(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110890101 B
(45)授权公告日 2024.01.12

(21)申请号 201911328515.3
(22)申请日 2014.08.27
(65)同一申请的已公布的文献号
申请公布号 CN 110890101 A
(43)申请公布日 2020.03.17
(30)优先权数据
61/870,933 2013.08.28 US
61/895,959 2013.10.25 US
61/908,664 2013.11.25 US
(62)分案原申请数据
201480048109.0 2014.08.27
(73)专利权人 杜比实验室特许公司
地址 美国加利福尼亚州
专利权人 杜比国际公司
(72)发明人 耶伦·科庞 汉内斯·米施
(74)专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 杜诚 马骁
(51)Int.Cl.
G10L 21/0324(2013.01)
G10L 21/0364(2013.01)
G10L 19/008(2013.01)
G10L 19/20(2013.01)
G10L 19/22(2013.01)
H04R 5/04(2006.01)
(56)对比文件
JP 2004004952 A,2004.01.08
US 2003152152 A1,2003.08.14
US 2008243277 A1,2008.10.02
US 6167375 A,2000.12.26
CN 101606195 A,2009.12.16
CN 102947880 A,2013.02.27
EP 2544465 A1,2013.01.09
US 6691082 B1,2004.02.10
US 2008319739 A1,2008.12.25
US 2010106507 A1,2010.04.29
审查员 杜培笑

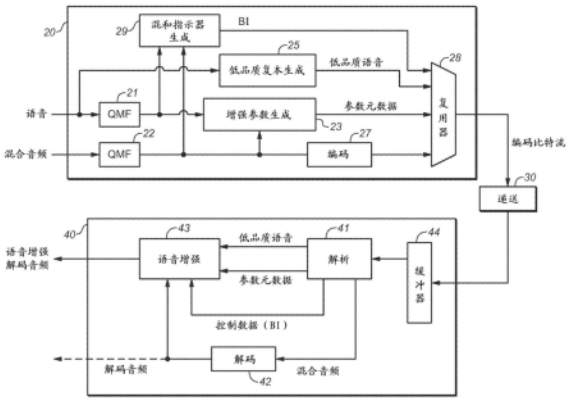
权利要求书2页 说明书39页 附图9页

(54)发明名称

用于基于语音增强元数据进行解码的方法和设备

(57)摘要

一种用于混合语音增强的方法,该方法在一些信号条件下使用参数编码增强(或者参数编码增强和波形编码增强的混和)并且在其他信号条件下使用波形编码增强(或者参数编码增强和波形编码增强的不同混和)。其他方面是:用于生成指示包括语音内容和其他内容的音频节目的比特流以使得能够对节目执行混合语音增强的方法;包括对由本发明方法的任何实施方式所生成的编码音频比特流的至少一个片段进行存储的缓冲器的解码器;以及被配置(例如,被编程)为执行本发明方法的任何实施方式的系统或装置(例如,编码器或解码器)。由接收方音频解码器利用由上游音频编码器所生成的中间/侧语音增强元数据来执行语音增强操作中的至少一些。



CN 110890101 B

1.一种音频信号处理方法,包括:

接收混合音频内容,其中,所述混合音频内容至少包括中间通道混合内容信号和侧通道混合内容信号,其中,所述中间通道信号表示参考音频通道表示的两个通道的加权和或非加权和,其中,所述侧通道信号表示所述参考音频通道表示的两个通道的加权差或非加权差,并且,所述混合音频内容具有语音内容与非语音音频内容的混合;

由音频解码器将所述中间通道信号和所述侧通道信号解码成左通道信号和右通道信号,其中,所述解码包括基于语音增强元数据进行解码,其中,所述语音增强元数据包括指示在解码期间要对所述中间通道信号和所述侧通道信号执行的至少一种类型的语音增强操作的优选标记,并且其中,所述语音增强元数据还指示用于所述中间通道信号的第一类型的语音增强和所述中间通道信号的第二类型的语音增强;以及

生成音频信号,所述音频信号包括针对所述混合音频内容的解码的中间通道信号和侧通道信号的一个或更多部分的所述左通道信号和所述右通道信号,

其中,所述方法由一个或更多个计算装置来执行;

所述第一类型的语音增强为波形编码语音增强;并且

所述第二类型的语音增强为参数语音增强。

2.根据权利要求1所述的方法,其中,所述混合音频内容包括参考音频通道表示,所述参考音频通道表示包括与环绕扬声器有关的音频通道。

3.根据权利要求1所述的方法,其中,所述语音增强元数据包括与所述中间通道信号有关的单个语音增强元数据的集合。

4.根据权利要求1所述的方法,其中,所述语音增强元数据表示所述混合音频内容的全部音频元数据的一部分。

5.根据权利要求1所述的方法,其中,编码在所述混合音频内容中的音频元数据包括指示所述语音增强元数据的存在的数据字段。

6.根据权利要求1所述的方法,其中,所述混合音频内容是音视频信号的一部分。

7.一种非暂态计算机可读存储介质,包括当由一个或更多个处理器执行时使得执行权利要求1的软件指令。

8.一种音频信号处理设备,包括:

接收器,被配置成接收混合音频内容,其中,所述混合音频内容至少包括中间通道混合内容信号和侧通道混合内容信号,其中,所述中间通道信号表示参考音频通道表示的两个通道的加权和或非加权和,其中,所述侧通道信号表示所述参考音频通道表示的两个通道的加权差或非加权差,并且,所述混合音频内容具有语音内容与非语音音频内容的混合;

解码器,被配置成将所述中间通道信号和所述侧通道信号解码成左通道信号和右通道信号,其中,所述解码包括基于语音增强元数据进行解码,其中,所述语音增强元数据包括指示在解码期间要对所述中间通道信号和所述侧通道信号执行的至少一种类型的语音增强操作的优选标记,并且其中,所述语音增强元数据还指示用于所述中间通道信号的第一类型的语音增强和所述中间通道信号的第二类型的语音增强;以及

处理器,被配置成生成音频信号,所述音频信号包括针对所述混合音频内容的解码的中间通道信号和侧通道信号中的一个或更多部分的所述左通道信号和所述右通道信号;

其中,所述第一类型的语音增强为波形编码语音增强;并且

所述第二类型的语音增强为参数语音增强。

9.根据权利要求8所述的设备,其中,所述混合音频内容包括参考音频通道表示,所述参考音频通道表示包括与环绕扬声器有关的音频通道。

10.根据权利要求8所述的设备,其中,所述语音增强元数据包括与所述中间通道信号有关的单个语音增强元数据的集合。

11.根据权利要求8所述的设备,其中,所述语音增强元数据表示所述混合音频内容的全部音频元数据的一部分。

12.根据权利要求8所述的设备,其中,编码在所述混合音频内容中的音频元数据包括指示所述语音增强元数据的存在的数据字段。

13.根据权利要求8所述的设备,其中,所述混合音频内容是音视频信号的一部分。

用于基于语音增强元数据进行解码的方法和设备

[0001] 本申请是申请号为201480048109.0、申请日为2014年8月27日、发明名称为“混合波形编码和参数编码语音增强”的中国发明专利申请的分案申请。

[0002] 相关申请的交叉引用

[0003] 本申请要求2013年8月28日提交的美国临时专利申请第61/870,933号、2013年10月25日提交的美国临时专利申请第61/895,959号以及2013年11月25日提交的美国临时专利申请第61/908,664号的优先权,上述美国临时专利申请中的每一个的全部内容通过引用合并到本文中。

技术领域

[0004] 本发明涉及音频信号处理,更具体地,涉及音频节目的语音内容相对于节目的其他内容的增强,其中,语音增强就以下这种意义而言是“混合的”:所述语音增强在一些信号条件下包括波形编码增强(或者相对较多的波形编码增强)以及在其他信号条件下包括参数编码增强(或者相对较多的参数编码增强)。其他方面是对包括足以使得能够实现这样的混合语音增强的数据的音频节目的编码、解码和呈现(render)。

背景技术

[0005] 在电影和电视中,对话和叙述经常与其他的非语音音频如来自体育赛事的音乐、效果或氛围一起呈现。在许多情况下,语音和非语音声音在声音工程师的控制下被分别捕获并且混合在一起。声音工程师以适合于大多数收听者的方式来选择相对于非语音的水平的语音的水平。然而,一些收听者——例如听力损伤的那些收听者——在理解音频节目的语音内容(具有工程师确定的语音与非语音混合比)时体验到困难,并且更偏好以更高的相对水平混合语音。

[0006] 在使得这些收听者能够相对于非语音音频内容的可听度增大音频节目语音内容的可听度时存在要解决的问题。

[0007] 一种当前的方法是向收听者提供两个高品质音频流。一个流携载主内容音频(主要是语音)而另外的流携载次内容音频(剩余音频节目,其将语音排除在外),并且赋予用户对混合处理的控制。遗憾的是,该方案是不实用的,原因是该方案并不建立在传输完全混合的音频节目的当前实践上。另外,该方案要求当前广播实践的带宽的大约两倍,原因是两个独立音频流——广播品质中的每一个——必须被递送至用户。

[0008] 在受让于杜比实验室公司并且将Hannes Muesch指定为发明人的、2010年4月29日公开的美国专利申请公开第2010/0106507A1号中描述了另一种语音增强方法(在本文中被称为“波形编码”增强)。在波形编码增强中,通过将已经与主混合一起被发送至接收器的纯净语音信号(clean speech signal)的降低品质版本(低品质复本)添加至主混合来增大语音与非语音内容的原始音频混合(有时称为主混合)的语音与背景(非语音)比。为了减少带宽开销,通常以非常低的比特率对低品质复本进行编码。由于低比特率编码,编码伪声与低品质复本相关联,并且当低品质复本被单独地呈现和试听时,编码伪声是清楚地听得见的。

因此,当被单独地试听时,低品质复本具有令人讨厌的品质。仅在非语音分量的水平高而使得编码伪声被非语音分量掩蔽的时间期间,波形编码增强试图通过将低品质复本添加至主混合来隐藏这些编码伪声。如稍后将详细描述的,该方法的限制包括以下:语音增强的量通常不能在时间上恒定,并且当主混合的背景(非语音)分量弱或者它们的频率幅度频谱与编码噪声的频率幅度频谱有很大不同时,音频伪声会变得可听见。

[0009] 根据波形编码增强,音频节目(用于递送至解码器以进行解码和随后的呈现)被编码为包括作为主混合的侧流的低品质语音复本(或者其编码版本)的比特流。比特流可以包括指示确定要执行的波形编码语音增强的量的缩放参数的元数据(即,缩放参数确定在缩放的低品质语音复本与主混合组合之前要应用于低品质语音复本的缩放因子,或者将确保对编码伪声的掩蔽的这样的缩放因子的最大值)。当缩放因子的当前值为0时,解码器不对主混合的相应片段执行语音增强。虽然缩放参数的当前值(或者缩放参数可以达到的当前最大值)通常在编码器中被确定(由于缩放参数通常由计算密集型心理声学模型生成),但是其也可以在解码器中生成。在后一种情况下,不需要将指示缩放参数的元数据从编码器发送至解码器,并且替代地,解码器可以根据主混合确定混合的语音内容的功率与混合的功率之比,并且响应于功率比的当前值来实现确定缩放参数的当前值的模型。

[0010] 用于在存在竞争音频(背景)的情况下增强语音的可理解度的另一种方法(在本文中要被称为“参数编码”增强)是:将原始音频节目(通常是音轨)分割成时间/频率分块(tile)并且根据它们的语音内容与背景内容的功率(或水平)的比率来增强分块以实现语音分量相对于背景的增强。该方法的基本构思类似于指引频谱减少噪声抑制的基本构思。在该方法的极端示例中,其中,SNR(即,语音分量的功率或水平与竞争声音内容的功率或水平的比率)在预定阈值以下的所有分块被完全抑制,已经显示该方法提供鲁棒的语音可理解度增强。在该方法应用于广播时,可以通过将原始音频混合(语音与非语音内容的)与混合的语音分量进行比较来推断语音与背景比(SNR)。然后,可以将所推断的SNR转换成与原始音频混合一起被发送的增强参数的适当的集合。在接收器处,可以(可选地)将这些参数应用于原始音频混合以获得指示增强语音的信号。如稍后将详细描述的,当语音信号(混合的语音分量)比背景信号(混合的非语音分量)占优势时,参数编码增强最优地发挥作用。

[0011] 波形编码增强要求递送音频节目的语音分量的低品质复本在接收器处可用。为了限制在与主音频混合一起发送该复本时引起的数据开销,以非常低的比特率对该复本进行编码并且该复本呈现编码失真。当非语音分量的水平高时,这些编码失真很可能被原始音频掩蔽。当编码失真被掩蔽时,所得到的增强音频的品质非常好。

[0012] 参数编码增强是基于将主音频混合信号解析成时间/频率分块并且向这些分块中的每一个应用适当的增益/衰减。当与波形编码增强的数据率相比时,将这些增益转发至接收器所需的数据率较低。然而,由于参数的有限的时间频谱分辨率,当语音与非语音音频混合时,语音不能被操纵,也不会影响非语音音频。因此,音频混合的语音内容的参数编码增强在混合的非语音内容中引入调制,并且当回放语音增强混合时,该调制(“背景调制”)会变得令人讨厌。当语音与背景比非常低时,背景调制最可能令人讨厌。

[0013] 在本部分中描述的方法是能够被执行的方法,但是不一定是先前已经被构思或执行的方法。因此,除非另有说明,否则不应该假定在本部分中描述的任何方法仅因其被包括在本部分中而被认为是现有技术。类似地,除非另有说明,否则不应该假定在基于本部分的

