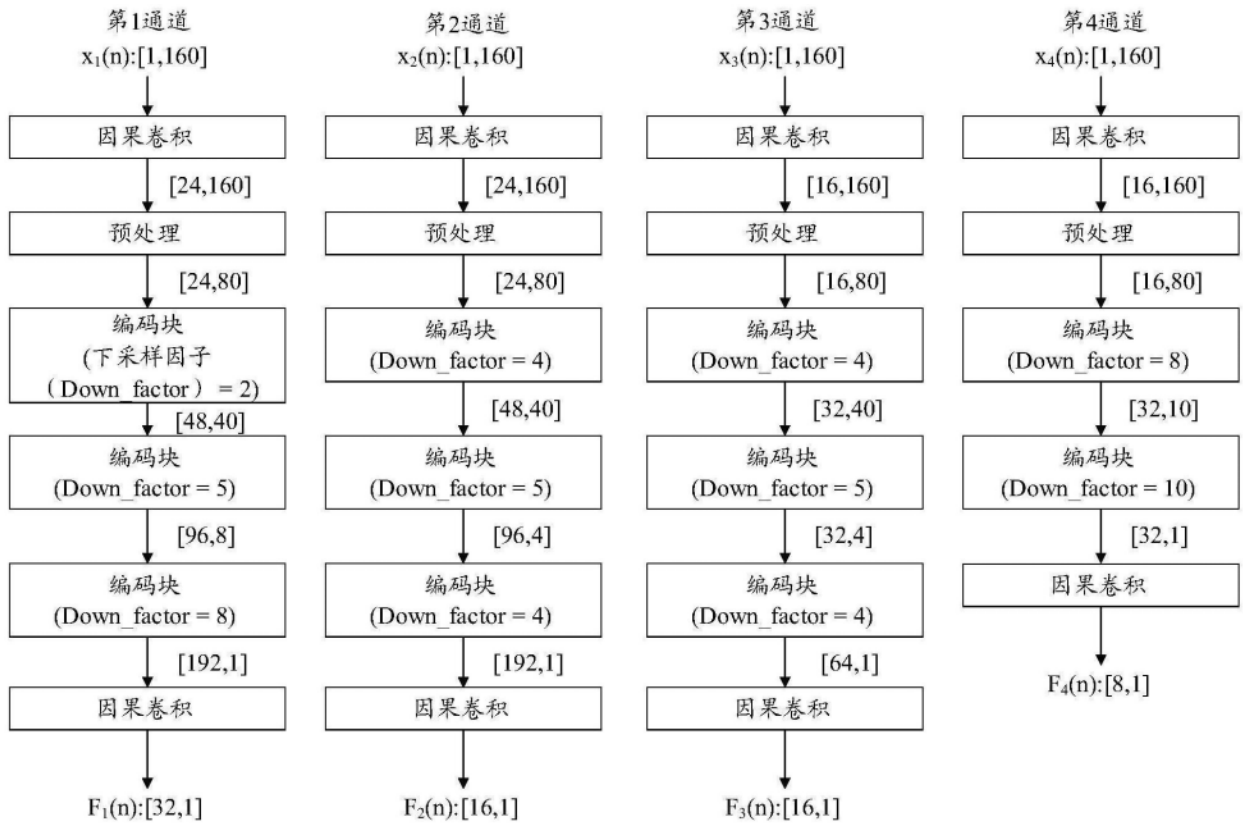


图11

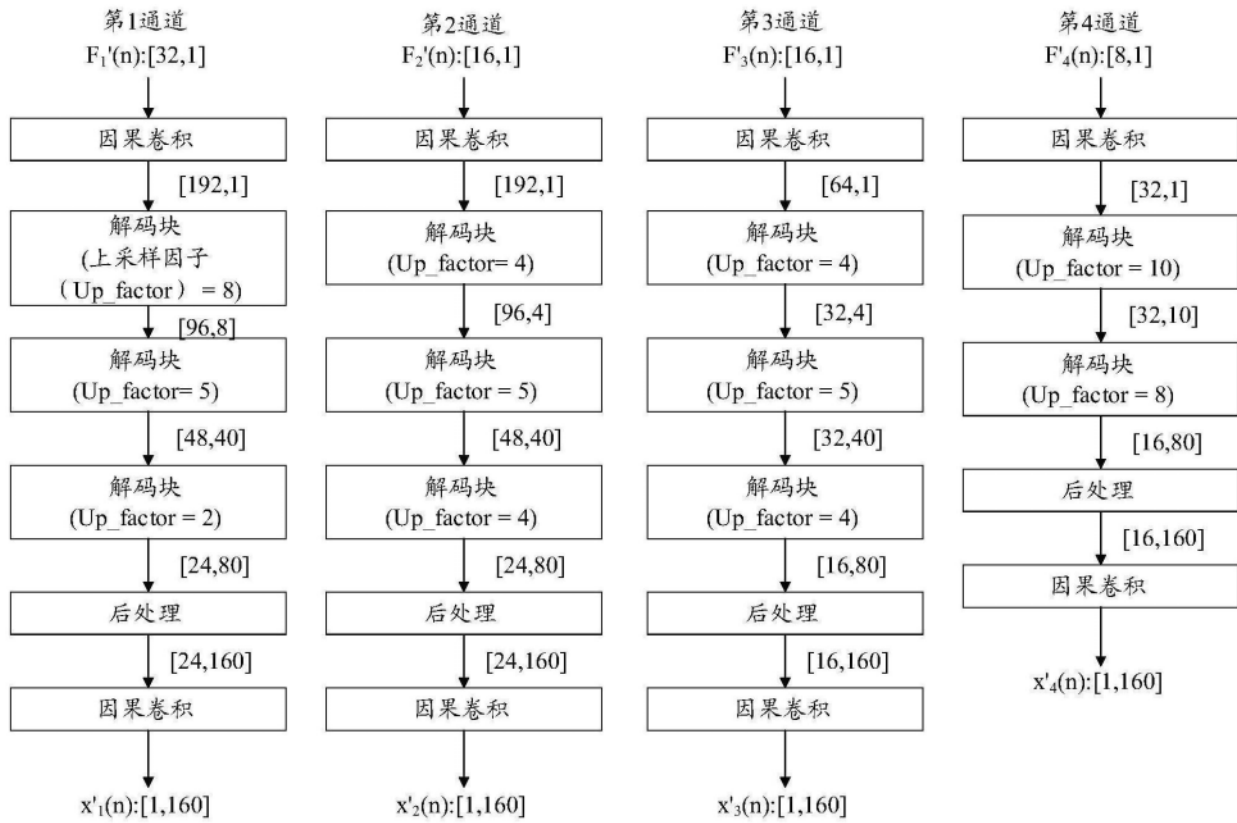


图12