任何现有技术中已经意识到关于一种或更多种方法而识别出的问题。

附图说明

[0014] 在附图的图中以示例性方式而非限制性方式来说明本发明,并且在附图中相似的附图标记指代类似的要素,并且其中:

[0015] 图1是被配置成生成用于重构单通道混合内容信号(具有语音内容和非语音内容)的语音内容的预测参数的系统的框图。

[0016] 图2是被配置成生成用于重构多通道混合内容信号(具有语音内容和非语音内容)的语音内容的预测参数的系统的框图。

[0017] 图3是包括被配置成执行本发明的编码方法的实施方式以生成指示音频节目的编码音频比特流的编码器,以及被配置成对编码音频比特流进行解码并执行语音增强(根据本发明方法的实施方式)的解码器的系统的框图。

[0018] 图4是被配置成呈现包括通过对其执行常规语音增强的多通道混合内容音频信号的系统的框图。

[0019] 图5是被配置成呈现包括通过对其执行常规参数编码语音增强的多通道混合内容音频信号的系统的框图。

[0020] 图6和图6A是被配置成呈现包括通过对其执行本发明的语音增强方法的实施方式的多通道混合内容音频信号的系统的框图。

[0021] 图7是用于使用听觉掩蔽模型来执行本发明的编码方法的实施方式的系统的框图。

[0022] 图8A和图8B示出了示例处理流程,以及

[0023] 图9示出了在其上可以实现如本文中所描述的计算机或计算装置的示例硬件平台。

具体实施方式

[0024] 在本文中描述了涉及混合波形编码和参数编码语音增强的示例实施方式。在下面的描述中,出于说明的目的,阐述了大量具体细节以提供对本发明的透彻理解。然而,将会明白可以在没有这些具体细节的情况下实践本发明。在其他实例中,并未详尽地描述已知的结构和装置,以避免不必要地封闭本发明、模糊或者混淆本发明。

[0025] 在本文中根据以下概要来描述示例实施方式:

[0026] 1.一般概述

[0027] 2.符号和术语

[0028] 3.预测参数的生成

[0029] 4.语音增强操作

[0030] 5.语音呈现

[0031] 6.中间/侧表示

[0032] 7.示例处理流程

[0033] 8.实现机构——硬件概述

[0034] 9.等同方案、扩展方案、替代方案和其他方案

[0035] 1.一般概述

[0036] 本概述提供对本发明的实施方式的一些方面的基本描述。应当注意,该概述并非对实施方式的各方面的广泛或详尽概括。此外,应当注意,此概述并非意在被理解为识别实施方式的任何特别显著的方面或要素,也并非意在被理解为划定一般为本发明、特别是实施方式的任何范围。此概述仅以扼要和简化的形式提供与示例实施方式有关的一些概念,并且应当被理解为仅是随后在下面描述的示例实施方式的更详细描述的概念性前序。注意,尽管本文中讨论了单独的实施方式,但是可以将本文中讨论的部分实施方式和/或实施方式的任意组合进行组合以形成另外的实施方式。

[0037] 发明人已经意识到参数编码增强和波形编码增强的各自的优势和弱势可以彼此抵消,并且意识到可以通过以下混合增强方法来显著改善常规语音增强,该混合增强方法在一些信号条件下使用参数编码增强(或者参数编码增强与波形编码增强的混和(blend))并且在其他信号条件下使用波形编码增强(或者参数编码增强与波形编码增强的不同的混和)。本发明的混合增强方法的典型实施方式提供比通过单独的参数编码增强或者波形编码增强可以实现的语音增强更稳定并且品质更好的语音增强。

[0038] 在一类实施方式中,本发明方法包括以下步骤:(a)接收指示包括具有未增强的波形的语音以及其他音频内容的音频节目的比特流,其中,比特流包括指示语音内容和其他音频内容的音频数据,指示语音的降低品质的版本的波形数据(其中,已经通过将语音数据与非语音数据混合而生成了音频数据,与语音数据相比,波形数据通常包括较少比特),其中降低品质的版本具有与未增强的波形类似(例如,至少基本上类似)的第二波形,如果被单独地试听,则降低品质的版本将具有令人讨厌的品质,以及比特流包括参数数据,其中参数数据与音频数据一起确定参数构造语音,并且参数构造语音是至少与该语音基本匹配(例如,是该语音的良好近似)的该语音的参数重构版本;以及(b)响应于混和指示符对比特流执行语音增强,从而生成指示语音增强音频节目的数据,包括通过将音频数据与根据波形数据确定的低品质语音数据和重构语音数据的组合进行组合,其中,组合由混和指示符来确定(例如,组合具有由混和指示符的当前值序列所确定的状态序列),响应于参数数据中的至少一些以及音频数据中的至少一些来生成重构语音数据,与通过将仅低品质语音数据(其指示语音的降低品质的版本)与音频数据组合所确定的纯波形编码语音增强音频节目或者根据参数数据和音频数据确定的纯参数编码语音增强音频节目相比,该语音增强音频节目具有较少的听得见的语音增强伪声(例如,当语音增强音频节目被呈现和试听时被较好地掩蔽并且从而较少听得见的语音增强伪声)。

[0039] 在本文中,“语音增强伪声”(或者“语音增强编码伪声”)表示由语音信号的表示(例如,连同混合内容信号一起的参数数据或者波形编码语音信号)所引起的音频信号(指示语音信号和非语音音频信号)的失真(通常是可测量的失真)。

[0040] 在一些实施方式中,混和指示符(其可以具有值序列,例如,针对比特流片段序列中的每一个有一个值序列)被包括在步骤(a)中所接收的比特流中。一些实施方式包括以下步骤:响应于在步骤(a)中所接收的比特流来生成混和指示符(例如,在接收比特流并且对比特流进行解码的接收器中)。

[0041] 应当理解,表达式“混和指示符”并非意在要求混和指示符是比特流的每个片段的单个参数或值(或者单个参数或值序列)。而是,可以想到,在一些实施方式中,混和指示符

(针对比特流的片段)可以是两个或更多个参数或值(例如,对于每个片段,参数编码增强控制参数以及波形编码增强控制参数)的集合,或者参数或值的集合的序列。

[0042] 在一些实施方式中,每个片段的混和指示符可以是指示片段的每频带的混和的值序列。

[0043] 无需针对比特流的每个片段设置(例如,包括)波形数据和参数数据,无需使用波形数据和参数数据两者来对比特流的每个片段执行语音增强。例如,在一些情况下,至少一个片段可以包括仅波形数据(并且由每个这样的片段的混和指示符所确定的组合可以包括仅波形数据)并且至少一个其他片段可以包括仅参数数据(并且由每个这样的片段的混和指示符所确定的组合可以包括仅重构语音数据)。

[0044] 通常可以想到,编码器生成比特流,其包括通过对音频数据进行编码(例如,压缩)而不对波形数据或参数数据应用相同的编码。因此,当比特流被递送至接收器时,接收器通常对比特流进行解析以提取音频数据、波形数据和参数数据(以及混和指示符,如果其在比特流中被递送),但是仅对音频数据进行解码。在不对波形数据或参数数据应用与对音频数据应用的解码处理相同的解码处理的情况下,接收器通常(使用波形数据和/或参数数据)对所解码的音频数据执行语音增强。

[0045] 通常,波形数据和重构语音数据的组合(由混和指示符所指示)随时间变化,其中,每个组合状态与比特流的相应片段的语音内容和其他音频内容有关。混和指示符被生成为使得(波形数据和重构语音数据的)当前组合状态至少部分地由比特流的相应片段中的语音内容和其他音频内容的信号特性(例如,语音内容的功率与其他音频内容的功率的比)来确定。在一些实施方式中,混和指示符被生成为使得当前组合状态由比特流的相应片段中的语音内容和其他音频内容的信号特性来确定。在一些实施方式中,混和指示符被生成为使得当前组合状态由比特流的相应片段中的语音内容和其他音频内容的信号特性,以及波形数据中的编码伪声的量两者来确定。

[0046] 步骤(b)可以包括以下步骤:通过将低品质语音数据中的至少一些与比特流的至少一个片段的音频数据进行组合(例如,混合或混和)来执行波形编码语音增强;以及通过将重构语音数据与比特流中的至少一个片段的音频数据进行组合来执行参数编码语音增强。通过将片段的低品质语音数据和参数构造语音两者与片段的音频数据进行混和来对比特流中的至少一个片段执行波形编码语音增强与参数编码语音增强的组合。在一些信号条件下,对比特流的片段(或者对多于一个片段中的每一个)(响应于混和指示符)执行波形编码语音增强和参数编码语音增强中的仅一个(而不是两者)。

[0047] 在本文中,将使用表达“SNR”(信噪比)来表示音频节目(或者整个节目)的片段的语音内容与片段或节目的非语音内容的功率比(或水平差),或者节目(或整个节目)的片段的语音内容与片段或节目的整个(语音和非语音)内容的功率比(或水平差)。

[0048] 在一类实施方式中,本发明方法实现了音频节目的片段的参数编码增强与波形编码增强之间的基于“盲”时间SNR的切换。在此上下文中,“盲”表示切换并不由(例如,本文中要描述的类型的)复杂听觉掩蔽模型感知地指引,而是由与节目的片段相对应的SNR值序列(混和指示符)指引。在该类中的一种实施方式中,通过参数编码增强与波形编码增强之间的时间切换来实现混合编码语音增强,使得对执行语音增强的音频节目的每个片段执行参数编码增强或波形编码增强(而非参数编码增强和波形编码增强两者)。意识到波形编码增

强在低SNR条件下(对具有低SNR值的片段)最优地执行并且参数编码增强在良好的SNR下(对具有高SNR值的片段)最优地执行,切换决定通常基于原始音频混合中的语音(对话)与剩余音频的比率。

[0049] 实现基于“盲”时间SNR的切换的实施方式通常包括以下步骤:将未增强的音频信号(原始音频混合)分割成连续的时间片(片段),以及针对每个片段来确定片段的语音内容与其他音频内容之间(或者语音内容与总音频内容之间)的SNR;以及对于每个片段,将SNR与阈值进行比较,并且当SNR大于阈值时,针对片段(即,该片段的混和指示符指示应执行参数编码增强)设置参数编码增强控制参数,或者当SNR不大于阈值时,针对片段(即,混和指示符表示应执行该片段的波形编码增强)设置波形编码增强控制参数。通常,未增强的音频信号与作为元数据所包括的控制参数一起被递送(例如,被发送)至接收器,接收器(对每个片段)执行由片段的控制参数所指示的类型的语音增强。因此,接收器对控制参数是参数编码增强控制参数的每个片段执行参数编码增强,并且接收器对控制参数是波形编码增强控制参数的每个片段执行波形编码增强。

[0050] 如果愿意承担与原始(未增强)混合一起来传输(与原始音频混合的每个片段一起)波形数据(用于实现波形编码语音增强)和参数编码增强参数两者的成本,那么通过对混合的各个片段应用波形编码增强和参数编码增强两者可以获得较高程度的语音增强。因此,在一类实施方式中,本发明方法实现音频节目的片段的参数编码增强与波形编码增强之间的基于“盲”时间SNR的混和。在此上下文中,“盲”还表示切换不是由复杂听觉掩蔽模型(例如,要在本文中描述的类型的)感知地指引,而是由与节目的片段相对应的SNR值序列指引。

[0051] 实现基于“盲”时间SNR的混和的实施方式通常包括以下步骤:将未增强的音频信号(原始音频混合)分割成连续的时间片(片段);针对每个片段来确定片段的语音内容与其他音频内容之间(或者语音内容与总音频内容之间)的SNR;以及针对每个片段来设置混和控制指示符,其中,混和控制指示符的值由片段的SNR确定(是片段的SNR的函数)。

[0052] 在一些实施方式中,方法包括确定(例如,接收请求)语音增强的总量(“T”)的步骤,混和控制指示符是使得T=αPw+(1-α)Pp的每个片段的参数α,其中,Pw是下述的片段的波形编码增强:如果使用针对片段所设置的波形数据将该片段的波形编码增强应用于片段的未增强的音频内容则将产生预定的总增强量T(其中,片段的语音内容具有未增强的波形,片段的波形数据指示片段的语音内容的降低品质的版本,该降低品质的版本具有与未增强的波形类似(例如,至少基本上类似)的波形,并且当被单独地呈现和感知时,语音内容的降低品质的版本具有令人讨厌的品质),Pp是下述的参数编码增强:如果使用针对片段所设置的参数数据将该参数编码增强应用于片段的未增强的音频内容则将产生预定总增强量T(其中,片段的参数数据与片段的未增强的音频内容一起来确定片段的语音内容的参数重构版本)。在一些实施方式中,片段中的每一个的混和控制指示符是包括相关片段的每个频带的参数的这样的参数的集合。

[0053] 当未增强的音频信号与作为元数据的控制参数一起被递送(例如,被发送)至接收器时,接收器可以(对每个片段)执行由片段的控制参数所指示的混合语音增强。替选地,接收器根据未增强的音频信号生成控制参数。

[0054] 在一些实施方式中,接收器(对未增强的音频信号的每个片段)执行参数编码增强

(以通过由片段的参数α所缩放的增强Pp所确定的量)和波形编码增强(以通过由片段的值(1-α)所缩放的增强Pw所确定的量)的组合,使得参数编码增强与波形编码增强的组合生成预定的总增强量:

[0055] $T=\alpha Pw+(1-\alpha)Pp$ (1)

[0056] 在另一类实施方式中,通过听觉掩蔽模型来确定要对音频信号的每个片段执行的波形编码增强和参数编码增强的组合。在该类的一些实施方式中,要对音频节目的片段执行的波形编码增强和参数编码增强的混和的最佳混和比率使用刚好防止编码噪声变得可听见的最高的波形编码增强量。应当理解,解码器中的编码噪声可得性总是统计估计的形式,并且不能被精确地确定。

[0057] 在该类中的一些实施方式中,音频数据的每个片段的混和指示符指示要对片段执行的波形编码增强和参数编码增强的组合,并且该组合至少基本上等于由听觉掩蔽模型针对片段所确定的波形编码最大化组合,其中,波形编码最大化组合指定确保语音增强音频节目的相应片段中的编码噪声(由于波形编码增强而引起)并非令人讨厌地听得见(例如,听不见的)的最大相对波形编码增强量。在一些实施方式中,确保语音增强音频节目的片段中的编码噪声不听起来令人讨厌的最大相对形编码增强量是以下最大相对量,该最大相对量确保(对音频数据的相应片段)要执行的波形编码增强和参数编码增强的组合生成片段的预定总量的语音增强,和/或(其中,参数编码增强的伪声被包括在由听觉掩蔽模型所执行的评估中)其可以使得(由于波形编码增强而引起的)编码伪声能够超过参数编码增强的伪声而听得见(当这是良好的时)(例如,在(由于波形编码增强而引起的)听得见的编码伪声与参数编码增强的听得见的伪声相比而较不令人讨厌的情况下)。

[0058] 在通过使用听觉掩蔽模型来更精确地预测降低品质的语音复本(要用于实现波形编码增强)中的编码噪声如何被主要节目的音频混合掩蔽并且据此选择混和比率,来确保编码噪声不变得令人讨厌地听得见(例如,不变得听得见)的同时,可以增大本发明的混合编码方案中的波形编码增强的贡献。

[0059] 使用听觉掩蔽模型的一些实施方式包括以下步骤:将未增强音频信号(原始音频混合)分割成连续的时间片(片段);提供每个片段中的语音的降低品质的复本(用于波形编码增强)以及每个片段的参数编码增强参数(用于参数编码增强);对于每个片段,使用听觉掩蔽模型来确定在编码伪声不变得令人讨厌地听得见的情况下可以应用的最大量的波形编码增强;以及生成波形编码增强(以不超过使用片段的听觉掩蔽模型所确定的最大量的波形编码增强以及至少基本上与使用片段的听觉掩蔽模型所确定的最大量的波形编码增强匹配的量)和参数编码增强的组合的指示符(针对未增强音频信号的每个片段),使得波形编码增强和参数编码增强的组合生成片段的预定总量的语音增强。

[0060] 在一些实施方式中,每个指示符被包括(例如,由编码器)在比特流中,该比特流还包括指示未增强音频信号的编码音频数据。

[0061] 在一些实施方式中,未增强音频信号被分割成连续的时间片并且每个时间片被分割成频带,对于每个时间片中的每个频带,使用听觉掩蔽模型确定在编码伪声不变得令人讨厌地听得见的情况下可以应用的最大量的波形编码增强,针对未增强音频信号的每个时间片的每个频带生成指示符。

[0062] 可选地,方法还包括以下步骤:响应于每个片段的指示符来(对未增强音频信号的

每个片段)执行由指示符所确定的波形编码增强和参数编码增强的组合,使得波形编码增强和参数编码增强的组合生成片段的预定总量的语音增强。

[0063] 在一些实施方式中,将音频内容编码在诸如环绕声配置、5.1扬声器配置、7.1扬声器配置、7.2扬声器配置等的参考音频通道配置(或表示)的编码音频信号中。参考配置可以包括音频通道如立体声通道、左前通道和右前通道、环绕通道、扬声器通道、对象通道等。携载语音内容的通道中的一个或更多个可以不是中间/侧(M/S)音频通道表示的通道。如本文中所使用的,M/S音频通道表示(或简称为M/S表示)包括至少中间通道和侧通道。在示例实施方式中,中间通道表示左通道和右通道(例如,等同地被加权等)之和,而侧通道表示左通道和右通道之差,其中,左通道和右通道可以被视为两个通道例如前中央通道和前左通道的任意组合。

[0064] 在一些实施方式中,节目的语音内容可以与非语音内容混合,并且可以被分布在参考音频通道配置中的两个或更多个非M/S通道如左通道和右通道、左前通道和右前通道等上。语音内容可以但并不要求被表示在立体声内容中的幻象中心处,在所述立体声内容中,语音内容在两个非M/S通道如左通道和右通道等中同样响亮。立体声内容可以包括不一定同样响亮或者甚至出现在两个通道中的非语音内容。

[0065] 在一些方法中,用于与在其上分布有语音内容的多个非M/S音频通道相对应的用于语音增强的非M/S控制数据、控制参数等的多个集合作为全部音频元数据的一部分从音频编码器被发送至下游音频解码器。用于语音增强的非M/S控制数据、控制参数等的多个集合中的每一个与在其上分布有语音内容的多个非M/S音频通道的特定音频通道相对应,并且可以由下游音频解码器使用来控制与特定音频通道有关的语音增强操作。如本文中所使用的,非M/S控制数据、控制参数等的集合指代用于非M/S表示如在其中如本文中所描述的音频信号被编码的参考配置的音频通道中的语音增强操作的控制数据、控制参数等。

[0066] 在一些实施方式中,M/S语音增强元数据——除了非M/S控制数据、控制参数等的一个或更多个集合以外或者代替非M/S控制数据、控制参数等的一个或更多个集合——作为音频元数据的一部分从音频编码器被发送至下游音频解码器。M/S语音增强元数据可以包括用于语音增强的M/S控制数据、控制参数等的一个或更多个集合。如本文中所使用的,M/S控制数据、控制参数等的集合指代用于M/S表示的音频通道中的语音增强操作的控制数据、控制参数等。在一些实施方式中,用于语音增强的M/S语音增强元数据与编码在参考音频通道配置中的混合内容一起被音频编码器发送至下游音频解码器。在一些实施方式中,用于M/S语音增强元数据中的语音增强的M/S控制数据、控制参数等的集合的数目可以比在其上分布有混合内容中的语音内容的参考音频通道表示中的多个非M/S音频通道的数目少。在一些实施方式中,甚至当混合内容中的语音内容被分布在参考音频通道配置中的两个或更多个非M/S音频通道如左通道和右通道等上时,用于语音增强的M/S控制数据、控制参数等的仅一个集合——例如,与M/S表示的中间通道相对应——作为M/S语音增强元数据被音频编码器发送至下游解码器。可以使用用于语音增强的M/S控制数据、控制参数等的单个集合来实现针对两个或更多个非M/S音频通道如左通道和右通道等中的所有通道的语音增强操作。在一些实施方式中,可以使用参考配置与M/S表示之间的转换矩阵来应用用于如本文中所描述的语音增强的基于M/S控制数据、控制参数等的语音增强操作。

[0067] 如本文中所描述的技术可以用于以下情况中:语音内容被平移在左通道和右通道

的幻象中心处,语音内容未被完全平移至中央(例如,左通道和右通道两者中不同样响亮)等。在示例中,这些技术可以用于以下情况中:语音内容的大百分比(例如,70+%、80+%、90+%等)的能量在中间信号或M/S表示的中间通道中。在另一个示例中,(例如,空间等)转换如平移、旋转等可以用来将参考配置中的不等同的语音内容转换成M/S配置中的等同或基本上等同的语音内容。表示平移、旋转等的呈现向量、转换矩阵等可以用作语音增强操作的一部分或者可以与语音增强操作结合使用。

[0068] 在一些实施方式中(例如,混合模式等),语音内容的版本(例如,降低的版本等)作为M/S表示中的仅中间通道信号或者中间通道信号和侧通道信号两者,连同可能具有非M/S表示的参考音频通道配置中所发送的混合内容一起被发送至下游音频解码器。在一些实施方式中,当语音内容的版本作为M/S表示中的仅中间通道信号被发送至下游音频解码器时,对中间通道信号进行操作(例如,执行转换等)以基于中间通道信号来生成非M/S音频通道配置(例如,参考配置等)的一个或更多个非M/S通道中的信号部分的、相应呈现向量也被发送至下游音频解码器。

[0069] 在一些实施方式中,实现音频节目的片段的参数编码增强(例如,独立通道对话预测、多通道对话预测等)与波形编码增强之间的基于“盲”时间SNR切换的对话/语音增强算法(例如,在下游音频解码器等中)至少部分地在M/S表示中操作。

[0070] 如本文中所描述的至少部分地在M/S表示中实现语音增强操作的技术可以用于独立通道预测(例如,在中间通道等中)、多通道预测(例如,在中间通道和侧通道等中)等。这些技术还可以用来同时支持对一个对话、两个或更多个对话的语音增强。控制参数、控制数据等如预测参数、增益、呈现向量等的零个集合、一个或更多个另外的集合可以作为M/S语音增强元数据的一部分被设置在编码音频信号中以支持另外的对话。

[0071] 在一些实施方式中,(例如,从编码器输出等的)编码音频信号的语义支持M/S标记从上游音频编码器至下游音频解码器的传输。当要至少部分地使用利用M/S标记所发送的M/S控制数据、控制参数等来执行语音增强操作时,M/S标记出现/被设置。例如,当M/S标记被设置时,在根据语音增强算法(例如,独立通道对话预测、多通道对话预测、基于波形的、波形参数混合等)中的一个或更多个、使用如利用M/S标记所接收的M/S控制数据、控制参数等应用M/S语音增强操作之前,接收方音频解码器可以首先将非M/S通道中的立体声信号(例如,来自左通道和右通道等)转换成M/S表示的中间通道和侧通道。在执行M/S语音增强操作之后,可以将M/S表示中的语音增强信号转换回至非M/S通道。

[0072] 在一些实施方式中,要根据本发明来增强其语音内容的音频节目包括扬声器通道但是不包括任何对象通道。在其他实施方式中,要根据本发明增强其语音内容的音频节目是包括至少一个对象通道以及可选地至少一个扬声器通道的基于对象的音频节目(典型地为基于多通道对象的音频节目)。

[0073] 本发明的另一个方面是以下系统,该系统包括:编码器,其被配置(例如,被编程)为响应于指示包括语音内容和非语音内容的节目的音频数据,执行本发明编码方法的任何实施方式以生成包括编码音频数据、波形数据和参数数据(以及此外可选地音频数据的每个片段的混和指示符(例如,混和指示数据))的比特流;以及解码器,其被配置成对比特流进行解析以恢复编码音频数据(以及此外可选地每个混和指示符)并且对编码音频数据进行解码以恢复音频数据。替代地,解码器被配置成响应于所恢复的音频数据而生成音频数

据的每个片段的混和指示符。解码器被配置成响应于每个混和指示符对所恢复的音频数据执行混合语音增强。

[0074] 本发明的另一个方面是被配置成执行本发明方法的任何实施方式的解码器。在另一类实施方式中,本发明是包括存储(例如,以非暂态方式)已经通过本发明方法的任何实施方式所生成的编码音频比特流的至少一个片段(例如,帧)的缓冲存储器(缓冲器)的解码器。

[0075] 本发明的其他方面包括被配置(例如,被编程)成执行本发明方法的任何实施方式的系统或装置(例如,编码器、解码器或处理器)以及存储用于实现本发明方法或其步骤的任何实施方式的代码的计算机可读介质(例如,磁盘)。例如,本发明系统可以是或者包括使用软件或固件被编程成和/或以其他方式被配置成对数据执行包括本发明方法或其步骤的实施方式的多种操作中的任何操作的可编程通用处理器、数字信号处理器或微处理器。这样的通用处理器可以是或者包括以下计算机系统,该计算机系统包括被编程(和/或以其他方式被配置)成响应于设定(assert)至该计算机系统的数据来执行本发明方法(或其步骤)的实施方式的输入装置、存储器和处理电路。

[0076] 在一些实施方式中,如本文中所描述的机构形成媒体处理系统的一部分,包括但不限于:音视频装置、平板TV、手持装置、游戏机、电视、家庭影院系统、平板、移动装置、膝上型计算机、笔记本计算机、蜂窝无线电话、电子书阅读器、销售端的点、桌面型计算机、计算机工作站、计算机信息站、各种其他种类的终端和媒体处理单元等。

[0077] 对本领域的技术人员而言,对本文中所描述的一般原理和特征和优选实施方式的各种修改将是显而易见的。因此,本公开内容并不意在受限于所示的实施方式,而是意在符合与本文中所描述的原理和特征一致的最宽的范围。

[0078] 2.符号和术语

[0079] 贯穿包括权利要求在内的本公开内容,术语“对话”和“语音”作为同义词可互换地被用来表示作为由人类(或者虚拟世界中的角色)沟通的形式所感知的音频信号内容。

[0080] 贯穿包括权利要求在内的本公开内容,表达“对”信号或数据执行操作(例如,对信号或数据进行滤波、缩放、转换、或者应用增益)在广义上被用来表示对信号或数据直接执行操作或者对信号或数据的经处理的版本(例如,对在对其执行操作之前已经经历初步滤波或预处理的信号的版本)执行操作。

[0081] 贯穿包括权利要求在内的本公开内容,表达“系统”在广义上被用来表示装置、系统或子系统。例如,实现解码器的子系统可以被称为解码器系统,包括这样的子系统(例如,响应于多个输入生成X输出信号的系统,其中,子系统生成M个输入,从外部源接收另外X-M个输入)的系统还可以被称为解码器系统。

[0082] 贯穿包括权利要求在内的本公开内容,术语“处理器”在广义上被用来表示可编程或者以其他方式可配置(例如,使用软件或固件)成对数据(例如,音频、或者视频或其他图像数据)执行操作的系统或装置。处理器的示例包括现场可编程门阵列(或其他可配置集成电路或芯片组)、被编程和/或以其他方式被配置成对音频或其他声音数据执行流水线处理的数字信号处理器、可编程通用处理器或计算机、以及可编程微处理器芯片或芯片组。

[0083] 贯穿包括权利要求在内的本公开内容,表达“音频处理器”和“音频处理单元”可互换地被使用,并且在广义上,表示被配置成处理音频数据的系统。音频处理单元的示例包括

但不限于编码器(例如,转码器)、解码器、编解码器、预处理系统、后处理系统、以及比特流处理系统(有时称为比特流处理工具)。

[0084] 贯穿包括权利要求在内的本公开内容,表达“元数据”指代与相应音频数据(还包括元数据的比特流的音频内容)分立且不同的数据。元数据与音频数据相关联,并且表示音频数据的至少一个特征或特性(例如,已经对音频数据或者由音频数据所指示的对象的轨迹执行了什么类型的处理或者应该执行什么类型的处理)。元数据与音频数据的关联是时间同步的。因此,当前(最近所接收或更新的)元数据可以指示相应音频数据同时具有所指示的特征和/或包括音频数据处理的所指示类型的结果。

[0085] 贯穿包括权利要求在内的本公开内容,术语“耦接(couples)”或“耦接(coupled)”被用来表示直接或间接连接。因此,如果第一装置耦接至第二装置,则连接可以通过直接连接或者通过经由其他装置和连接的间接连接。

[0086] 贯穿包括权利要求在内的本公开内容,以下表达具有下面的定义:

[0087] -扬声器(speaker)和扩音器(loudspeaker)同义地被用来表示任何发出声音的转换器。该定义包括实现为多个转换器(例如,低频扬声器和高频扬声器)的扩音器;

[0088] -扬声器馈送:要直接应用于扩音器的音频信号,或者要应用于串联的放大器和扩音器的音频信号;

[0089] -通道(或“音频通道”):单通道音频信号。通常,这样的信号可以以这样的方式被呈现,使得等同于将信号直接应用于在期望位置或标称位置处的扩音器。如通常是具有物理扩音器的情况,期望位置可以是静止的,或可以是动态的;

[0090] -音频节目:一个或更多个音频通道的集合(至少一个扬声器通道和/或至少一个对象通道)以及此外可选地相关联的元数据(例如,描述期望的空间音频表示的元数据);

[0091] -扬声器通道(或者“扬声器馈送通道”):与命名扩音器(在期望位置或标称位置处)相关联的或者与限定的扬声器配置内的命名扬声器区相关联的音频通道。扬声器通道以这样的方式被呈现,使得等同于直接向命名扩音器(在期望位置或标称位置处)或者命名扬声器区中的扬声器应用音频信号;

[0092] -对象通道:指示由音频源(有时称为音频“对象”)发出的声音的音频通道。通常,对象通道确定参数音频源描述(例如,指示参数音频源描述被包括在对象通道中或者设置有对象通道的元数据)。源描述可以确定由源发出的声音(作为时间的函数)、作为时间的函数的源的表观位置(例如,三维空间坐标)、以及可选地至少一个表征源的附加参数(例如,表观源大小或宽度);

[0093] -基于对象的音频节目:包括一个或更多个对象通道的集合(以及此外可选地包括至少一个扬声器通道)以及此外可选地相关联的元数据(例如,指示发出由对象通道所指示的声音的音频对象的轨迹的元数据,或者以其他方式指示由对象通道所指示的声音的期望的空间音频表示的元数据,或者指示作为由对象通道所指示的声音的源的至少一个音频对象的标识的元数据)的音频节目;以及

[0094] -呈现:将音频节目转变成一个或更多个扬声器馈送的处理,或者将音频节目转变成一个或更多个扬声器馈送并且使用一个或更多个扩音器将该扬声器馈送转变成声音的处理(在后一种情况下,在本文中呈现有时被称为“由”扩音器呈现)。可以通过直接对在期望位置处的物理扬声器应用信号来平常地呈现(“在”期望位置处)音频通道,或者可以使用

要被设计成基本上等同于(对于听者而言)这样的平常呈现的多个虚拟化技术之一来呈现一个或更多个音频通道。在该后一种情况下,每个音频通道可以被转变成要应用于一般不同于期望位置的已知位置中的扩音器的一个或更多个扬声器馈送,使得由扩音器响应于馈送所发出的声音将被感知为从期望位置发出。这样的虚拟化技术的示例包括经由耳机的双耳呈现(例如,使用为耳机佩戴者模拟高达7.1环绕声通道的杜比耳机处理)以及波场合成。

[0095] 将参照图3、图6和图7来描述本发明的编码、解码和语音增强方法的实施方式以及被配置成实现方法的系统。

[0096] 3.预测参数的生成

[0097] 为了执行语音增强(包括根据本发明的实施方式的混合语音增强),需要访问要增强的语音信号。如果在要执行语音增强时语音信号不可用(与要增强的混合信号的语音内容和非语音内容的混合分立),则可以使用参数技术来创建可用混合的语音的重构。

[0098] 一种用于混合内容信号(指示语音内容与非语音内容的混合)的语音内容的参数重构的方法基于重构信号的每个时间-频率分块中的语音功率,并且根据以下公式生成参数:

[0099] $$p_{n,b}=\sqrt{\sum_{s\in n,f\in b}\frac{D_{s,f}^{2}}{M_{s,f}^{2}}} \qquad (2)$$

[0100] 其中,$p_{n,b}$是分块的参数(参数编码语音增强值),$p_{n,b}$具有时间索引n和频率带索引b,值$D_{s,f}$表示分块的时隙s和频率仓(bin)f中的语音信号,值$M_{s,f}$表示分块的同一时隙和频率仓中的混合内容信号,求和针对所有分块中的s和f的所有值。可以使用混合内容信号自身来递送(作为元数据)参数$p_{n,b}$,以使得接收器能够重构混合内容信号的每个片段的语音内容。

[0101] 如图1所描绘的,可以通过以下操作来确定每个参数$p_{n,b}$:对要增强的其语音内容的混合内容信号(“混合音频”)执行时域到频域的转换;对语音信号(混合内容信号的语音内容)执行时域到频域的转换;在分块中的所有时隙和频率仓对(具有语音信号的时间索引n和频率带索引b的每个时间-频率分块的)能量求积分;关于分块中的所有时隙和频率仓对混合内容信号的相应时间-频率分块的的能量求积分;以及将第一积分的结果除以第二积分的结果以生成分块的参数$p_{n,b}$。

[0102] 当将混合内容信号的每个时间-频率分块乘以分块的参数$p_{n,b}$时,所得到信号具有与混合内容信号的语音内容相似的频谱和时间包络。

[0103] 典型音频节目——例如立体声或5.1通道音频节目——包括多个扬声器通道。通常,每个通道(或者通道的子集中的每一个)指示语音内容和非语音内容,并且混合内容信号确定每个通道。可以将所描述的参数语音重构方法独立地应用于每个通道以重构所有通道的语音内容。可以使用每个通道的适当的增益将重构语音信号(针对通道中的每一个有一个重构语音信号)添加至相应混合内容通道信号,以获得对语音内容的期望的增强。

[0104] 多通道节目的混合内容信号(通道)可以被表示为信号向量的集合,其中,每个向量元素是与特定参数集合即帧(n)中的时隙(s)和参数带(b)中的所有频率仓(f)相对应的时间-频率分块的汇集。三通道混合内容信号的向量的这样的集合的示例是:

[0105] $$M_{n,b}=\begin{pmatrix}M_{c_1,n,b}\\M_{c_2,n,b}\\M_{c_3,n,b}\end{pmatrix} \quad (3)$$

[0106] 其中,c_i表示通道。该示例假定三个通道,但是通道的数目是任意量。

[0107] 类似地,多通道节目的语音内容可以被表示为1×1矩阵的集合(其中,语音内容包括仅一个通道)$D_{n,b}$。混合内容信号的每个矩阵元素与标量值的乘法产生每个子元素与标量值的乘积。因此,通过针对每个n和b计算下面的公式来获得每个分块的重构语音值

[0108] $$D_{r,n,b}=\mathrm{diag}(P)\cdot M_{n,b} \quad (4)$$

[0109] 其中,P是其元素是预测参数的矩阵。(所有分块的)重构语音还可以被表示为:

[0110] $$D_r=\mathrm{diag}(P)\cdot M \quad (5)$$

[0111] 多通道混合内容信号的多个通道中的内容引起可以使用其对语音信号做出较好的预测的通道之间相干。通过使用(例如,常规类型的)最小均方差(MMSE)预测器,可以将通道与预测参数进行组合以根据均方差(MSE)标准使用最小误差来重构语音内容。如图2所示,假定三通道混合内容输入信号,这样的MMSE预测器(在频域中操作)响应于混合内容输入信号以及指示混合内容输入信号的语音内容的单个输入语音信号来迭代地生成预测参数p_i(其中,索引i是1、2或3)的集合。

[0112] 根据混合内容输入信号的每个通道的分块(具有相同的索引n和索引b的每个分块)所重构的语音值是由每个通道的权重参数所控制的混合内容信号的每个通道(i=1,2或3)的内容($M_{ci,n,b}$)的线性组合。这些权重参数是具有相同的索引n和b的分块的预测参数p_i。因此,根据混合内容信号的所有通道的所有片重构的语音是:

[0113] $$D_r=p_1\cdot M_{c1}+p_2\cdot M_{c2}+p_3M_{c3} \quad (6)$$

[0114] 或者以下面的信号矩阵形式:

[0115] $$D_r=PM \quad (7)$$

[0116] 例如,当语音在混合内容信号的多个通道中相干地呈现而背景(非语音)声音在通道之间不相干时,通道的相加组合将有利于语音的能量。与通道独立重构相比,对于两个通道,这将导致3dB更好的语音分离。作为另一个示例,当语音内容在一个通道中呈现并且背景声音在多个通道中相干呈现时,通道的相减组合将(部分地)消除背景声音,而保留语音。

[0117] 在一类实施方式中,本发明方法包括以下步骤:(a)接收指示包括具有未增强的波形的语音以及其他音频内容的音频节目的比特流,其中,比特流包括:指示语音内容和其他音频内容的未增强的音频数据;指示语音的降低品质版本的波形数据,其中,语音的降低品质版本具有与未增强的波形相似(例如,至少基本上相似)的第二波形,并且如果单独地被试听则降低品质版本将具有令人讨厌的品质;以及参数数据,其中,与未增强音频数据一起的参数数据确定参数创建语音,并且该参数重构语音是至少基本上与语音匹配(例如,是良好近似)的、语音的参数重构版本;以及(b)响应于混和指示符对比特流执行语音增强,从而生成指示语音增强音频节目的数据,包括通过将未增强的音频数据与根据波形数据所确定的低品质语音数据和重构语音数据的组合进行组合,其中,该组合由混和指示符(例如,该组合具有由混和指示符的当前值序列所确定的状态序列)确定,重构的语音数据响应于参数数据中的至少一些以及未增强音频数据中的至少一些而生成,与通过将仅低品质语音数据与未增强的音频数据进行组合确定的纯波形编码语音增强音频节目或者根据参数数据

和未增强的音频数据所确定的纯参数编码语音增强音频节目相比，语音增强音频节目具有不太听得见语音增强编码伪声(例如，更好地被掩蔽的语音增强编码伪声)。

[0118] 在一些实施方式中，混和指示符(其可以具有值序列，例如针对比特流片段序列中的每一个的一个值序列)被包括在步骤(a)中所接收的比特流中。在其他实施方式中，混和指示符响应于比特流而生成(例如，在接收比特流并且对比特流进行解码的接收器中)。

[0119] 应当理解，表达“混和指示符”并不意在表示比特流的每个片段的单个参数或值(或者单个参数或值序列)。相反地，可以想到，在一些实施方式中，(比特流的片段的)混和指示符可以是两个或更多个参数或值的集合(例如，对于每个片段，参数编码增强控制参数和波形编码增强控制参数)。在一些实施方式中，每个片段的混和指示符可以是指示每片段的频带进行混和的值序列。

[0120] 无需为(例如，被包括在)比特流的每个片段设置波形数据和参数数据，或者无需被用于对比特流的每个片段执行语音增强。例如，在一些情况下，至少一个片段可以包括仅波形数据(以及由每个这样的片段的混和指示符所确定的组合可以包括仅波形数据)并且至少一个另外的片段可以包括仅参数数据(以及由每个这样的片段的混和指示符所确定的组合可以包括仅重构语音数据)。

[0121] 可以想到，在一些实施方式中，编码器生成比特流，包括通过对未增强音频数据而非波形数据或参数数据进行编码(例如，压缩)。因此，当比特流被递送至接收器时，接收器将对比特流进行解析以提取未增强的音频数据、波形数据以及参数数据(如果其在比特流中被递送，则以及混和指示符)，但是将对仅未增强的音频数据进行解码。在不对波形数据或参数数据应用与对音频数据应用的解码处理相同的解码处理的情况下，接收器将对经解码的、未增强的音频数据(使用波形数据和/或参数数据)执行语音增强。

[0122] 通常，波形数据与重构语音数据的组合(由混和指示符所指示)随时间而变化，具有与比特流的相对应的片段的语音内容和其他音频内容有关的每个组合状态。混和指示符被生成为：使得(波形数据和重构语音数据的)当前组合状态由比特流的相应片段中的语音内容和其他音频内容(例如，语音内容的功率与其他音频内容的功率的比)的信号特性确定。

[0123] 步骤(b)可以包括以下步骤：通过将低品质语音数据中的至少一些与比特流的至少一个片段的未增强的音频数据进行组合(例如，混合或混和)执行波形编码语音增强；以及通过将重构语音数据与比特流的至少一个片段的未增强的音频数据进行组合执行参数编码语音增强。通过将片段的低品质语音数据和重构语音数据两者与片段的未增强的音频数据进行混和对比特流的至少一个片段执行波形编码语音增强与参数编码语音增强的组合。在一些信号条件下，对比特流的片段(或者对多于一个片段中的每一个)执行(响应于混和指示符)波形编码语音增强和参数编码语音增强中的仅一个(而不是两者)。

[0124] 4.语音增强操作

[0125] 在本文中，“SNR”(信噪比)被用来表示对音频节目(或整个节目)的片段的语音分量(即，语音内容)的功率(或水平)与片段或节目的非语音分量(即，非语音内容)的功率(或水平)之比，或者与片段或节目的整个(语音和非语音)内容的功率(或水平)之比。在一些实施方式中，根据音频信号(以经历语音增强)以及指示音频信号的语音内容(例如，为了在波形编码增强中使用已经生成的语音内容的低品质复本)的分立的信号导出SNR。在一些实施

方式中,根据音频信号(以经历语音增强)并且根据参数数据(其为了在音频信号的参数编码增强中使用已经被生成)导出SNR。

[0126] 在一类实施方式中,本发明方法实现音频节目的片段的参数编码增强与波形编码增强之间基于“盲”时间SNR切换。在本上下文中,“盲”表示切换并不由(例如,本文中要描述的类型的)复杂听觉掩蔽模型感知地指引,而是由与节目的片段相对应的SNR值序列(混和指示符)指引。在该类的一种实施方式中,通过参数编码增强与波形编码增强(响应于混和指示符,例如,在图3的编码器的子系统29中所生成的混和指示符,其指示应当对相应音频数据执行仅参数编码增强或者波形编码增强)之间的时间切换实现混合编码语音增强,使得对执行了语音增强的音频节目的每个片段执行参数编码增强或者波形编码增强(而非参数编码增强和波形编码增强两者)。意识到在低SNR(对具有低SNR值的片段)的条件下波形编码增强表现地最好并且在良好的SNR(对具有高SNR值的片段)的条件下参数编码增强表现地最好,切换决定通常基于原始音频混合中的语音(对话)与剩余音频的比。

[0127] 实现基于“盲”时间SNR的切换的实施方式通常包括以下步骤:将未增强的音频信号(原始音频混合)分割成连续时间片(片段),为每个片段确定片段的语音内容与其他音频内容之间(或者语音内容与总音频内容之间)的SNR;以及对于每个片段,将SNR与阈值进行比较并且当SNR大于阈值时为片段设置参数编码增强控制参数(即,片段的混和指示符指示应当执行参数编码增强),或者当SNR不大于阈值时为参数设置波形编码增强控制参数(即,片段的混和指示符指示应当执行波形编码增强)。

[0128] 当未增强音频信号与作为元数据所包括的控制参数一起被递送(例如,发送)至接收器时,接收器可以(对每个片段)执行由片段的控制参数所指示的语音增强的类型。因此,接收器对控制参数是参数编码增强控制参数的每个片段执行参数编码增强,并且对控制参数是波形编码增强控制参数的每个片段执行波形编码增强。

[0129] 如果愿意承担传输(关于原始音频混合的每个片段)波形数据(用于实现波形编码语音增强)以及关于原始(未增强)混合的参数编码增强参数两者的成本,那么可以通过对混合的各个分量应用波形编码增强和参数编码增强两者实现较高程度的语音增强。因此,在一类实施方式中,本发明方法实现音频节目的片段的参数编码增强与波形编码增强之间的基于“盲”时间SNR混合。此外,在此上下文中,“盲”表示切换并不由(例如,本文中要描述的类型的)复杂听觉掩蔽模型感知地指引,而是由与节目的片段相对应的SNR值序列指引。

[0130] 实现基于的“盲”时间SNR混和的实施方式通常包括以下步骤:将未增强音频信号(原始音频混合)分割成连续时间片(片段),并且为每个片段确定片段的语音内容与其他音频内容之间(或者语音内容与总音频内容之间)的SNR;确定(例如,接收请求)语音增强的总量(“T”);以及为每个片段设置混和控制参数,求中,混和控制参数的值由片段的SNR确定(是片段的SNR的函数)。

[0131] 例如,音频节目的片段的混和指示符可以是在图3的编码器的子系统29中为片段所生成的混和指示符参数(或参数集合)。

[0132] 混和控制指示符可以是使得$T=\alpha Pw+(1-\alpha)Pp$的每个片段的参数α,其中,Pw是下述的波形的波形编码增强:如果使用针对片段所设置的波形数据将该波形的波形编码增强应用于片段的未增强音频内容则将产生预定的总增强量T(其中,片段的语音内容具有未增强的波形,片段的波形数据指示片段的语音内容的降低品质的版本,降低品质的版本具有与

未增强波形相似(例如,至少基本上相似)的波形,当被单独地呈现和感知时,语音内容的降低品质的版本具有令人讨厌的品质),Pp是下述的参数编码增强:如果使用针对片段所设置的参数数据将该参数编码增强应用于片段的未增强音频内容则将产生预定的总增强量T(其中,片段的参数数据与片段的未增强音频内容一起来确定片段的语音内容的参数重构版本)。

[0133] 当未增强音频信号与作为元数据的控制参数一起被递送(例如,发送)至接收器时,接收器可以(对每个片段)执行由片段的控制参数所指示的混合语音增强。替代地,接收器根据未增强音频信号生成控制参数。

[0134] 在一些实施方式中,接收器(对未增强音频信号的每个片段)执行参数编码增强Pp(由片段的参数α缩放)与波形编码增强Pw(由片段的值(1-α)缩放)的组合,使得所缩放的参数编码增强和所缩放的波形编码增强的组合生成如表达式(1)(T=αPw+(1-α)Pp)中的预定总量的增强。

[0135] 片段的SNR与α之间的关系的示例如下:α是SNR的非递减函数,α的范围是0到1,当片段的SNR小于或等于阈值("SNR_poor")时,α的值为0,当SNR大于或等于较大阈值("SNR_high")时,α的值为1。当SNR良好时,α高,导致很大部分的参数编码增强。当SNR不良时,α低,导致很大部分的波形编码增强。应当选择饱和点的位置(SNR_poor和SNR_high)以调节波形编码增强算法和参数编码增强算法两者的具体实现。

[0136] 在另一类实施方式中,要对音频信号的每个片段执行的波形编码增强和参数编码增强的组合由听觉掩蔽模型确定。在该类的一些实施方式中,要对音频节目的片段执行的波形编码增强和参数编码增强的混和的最佳混和比率使用刚好使编码噪声不变得听得见的最高波形编码增强量。

[0137] 在上述基于盲SNR的混和实施方式中,从SNR获得片段的混和比率,SNT被假定成指示掩蔽要为波形编码增强所使用的语音的降低品质版本(复本)中的编码噪声的音频混合的能力。基于盲SNR方法的优点是实现的简单性以及编码器处的低计算负荷。然而,SNR是以下不可靠的预测器:编码噪声在多大程度上将被掩蔽以及必须在多大程度上应用大的安全裕度以确保编码噪声将一直仍然被掩蔽。这意味着至少一些时间被混和的降低品质语音复本的水平低于其能够达到的水平,或者如果将裕度设置地较严格,则一些时候编码噪声变得听得见。当通过使用更准确地预测降低品质语音复本中的编码噪声如何被主要节目的音频混合掩蔽并且据此选择混和比率的听觉掩蔽模型确保编码噪声不变得听得见时,可以增大本发明的混和编码方案中的波形编码增强的贡献。

[0138] 使用听觉掩蔽模型的特定实施方式包括以下步骤:将未增强音频信号(原始音频混合)分割成连续时间片(片段),设置每个片段中的语音的降低品质复本(用于在波形编码增强中使用)以及每个片段的参数编码增强参数(用于在参数编码增强中使用);使用听觉掩蔽模型针对片段中的每一个来确定可以被应用但伪声不变得听得见的最大波形编码增强量;以及生成波形编码增强(以不超过使用听觉掩蔽模型针对片段所确定的最大波形编码增强量以及优选地至少基本上与使用听觉掩蔽模型针对片段所确定的最大波形编码增强量匹配的量)和参数编码增强的组合的混和指示符(未增强音频信号的每个片段的),使得波形编码增强与参数编码增强的组合生成片段的预定语音增强总量。

[0139] 在一些实施方式中,每个这样的混和指示符被包括(例如,由编码器)在还包括指

示未增强音频信号的编码音频数据的比特流中。例如,图3的编码器20的子系统29可以被配置成生成这样的混和指示符,编码器20的子系统28可以被配置成包括要从编码器20输出的比特流中的混和指示符。又例如,可以根据由图7编码器的子系统14所生成的$g_{max}(t)$参数生成混和指示符(例如,图7编码器的子系统13中),图7编码器的子系统13可以被配置成包括要从图7编码器输出的比特流中的混和指示符(或者子系统13可以包括要从图7编码器输出的比特流中的由子系统14所生成的$g_{max}(t)$参数,接收并解析比特流的接收器可以被配置成响应于$g_{max}(t)$参数生成混和指示符)。

[0140] 可选地,所述方法还包括的步骤:响应于每个片段的混和指示符(对未增强音频信号的每个片段)执行由混和指示符所确定的波形编码增强和参数编码增强的组合,使得波形编码增强和参数编码增强的组合生成片段的预定语音增强总量。

[0141] 将参照图7来描述使用听觉掩蔽模型的本发明方法的实施方式的示例。在该示例中,语音和背景音频的混和A(t)(未增强音频混合)被确定(在图7的元件10中)并且被传递至预测未增强音频混合的每个片段的掩蔽阈值$\Theta(f,t)$的听觉掩蔽模型(由图7的元件11实现)。未增强音频混合A(t)还被提供至用于编码的编码元件13以供传输。

[0142] 由模型所生成的掩蔽阈值指示为任何信号必须超过以成为听得见的频率和时间听觉激励的函数。这样的掩蔽模型是本领域公知的。对未增强音频混合A(t)的每个片段的语音分量s(t)进行编码(以低比特率音频编码器15)以生成片段的语音复本的降低品质复本s'(t)。降低品质复本s'(t)(与原始语音s(t)相比,其包括较少的比特)可以被概念化为原始语音s(t)与编码噪声n(t)之和。编码噪声可以通过从降低品质复本减去(在元件16中)时间对准语音信号s(t)与降低品质复本分离以供分析。

[0143] 在元件17中将编码噪声n与缩放因子g(t)相乘,并且将所缩放的编码噪声传递至预测由所缩放的编码噪声所生成的听觉激励N(f,t)的听觉模型(由元件18实现)。这样的激励模型是本领域已知的。在最终的步骤中,将听觉激励N(f,t)与所预测的掩蔽阈值$\Theta(f,t)$相比,并且确保编码噪声被掩蔽即确保$N(f,t)<\Theta(f,t)$的g(t)的最大值的最大缩放因子$g_{max}(t)$被找到(在元件14中)。如果听觉模型是非线性的,则可能需要通过在元件17中将向编码噪声应用的值g(t)迭代n(t)来迭代地进行上述操作(如图2所示);如果听觉模型是线性的,则可以在简单前馈步骤中进行上述操作。所得到的缩放因子$g_{max}(t)$是其被添加至未增强音频混合A(t)的相应片段而所缩放的、降低品质的语音复本中的编码伪声并不在所缩放的、降低品质的语音复本$g_{max}(t)*s'(t)$与未增强音频混合A(t)的混合中变得听得见之前可以向降低品质语音复本s'(t)应用的最大缩放因子。

[0144] 图7系统还包括元件12,该元件12被配置成(响应于未增强音频混合A(t)和语音s(t))生成用于对未增强音频混合的每个片段执行参数编码语音增强的参数编码增强参数p(t)。

[0145] 针对音频节目的每个片段的参数编码增强参数p(t)以及在编码器15中所生成的降低品质语音复本s'(t)和在元件14中所生成的因子$g_{max}(t)$也被设定至编码元件13。元件13生成指示针对音频节目的每个片段的未增强音频混合A(t)、参数编码增强参数p(t)、降低品质语音复本s'(t)和因子$g_{max}(t)$的编码音频比特流,并且该编码音频比特流可以被发送或以其他方式被递送至接收器。

[0146] 在该示例中,如下对未增强音频混合A(t)的每个片段(例如,在元件13的编码输出

已经被递送至的接收器中)执行语音增强,以使用片段的缩放因子$g_{max}(t)$应用预定的(例如,所要求的)总增强量T。对编码音频节目进行解码,以提取针对音频节目的每个片段的未增强音频混合A(t)、参数编码增强参数p(t)、降低品质语音复本s'(t)以及因子$g_{max}(t)$。对于每个片段,波形编码增强Pw被确定成下述的波形编码增强:如果使用片段的降低品质的语音复本s'(t)将该波形编码增强应用于片段的未增强音频内容则将产生预定的总增强量T。参数编码增强Pp被确定成下述参数编码增强:如果使用针对片段设置的参数数据将该参数编码增强应用于片段的未增强音频内容,则将产生预定的总增强量T(其中,关于片段的未增强音频内容,片段的参数数据确定片段的语音内容的参数重构版本)。对于每个片段,执行参数编码增强(以由片段的参数α_2缩放的量)与波形编码增强(以由片段的值α_1所确定的量)的组合,使得参数编码增强与波形编码增强的组合使用由以下模型所允许的最大波形编码增强量来生成预定总增强量:$T=(\alpha_1(Pw)+\alpha_2(Pp)$,在$T=(\alpha_1(Pw)+\alpha_2(Pp)$中,因子$\alpha_1$是不超过片段的$g_{max}(t)$并且使得能够实现所指示的等式($T=(\alpha_1(Pw)+\alpha_2(Pp))$的最大值,参数$\alpha_2$是使得能够实现所指示的等式($T=(\alpha_1(Pw)+\alpha_2(Pp))$的最小非负值。

[0147] 在替选实施方式中,参数编码增强的伪声被包括在(由听觉掩蔽模型执行的)评估中,以使得当(由于波形编码增强而引起的)编码伪声比参数编码增强的伪声有利时,其变得听得见。

[0148] 在对图7实施方式(以及类似于使用听觉掩蔽模型的图7的实施方式的实施方式)的变型中,有时被称为听觉模型指引的多带划分实施方式,降低品质语音复本的波形编码增强编码噪声N(f,t)与掩蔽阈值Θ(f,t)之间的关系可以跨所有频带而不一致。例如,波形编码增强编码噪声的频谱特征可以是使得在第一频率区中掩蔽噪声即将超过掩蔽阈值,而在第二频率区中掩蔽噪声远低于掩蔽阈值。在图7实施方式中,通过第一频率区中的编码噪声确定波形编码增强的最大贡献,并且通过第一频率区中的编码噪声和掩蔽特性来确定可被应用于降低品质的语音复本的最大缩放因子g。其小于在最大缩放因子的确定仅基于第二频率区的情况下可应用的最大缩放因子g。如果在两个频率区中分别应用时间混和的原理,则可以改进整体性能。

[0149] 在听觉模型指引的多带划分的一种实施方式中,未增强的音频信号被划分成M个连续的非交叠频带并且在M个带中的每一个中独立地应用时间混和的原理(即,根据本发明的实施方式使用波形编码增强与参数编码增强的混和的混合语音增强)。替选实现将频谱划分成截止频率fc以下的低频带以及截止频率fc以上的高频带。总是使用波形编码增强来增强低频带,并且总是使用参数编码增强来增强高频带。截止频率随着时间变化并且总是在以下约束下选择尽可能高的截止频率:在预定的总语音增强量T处的波形编码增强编码噪声在掩蔽阈值以下。换言之,在任意时刻的最大截止频率是:

[0150] $\max(fc\,|\,T*N(f<fc,t)<\Theta(f,t))$ (8)

[0151] 上述实施方式已经假定可用来防止波形编码增强编码伪声变得听见的方法调整(波形编码增强与参数编码增强的)混和比率,或者缩减总增强量。替选方法通过比特率的可变分配对波形编码增强编码噪声的量进行控制以生成降低品质语音复本。在该替选实施方式的示例中,应用参数编码增强的恒定基本量并且应用另外的波形编码增强以达到所期望的(预定的)总量增强。使用可变比特流对降低品质语音复本进行编码,并且该比特率被选作保持波形编码增强编码噪声在参数编码增强主音频的掩蔽阈值以下的最低比特率。

[0152] 在一些实施方式中,要根据本发明增强其语音内容的音频节目包括扬声器通道,但是不包括任何对象通道。在其他实施方式中,要根据本发明增强其语音内容的音频节目是包括至少一个对象通道以及此外可选地至少一个扬声器通道的基于对象的音频节目(通常多通道基于对象的音频节目)。

[0153] 本发明的其他方面包括:编码器,其被配置成执行本发明编码方法的任何实施方式,以响应于音频输入信号(例如,响应于指示多通道音频输入信号的音频数据)而生成编码音频信号;解码器,其被配置成对这样的编码信号进行解码并且对解码音频内容执行语音增强;以及包括这样的编码器和这样的解码器的系统。图3系统是这样的系统的示例。

[0154] 图3的系统包括编码器20,该编码器20被配置(例如,被编程)为执行本发明编码方法的实施方式,以响应于指示音频节目的音频数据而生成编码音频信号。通常,节目是多通道音频节目。在一些实施方式中,多通道音频节目包括仅扬声器通道。在其他实施方式中,多通道音频节目是包括至少一个对象通道以及此外可选地至少一个扬声器通道的基于对象的音频节目。

[0155] 音频数据包括指示混合音频内容(语音内容与非语音内容的混合)的数据(在图3中被标识为“混合音频”数据),以及指示混合音频内容的语音内容的数据(在图3中被标识为“语音”数据)。

[0156] 语音数据在级21中进行时域至频域(QMF)转换,所得到的QMF分量被设定至增强参数生成元件23。混合音频数据在级22中进行时域至频域(QMF)转换,所得到的QMF分量被设定至元件23并且被设定至编码子系统27。

[0157] 语音数据还被设定至被配置成生成指示语音数据的低品质复本的波形数据(在本文中有时被称为“降低品质”或者“低品质”语音复本)的子系统25,以供在由混合音频数据所确定的混合(语音与非语音)内容的波形编码语音增强中使用。与原始语音数据相比,低品质语音复本包括更少的比特,当单独地被呈现和感知时以及当呈现指示具有与由原始语音数据所指示的语音的波形相似(例如,至少基本上相似)的波形的语音时,低品质语音复本具有令人讨厌的品质。实现子系统25的方法是本领域已知的。示例是通常以低比特率(例如,20kbps)所操作的码激励线性预测(CELP)语音编码器如AMR和G729.1、或者现代混合编码器如MPEG统一语音和音频编码(USAC)。替选地,可以使用频域编码器,示例包括Siren(G722.1)、MPEG 2层II/III、MPEG AAC。

[0158] 根据本发明的典型实施方式所执行(例如,在解码器40的子系统43中)的混合语音增强包括以下步骤:(对波形数据)执行所执行(例如,在编码器20的子系统25中)的编码的逆操作以生成波形数据,来恢复要增强的混合音频信号的语音内容的低品质复本。然后,(通过参数数据,以及指示混合音频信号的数据)使用所恢复的语音的低品质复本,来执行语音增强的剩余步骤。

[0159] 元件23被配置成响应于从级21和级22输出的数据生成参数数据。参数数据与原始混合音频数据一起确定作为由原始语音数据(即,混合音频数据的语音内容)所指示的语音的参数重构版本的参数构造语音。语音的参数重构版本至少基本上与由原始语音数据所指示的语音匹配(例如,是由原始语音数据所指示的语音的良好近似)。参数数据确定用于对由混合音频数据所确定的未增强的混合内容的每个片段执行参数编码语音增强的参数编码增强参数p(t)的集合。

[0160] 混和指示符生成元件29被配置成响应于从级21和级22输出的数据生成混和指示符(“BI”)。可以想到,由从编码器20输出的比特流所指示的音频节目将进行混合语音增强(例如,在解码器40中)以确定语音增强音频节目,包括通过将原始节目的未增强音频数据与(根据波形数据所确定的)低品质语音数据和参数数据的组合进行组合。混和指示符确定这样的组合(例如,该组合具有由混和指示符的当前值序列所确定的状态序列),使得与通过将仅低品质语音数据与未增强的音频数据进行组合所确定的纯波形编码语音增强音频节目或者通过将仅参数构造语音与未增强的音频数据进行组合所确定的纯参数编码语音增强音频节目相比,该语音增强音频节目具有更少听得见的语音增强编码伪声(例如,被更好掩蔽的语音增强编码伪声)。

[0161] 在对图3实施方式的变型中,本发明混合语音增强所使用的混和指示符没有在本发明的编码器中被生成(并且没有包括在从编码器输出的比特流中),而替代地响应于从编码器输出的比特流(该比特流包括波形数据和参数数据)而被生成(例如,在对接收器40的变型中)。

[0162] 应当理解,表达“混和指示符”并不意在表示比特流的每个片段的单个参数或值(或者单个参数或值序列)。而是,可以想到,在一些实施方式中,(比特流的片段的)混和指示符可以是两个或更多个参数或值(例如,对于每个片段,参数编码增强控制参数和波形编码增强控制参数)的集合。

[0163] 编码子系统27生成指示混合音频数据(通常,混合音频数据的压缩版本)的音频内容的编码音频数据。编码子系统27通常实现在级22中所执行的转换的逆操作以及其他编码操作。

[0164] 格式化级28被配置成将从元件23输出的参数数据、从元件25输出的波形数据、在元件29中所生成的混和指示符以及从子系统27输出的编码音频数据汇编成指示音频节目的编码比特流。比特流(在一些实现方式中,其可以具有E-AC-3或者AC-3格式)包括未编码的参数数据、波形数据和混和指示符。

[0165] 从编码器20输出的编码音频比特流(编码音频信号)被提供至递送子系统30。递送子系统30被配置成存储由编码器20生成的编码音频信号(例如,以存储指示编码音频信号的数据)和/或传送编码音频信号。

[0166] 解码器40被耦接并配置(例如,被编程)为:从子系统30接收编码音频信号(例如,通过从子系统30中的存储装置读取或取回指示编码音频信号的数据,或者接收已经被子系统30发送的编码音频信号);对指示编码音频信号的混合(语音与非语音)音频内容的数据进行解码;以及对经解码的混合音频内容执行混合语音增强。解码器40通常被配置成生成并且输出指示输入至编码器20的混合音频内容的语音增强版本的语音增强的解码音频信号(例如,至呈现系统,在图3中未示出)。替选地,其包括被耦接成接收子系统43的输出的这样的呈现系统。

[0167] 解码器40的缓冲器44(缓冲存储器)(例如,以非暂态方式)存储由解码器40接收的编码音频信号(比特流)的至少一个片段(例如,帧)。在典型操作中,编码音频比特流的片段序列被提供至缓冲器44并且从缓冲器44被设定至去格式化级41。

[0168] 解码器40的去格式化(解析)级41被配置成对来自递送子系统30的编码比特流进行解析,以从编码比特流提取参数数据(由编码器20的元件23所生成)、波形数据(由编码器

20的元件25所生成)、混和指示符(在编码器20的元件29中所生成)、以及编码混合(语音与非语音)音频数据(在编码器20的编码子系统27中所生成)。

[0169] 编码混合音频数据在解码器40的解码子系统42中被解码,所得到的经解码的混合(语音与非语音)音频数据被设定至混合语音增强子系统43(并且可选地从解码器40输出而未经历语音增强)。

[0170] 响应于由级41从比特流所提取(或者响应于比特流中所包括的元数据在级41中所生成)的控制数据(包括混和指示符),并且响应于由级41所提取的参数数据和波形数据,语音增强子系统43根据本发明的实施方式对来自解码子系统42的解码混合(语音与非语音)音频数据执行混合语音增强。从子系统43输出的语音增强音频信号指示输入至编码器20的混合音频内容的语音增强版本。

[0171] 在图3的编码器20的各种实现中,子系统23可以生成混合音频输入信号的每个通道的每个分块的预测参数p_i的所描述的示例中的任何示例,以用于(例如,在解码器40中)解码混合音频信号的语音分量的重构。

[0172] 使用指示解码混合音频信号的语音内容(例如,由编码器20的子系统25所生成的语音的低品质复本,或者使用由编码器20的子系统23所生成的预测参数p_i所生成的语音内容的重构)的语音信号,可以通过将语音信号与解码混合音频信号进行混合来(例如,在图3的解码器40的子系统43中)执行语音增强。通过对要添加(被混合)的语音应用增益,可以控制语音增强量。对于6dB增强,可以向语音添加0dB增益(假定语音增强混合中的语音具有与所发送或所重构的语音信号相同的水平)。

[0173] 语音增强信号是:

[0174] $M_e = M + g \cdot D_r$ (9)

[0175] 在一些实施方式中,为了获得语音增强增益G,应用下面的混合增益:

[0176] $g = 10^{G/20} - 1$ (10)

[0177] 在独立通道语音重构的情况下,获得语音增强混合M_e作为:

[0178] $M_e = M \cdot (1 + \mathrm{diag}(P) \cdot g)$ (11)

[0179] 在上述示例中,使用相同的能量来重构混合音频信号的每个通道中的语音贡献。当语音已经作为侧信号(例如,作为混合音频信号的语音内容的低品质复本)被发送时或者当使用多个通道(如使用MMSE预测器)重构语音时,语音增强混合需要语音呈现信息,以将与要增强的混合音频信号中已经存在的语音分量在不同通道上具有相同分布的语音进行混合。

[0180] 该呈现信息可以由每个通道的呈现参数r_i来设置,当存在三个通道时,可以将该呈现信息表达为具有以下形式的呈现向量R。

[0181] $$R = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \quad (12)$$

[0182] 语音增强混合为:

[0183] $M_e = M + R \cdot g \cdot D_r$ (13)

[0184] 在存在多个通道的情况下,使用预测参数p_i重构(要与混合音频信号的每个通道进行混合的)语音,先前的等式可以被写为:

[0185] $Me = M + R \cdot g \cdot P \cdot M = (I + R \cdot g \cdot P) \cdot M$ (14)

[0186] 其中,I是单位矩阵。

[0187] 5.语音呈现

[0188] 图4是实现以下形式的常规语音增强混合的语音呈现系统的框图:

[0189] $M_e = M + R \cdot g \cdot D_r$ (15)

[0190] 在图4中,要增强的三通道混合音频信号处于(或者被转换成)频域中。左通道的频率分量被设定至混合元件52的输入,中央通道的频率分量被设定至混合元件53的输入,右通道的频率分量被设定至混合元件54的输入。

[0191] 要与混合音频信号进行混合(以增强混合音频信号)的语音信号可以作为侧信号(例如,作为混合音频信号的语音内容的低品质复本)已经被发送或者可以根据与混合音频信号一起被发送的预测参数p_i被重构。语音信号由频域数据(例如,其包括通过将时域信号转换至频域生成的频率分量)表示,这些频率分量被设定至混合元件51的输入,在混合元件51中,将这些频率分量与增益参数g相乘。

[0192] 元件51的输出被设定至呈现子系统50。还被设定至呈现子系统50的是已经与混合音频信号一起被发送的CLD(通道水平差)参数、CLD_1和CLD_2。(针对混合音频信号的每个片段的)CLD参数描述如何将语音信号混合至混合音频信号内容的所述片段的通道。CLD_1表示一对扬声器通道的平移系数(例如,其限定语音在左通道与中央通道之间的平移),CLD_2表示另一对扬声器通道的平移系数(例如,其限定语音在中央通道与右通道之间的平移)。因此,呈现子系统50设定(至元件52)指示左通道的$R \cdot g \cdot D_r$的数据(语音内容,由左通道的增益参数和呈现参数进行缩放),并且在元件52中将该数据与混合音频信号的左通道进行求和。呈现子系统50设定(至元件53)指示中央通道的$R \cdot g \cdot D_r$的数据(语音内容,由中央通道的增益参数和呈现参数进行缩放),并且在元件53中将该数据与混合音频信号的中央通道进行求和。呈现子系统50设定(至元件54)指示右通道的$R \cdot g \cdot D_r$的数据(语音内容,由右通道的增益参数和呈现参数进行缩放),并且在元件54中将该数据与混合音频信号的右通道进行求和。

[0193] 分别使用元件52、53和54的输出来驱动左扬声器L、中央扬声器C和右扬声器“右”。

[0194] 图5是实现以下形式的常规语音增强混合的语音呈现系统的框图:

[0195] $M_e = M + R \cdot g \cdot P \cdot M = (I + R \cdot g \cdot P) \cdot M$ (16)

[0196] 在图5中,要增强的三通道混合音频信号处于(或者被转换成)频域中。左通道的频率分量被设定至混合元件52的输入,中央通道的频率分量被设定至混合元件53的输入,右通道的频率分量被设定至混合元件54的输入。

[0197] 根据与混合音频信号一起被发送的预测参数p_i来重构(如所指示的)要与混合音频信号进行混合的语音信号。使用预测参数p_1来重构来自混合音频信号的第一(左)通道的语音,使用预测参数p_2来重构来自混合音频信号的第二(中央)通道的语音,使用预测参数p_3来重构来自混合音频信号的第三(右)通道的语音。语音信号由频域数据表示,这些频率分量被设定至混合元件51的输入,在混合元件51中,将这些频率分量与增益参数g相乘。

[0198] 元件51的输出被设定至呈现子系统55。还被设定至呈现子系统的是已经与混合音频信号一起被发送的CLD(通道水平差)参数、CLD_1和CLD_2。(针对混合音频信号的每个片段的)CLD参数描述了如何将语音信号混合至混合音频信号内容的所述片段的通道。CLD_1表示一对扬声器通道的平移系数(例如,其限定语音在左通道与中央通道之间的平移),CLD_2表

示另一对扬声器通道的平移系数(例如,其限定语音在中央通道与右通道之间的平移)。因此,呈现子系统55设定(至元件52)指示左通道的R•g•P•M的数据(与混合音频内容的左通道进行混合的重构语音内容,由左通道的增益参数和呈现参数进行缩放,与混合音频内容的左通道进行混合),并且在元件52中将该数据与混合音频信号的左通道进行求和。呈现子系统55设定(至元件53)指示中央通道的R•g•P•M的数据(与混合音频内容的中央通道进行混合的重构语音内容,由中央通道的增益参数和呈现参数进行缩放),并且在元件53中将该数据与混合音频信号的中央通道进行求和。呈现子系统55设定(至元件54)指示右通道的R•g•P•M的数据(与混合音频内容的右通道进行混合的重构语音内容,由右通道的增益参数和呈现参数进行缩放),并且在元件54中将该数据与混合音频信号的右通道进行求和。

[0199] 分别使用元件52、53和54的输出来驱动左扬声器L、中央扬声器C和右扬声器"右"。

[0200] CLD(通道水平差)参数通常与扬声器通道信号一起被发送(例如,以确定应当呈现不同通道的水平之间的比率)。在本发明的一些实施方式中以新颖的方式使用CLD参数(例如,以在语音增强音频节目的扬声器通道之间平移所增强的语音)。

[0201] 在典型实施方式中,呈现参数r_i是(或者指示)语音的上混合系数,描述语音信号如何被混合至要增强的混合音频信号的通道。可以使用通道水平差参数(CLD)将这些系数有效地发送至语音增强器。一个CLD表示两个扬声器的平移系数。例如,

[0202] $$\beta_1 = \sqrt{\frac{1}{1+10^{\frac{CLD}{10}}}} \tag{17}$$

[0203] $$\beta_2 = \sqrt{\frac{10^{\frac{CLD}{10}}}{1+10^{\frac{CLD}{10}}}} \tag{18}$$

[0204] 其中,β_1表示在平移期间瞬时的第一扬声器的扬声器馈送的增益,β_2表示在平移期间瞬时的第二扬声器的扬声器馈送的增益。当CLD=0时,平移完全针对第一扬声器,而当CLD接近无穷大时,平移完全朝向第二扬声器。使用在dB范围中所限定的CLD,有限数目的量化水平可以足够描述平移。

[0205] 使用两个CLD可以限定在三个扬声器之间进行平移。可以如下根据呈现系数来导出CLD:

[0206] $$CLD_1 = 10 \cdot \log_{10}\left(\frac{\bar{r}_2^2}{\bar{r}_1^2}\right) \tag{19}$$

[0207] $$CLD_2 = 10 \cdot \log_{10}\left(\frac{\bar{r}_3^2}{\bar{r}_1^2+\bar{r}_2^2}\right) \tag{20}$$

[0208] 其中,$\bar{r}_x^2 = \frac{r_x^2}{\sum_i r_i^2}$是归一化呈现系数,使得

[0209] $$\bar{r}_1^2 + \bar{r}_2^2 + \bar{r}_3^2 = 1 \tag{21}$$

[0210] 然后,可以通过以下等式根据CLD重构呈现系数:

[0211] $$R=\begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}=\begin{pmatrix} \sqrt{\frac{1}{\left(1+10^{\frac{CLD_1}{10}}\right)\left(1+10^{\frac{CLD_2}{10}}\right)}} \\ \sqrt{\frac{10^{\frac{CLD_1}{10}}}{\left(1+10^{\frac{CLD_1}{10}}\right)\cdot\left(1+10^{\frac{CLD_2}{10}}\right)}} \\ \sqrt{\frac{10^{\frac{CLD_2}{10}}}{1+10^{\frac{CLD_2}{10}}}} \end{pmatrix} \quad (22)$$

[0212] 如在本文中别处所指出的,波形编码语音增强使用要增强的混合内容信号的语音内容的低品质复本。低品质复本通常以低比特率被编码并且作为侧信号与混合内容信号一起被发送,因此,低品质复本通常包括显著的编码伪声。因此,在具有低SNR(即,由混合内容信号所指示的语音与所有其他声音之间的低比率)的情况下,波形编码语音增强提供良好的语音增强性能,而在具有高SNR的情况下通常提供差的性能(即,导致不期望的听得见的编码伪声)。

[0213] 相反地,当挑选出(要增强的混合内容信号的)语音内容(例如,其被设置为多通道混合内容信号中的仅中央通道的内容)或者混合内容信号以其他方式具有高SNR时,参数编码语音增强提供良好的语音增强性能。

[0214] 因此,波形编码语音增强和参数编码语音增强具有互补的性能。基于要增强其语音内容的信号的特性,本发明的一类实施方式将两种方法进行混和以利用它们的性能。

[0215] 图6是该类实施方式中的被配置成执行混合语音增强的语音呈现系统的框图。在一种实现中,图3的解码器40的子系统43实现图6系统(除了图6中所示的三个扬声器以外)。混合(hybrid)语音增强(混合(mixing))可以由下式来描述

[0216] $M_e=R\cdot g_1\cdot D_r+(I+R\cdot g_2\cdot P)\cdot M$ (23)

[0217] 其中,$R\cdot g_1\cdot D_r$是由常规的图4系统所实现的类型的波形编码语音增强,$R\cdot g_2\cdot P\cdot M$是由常规的图5系统所实现的类型的参数编码语音增强,参数g_1和g_2控制整体增强增益以及两种语音增强方法之间的平衡(trade-off)。参数g_1和g_2的定义的示例是:

[0218] $g_1=\alpha_c\cdot(10^{G/20}-1)$ (24)

[0219] $g_2=(1-\alpha_c)\cdot(10^{G/20}-1)$ (25)

[0220] 其中,参数α_c限定参数编码语音增强方法与参数编码语音增强方法之间的平衡。当值$\alpha_c=1$时,仅语音的低品质复本用于波形编码语音增强。当$\alpha_c=0$时,参数编码增强模式对增强作出全部贡献。0到1之间的α_c值对两种方法进行混和。在一些实现中,α_c是宽带参数(应用于音频数据的所有频带)。可以在各个频带内应用相同的原理,使得使用每个频带的参数α_c的不同值以频率相关方式对混和进行优化。

[0221] 在图6中,要增强的三通道混合音频信号处于(或者被转换成)频域中。左通道的频率分量被设定至混合元件65的输入,中央通道的频率分量被设定至混合元件66的输入,右通道的频率分量被设定至混合元件67的输入。

[0222] 要与混合音频信号进行混合(以增强混合音频信号)的语音信号包括:已经根据与混合音频信号(例如,作为侧信号)一起(根据波形编码语音增强)被传输的波形数据而生成

的混合音频信号的语音内容的低品质复本(在图6中标识为“语音”),以及根据混合音频信号和与混合音频信号一起(根据参数编码语音增强)被传输的预测参数p_i所重构的重构语音信号(其从图6的参数编码语音重构元件68输出)。语音信号由频域数据(例如,其包括通过将时域信号转换成频域所生成的频率分量)表示。低品质语音复本的频率分量被设定至混合元件61的输入,在混合元件61中,将低品质语音复本的频率分量乘以增益参数g_2。参数重构语音信号的频率分量从元件68的输出被设定至混合元件62的输入,在混合元件62中,将参数重构语音信号的频率分量乘以增益参数g_1。在替选实施方式中,在时域中而不是在如图6实施方式中的频域中执行要实现语音增强所执行的混合。

[0223] 求和元件63对元件61和元件62的输出进行求和以生成要与混合音频信号进行混合的语音信号,并且该语音信号从元件63的输出被设定至呈现子系统64。还被设定至呈现子系统64的是已经与混合音频信号一起被发送的CLD(通道水平差)参数、CLD_1和CLD_2。(针对混合音频信号的每个片段的)CLD参数描述了如何将语音信号混合至混合音频信号内容的所述片段的通道。CLD_1表示一对扬声器通道的平移系数(例如,其限定语音在左通道与中央通道之间的平移),CLD_2表示另一对扬声器通道的平移系数(例如,其限定语音在中央通道与右通道之间的平移)。因此,呈现子系统64设定(至元件52)指示左通道的$R \cdot g_1 \cdot D_r + (R \cdot g_2 \cdot P) \cdot M$的数据(与混和音频内容的左通道进行混合的重构语音内容,由左通道的增益参数和呈现参数缩放,与混合音频内容的左通道进行混合),并且在元件52中将该数据与混合音频信号的左通道进行求和。呈现子系统64设定(至元件53)指示中央通道的$R \cdot g_1 \cdot D_r + (R \cdot g_2 \cdot P) \cdot M$的数据(与混合音频内容的中央通道进行混合的重构语音内容,由中央通道的增益参数和呈现参数进行缩放),并且在元件53中将该数据与混合音频信号的中央通道进行求和。呈现子系统64设定(至元件54)指示右通道的$R \cdot g_1 \cdot D_r + (R \cdot g_2 \cdot P) \cdot M$的数据(与混和音频内容的右通道进行混合的重构语音内容,由右通道的增益参数和呈现参数进行缩放),并且在元件54中将该数据与混合音频信号的右通道进行求和。

[0224] 分别使用元件52、53和54的输出来驱动左扬声器L、中央扬声器C和右扬声器“右”。

[0225] 当参数α_c被约束成具有值$\alpha_c = 0$或者值$\alpha_c = 1$时,图6系统可以实现基于时间SNR的切换。在以下的强的比特率约束情况下这样的实现尤其有用:低品质语音复本数据可以被发送或者参数数据可以被发送,但是低品质语音复本数据和参数数据两者不能一起被发送。例如,在一种这样的实现中,仅在$\alpha_c = 1$的片段中将低品质语音复本与混合音频信号(例如,作为侧信号)一起发送,并且仅在$\alpha_c = 0$的片段中将预测参数p_i与混合音频信号(例如,作为侧信号)一起发送。

[0226] 切换(由图6的该实现中的元件61和62所实现)基于片段中的语音内容与所有其他音频内容之间的比率(SNR)(该比率又确定α_c的值)来确定要对每个片段执行波形编码增强还是参数编码增强。这样的实现可以使用SNR的阈值来决定要选择哪种方法:

[0227] $$\alpha_c = \begin{cases} 0 & \text{如果 } SNR > \tau \\ 1 & \text{如果 } SNR \leq \tau \end{cases} \quad (26)$$

[0228] 其中,τ是阈值(例如,τ可以等于0)。

[0229] 当SNR大约为几个帧的阈值时,图6的一些实现使用滞后作用来阻止在波形编码增强模式与参数编码增强模式之间快速交替切换。

[0230] 当使得参数α_c能够具有0到1范围内的任意实值(0和1也包括在内)时,图6系统可以实现基于时间SNR的混和。

[0231] 图6系统的一种实现使用(要增强的混合音频信号的片段的SNR的)两个目标值τ_1和τ_2,超过这两个目标值,一种方法(波形编码增强或者参数编码增强)总是被视为提供最佳性能。在这些目标之间,使用插值来确定片段的参数α_c的值。例如,可以使用线性插值来确定片段的参数α_c的值:

[0232] $$\alpha_c = \begin{cases} 0 & \text{如果 } SNR > \tau_2 \\ 1 - \dfrac{SNR - \tau_1}{\tau_2 - \tau_1} & \text{如果 } \tau_1 < SNR \leq \tau_2 \\ 1 & \text{如果 } SNR \leq \tau_1 \end{cases} \quad (27)$$

[0233] 替选地,可以使用其他适当的插值方案。当SNR不可用时,在许多实现中可以使用预测参数来提供SNR的近似值。

[0234] 在另一类实施方式中,通过听觉掩蔽模型确定要对音频信号的每个片段执行的波形编码增强和参数编码增强的组合。在该类的典型实施方式中,要对音频节目的片段执行的波形编码增强和参数编码增强的混和的最佳混和比率使用刚好防止编码噪声变得听见的最高波形编码增强量。在本文中,参照图7来描述使用听觉掩蔽模型的本发明方法的实施方式的示例。

[0235] 更一般地,下面的考虑涉及以下实施方式:使用听觉掩蔽模型来确定要对音频信号的每个片段执行的波形编码增强和参数编码增强的组合(例如,混和)。在这样的实施方式中,对指示要称为未增强音频混合的语音与背景音频的混合A(t)的数据进行设置并且根据听觉掩蔽模型(例如,由图7的元件11所实现的模型)对其进行处理。模型预测了未增强音频混合的每个片段的掩蔽阈值$\Theta(f,t)$。可以将具有时间索引n和频带索引b的未增强音频混合的每个时间-频率分块的掩蔽阈值表示为$\Theta_{n,b}$。

[0236] 掩蔽阈值$\Theta_{n,b}$指示:对于帧n和频带b,可以添加多少失真而不会听得见。令$\varepsilon_{D,n,b}$为低品质语音复本(要于波形编码增强)的编码误差(即,量化噪声),并且令$\varepsilon_{P,n,b}$为参数预测误差。

[0237] 该类中的一些实施方式实现到由未增强音频混合内容最佳掩蔽的方法(波形编码增强或参数编码增强)的硬切换:

[0238] $$\alpha_c = \begin{cases} 0 & \text{如果 } \sum_{n,b} \Theta_{n,b} - \varepsilon_{P,n,b} > \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} \\ 1 & \text{如果 } \sum_{n,b} \Theta_{n,b} - \varepsilon_{P,n,b} \leq \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} \end{cases} \quad (28)$$

[0239] 在许多实际情况中,在生成语音增强参数时准确的参数预测误差$\varepsilon_{P,n,b}$可能不可用,这是因为这些可能在未增强混合的混合被编码之前生成。特别地,参数编码方案可以对来自混合内容通道的语音的参数重构的误差具有显著影响。

[0240] 因此,当(要用于波形编码增强的)低品质语音复本中的编码伪声未被混合内容掩蔽时,一些替选实施方式在参数编码语音增强(与波形编码增强)中进行混合:

[0241] $$\alpha_c = \begin{cases} 1 & \text{如果 } \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} \geq 0 \\ 1 - \frac{\sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b}}{\tau_a} & \text{如果 } -\tau_a \leq \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} < 0 \\ 0 & \text{如果 } \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} < -\tau_a \end{cases} \quad (29)$$

[0242] 其中，τ_a是失真阈值，超出该失真阈值，仅应用参数编码增强。当整体失真大于整体掩蔽可能(potential)时，该解决方案开始波形编码增强和参数编码增强的混和。实际上，这意味着失真已经是听得见的。因此，可以使用具有比0更高的值的第二阈值。替选地，可以使用宁愿关注未被掩蔽的时间-频率分块而不是平均行为的情况。

[0243] 类似地，当(要用于波形编码增强的)低品质语音复本中的失真(编码伪声)太高时，可以将该方法与SNR指引的混和规则进行组合。该方法的优点在于：在SNR非常低的情况下，当其产生比低品质语音复本的失真更多听得见的噪声时，不使用参数编码增强模式。

[0244] 在另一种实施方式中，当在每个这样的时间-频率分块中检测到频谱空洞(spectral hole)时，对一些时间-频率分块执行的语音增强的类型偏离由上述示例方案(或类似方案)所确定的语音增强类型。例如通过在参数重构中对相应分块中的能量进行评估可以检测频谱空洞，而在(要用于波形编码增强的)低品质语音复本中能量为0。如果该能量超过阈值，则可以将其视为相关音频。在这些情况下，可以将分块的参数α_c设置成0(或者，取决于SNR，分块的参数α_c可以朝向0偏置)。

[0245] 在一些实施方式中，本发明的编码器能够在以下模式中的任意所选之一中操作：

[0246] 1.独立通道参数——在该模式下，传输包括语音的每个通道的参数集合。使用这些参数，接收编码音频节目的解码器可以对节目执行参数编码语音增强以将这些通道中的语音加强任意量。用于传输参数集合的示例比特率是0.75kbps至2.25kbps。

[0247] 2.多通道语音预测——在该模式下，以线性组合对混合内容的多个通道进行组合来预测语音信号。传输每个通道的参数集合。使用这些参数，接收编码音频节目的解码器可以对节目执行参数编码语音增强。将附加的位置数据与编码音频节目一起传输以使得能够将所加强的语音呈现回混合。用于传输参数集合和位置数据的示例比特率是每对话1.5kbps至6.75kbps。

[0248] 3.波形编码语音——在该模式下，通过任何适当的方式将音频节目的语音内容的低品质复本与常规音频内容(例如，作为分立的比特流)单独地并行传输。接收编码音频节目的解码器可以通过在语音内容的分立的低品质复本中与主混合进行混合对节目执行波形编码语音增强。当幅度加倍时，通常将语音的低品质复本与0dB的增益进行混合将使语音加强6dB。此外，对于该模式，位置数据被传输，使得将语音信号正确地分布在相关通道中。用于传输语音的低品质复本和位置数据的示例比特率大于每对话20kbps。

[0249] 4.波形参数混合——在该模式下，将音频节目的语音内容的低品质复本(用于对节目执行波形编码语音增强)和每个包括语音的通道的参数集合(用于对节目执行参数编码语音增强)两者与节目的未增强混合(语音与非语音)音频内容并行传输。当语音的低品质复本的比特率降低时，在该信号中更多编码伪声变得听得见，并且减小了传输所需要的带宽。此外，还传输了下述混和指示符：该混和指示符使用语音的低品质复本和参数集合来

确定要对节目的每个片段执行的波形编码语音增强和参数编码语音增强的组合。在接收器处,对节目执行混合语音增强,包括通过:执行由混和指示符所确定的波形编码语音增强和参数编码语音增强的组合,从而生成指示语音增强音频节目的数据。此外,还将位置数据与节目的未增强的混和音频内容一起传输以指示要在哪里呈现语音信号。该方法的优点在于:如果接收器/解码器丢弃语音的低品质复本并且仅应用参数集合来执行参数编码增强,则可以降低所要求的接收器/解码器复杂度。用于传输语音的低品质复本、参数集合、混和指示符和位置数据的示例比特率是每对话8至24kbps。

[0250] 出于实践原因,可以将语音增强增益限制成0至12dB范围。可以将编码器实现成:能够进一步借助于比特流字段来进一步减小该范围的上限。在一些实施方式中,(从编码器输出的)编码节目的语法将支持(除了节目的非语音内容以外的)多个同时的可增强对话,使得可以分立地重构和呈现每个对话。在这些实施方式中,在后面的模式下,将用于(来自不同空间位置处的多个源的)同时对话的语音增强呈现在单个位置处。

[0251] 在编码音频节目是基于对象的音频节目的一些实施方式中,可以选择(最大总数中的)一个或更多个对象簇来进行语音增强。可以将CLD值对包括在编码节目中以供语音增强和呈现系统使用,以在对象簇之间平移所增强的语音。类似地,在编码音频节目包括常规5.1格式的扬声器通道的一些实施方式中,可以选择前扬声器通道中的一个或更多个以进行语音增强。

[0252] 本发明的另一个方面是用于对已经根据本发明的编码方法的实施方式生成的编码音频信号进行解码并执行混合语音增强的方法(例如,由图3的解码器40所执行的方法)。

[0253] 可以以硬件、固件或软件或者两者的组合(例如,作为可编程逻辑阵列)来实现本发明。除非另有说明,否则作为本发明的一部分所包括的算法或处理并不固有地与任何特定计算机或其他设备相关。具体地,可以与根据本文中的教示所编写的程序一起使用各种通用机器,或者更便利的是,可以构造执行所要求的方法步骤的更专用的设备(例如,集成电路)。因此,可以以在一个或更多个可编程计算机系统(例如,实现图3的编码器20或图7的编码器或图3的解码器40的计算机系统)上执行的一个或更多个计算机程序来实现本发明,每个可编程计算机系统包括至少一个处理器、至少一个数据存储系统(包括易失性和非易失性存储器和/或存储元件)、至少一个输入装置或端口、以及至少一个输出装置或端口。对输入数据应用程序代码以执行本文中所描述的功能并且生成输出信息。以已知的方式对一个或更多个输出装置应用输出信息。

[0254] 可以以与计算机系统进行通信的任何期望的计算机语言(包括机器语言、汇编语言、或者高级过程语言、逻辑语言、或者面向对象的编程语言)来实现每个这样的程序。在任何情况下,语言可以是编译语言或解释语言。

[0255] 例如,当由计算机软件指令序列实现时,可以通过在适当的数字信号处理硬件中运行的多线程软件指令序列来实现本发明的实施方式的各种功能和步骤,在这种情况下,实施方式的各种装置、步骤和功能可以对应于软件指令的一部分。

[0256] 优选地,每个这样的计算机程序被存储在能够由通用或专用可编程计算机读取的存储介质或装置(例如,固态存储器或介质,或者磁介质或光介质)上或者被下载至能够由通用或专用可编程计算机读取的存储介质或装置(例如,固态存储器或介质,或者磁介质或光介质),以当执行本文中所描述的过程的计算机系统读取存储介质或装置时对计算机进

行配置和操作。还可以将本发明系统实现为配置有(即,存储)计算机程序的计算机可读存储介质,其中,如此配置的存储介质使计算机系统以特定且预定义方式进行操作以执行本文中所描述的功能。

[0257] 已经描述了本发明的许多实施方式。然而,应当理解,在不偏离本发明的精神和范围的情况下,可以作出各种修改。鉴于上面的教示,本发明的大量修改和变更是可以的。应当理解,在所附权利要求的范围内,可以以与如本文中具体描述的方式不同的方式来实践本发明。

[0258] 6.中间/侧表示

[0259] 音频解码器可以至少部分地基于M/S表示中的控制数据、控制参数等来执行如本文中所描述的语音增强操作。上游音频编码器可以生成M/S表示中的控制数据、控制参数等,并且音频解码器从由上游音频编码器所生成的编码音频信号中提取M/S表示中的控制数据、控制参数等。

[0260] 在根据混合内容预测语音内容(例如,一个或更多个对话等)的参数编码增强模式中,如以下表达式中所示,可以使用单个矩阵H一般地表示语音增强操作:

[0261] $$\begin{pmatrix} M_{s,c_1} \\ M_{s,c_2} \end{pmatrix} = H \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \quad (30)$$

[0262] 其中,左手侧(LHS)表示通过如由矩阵H所表示的语音增强操作对右手侧(RHS)的原始混合内容信号进行操作而生成的语音增强混合内容信号。

[0263] 出于说明的目的,语音增强混合内容信号(例如,表达式(30)的LHS等)和原始混合内容信号(例如,由表达式(30)中的H所操作的原始混合内容信号等)中的每个包括分别在两个通道c_1和c_2中具有语音增强混合内容和原始混合内容的两个分量信号。两个通道c_1和c_2可以是基于非M/S表示的非M/S音频通道(例如,左前通道、右前通道等)。应当注意,在各种实施方式中,语音增强混合内容信号和原始混合内容信号中的每个还可以包括在除了两个非M/S通道c_1和c_2以外的通道(例如,环绕通道、低频效果通道等)中具有非语音内容的分量信号。还应当注意,在各种实施方式中,语音增强混合内容信号和原始混合内容信号中的每个可能包括在一个通道、如表达式(30)中所示的两个通道、或者多于两个通道中具有语音内容的分量信号。如本文中所描述的语音内容可以包括一个对话、两个对话或更多个对话。

[0264] 在一些实施方式中,如由表达式(30)中的H所表示的语音增强操作可以用于(例如,如由SNR引导混和规则等所指引)混合内容中的语音内容与其他(例如,非语音等)内容之间的SNR值相对高的混合内容的时间片(片段)。

[0265] 如以下表达式所示,可以将矩阵H重写/扩展为表示M/S表示中的增强操作的矩阵H_{MS}在右边乘以从非M/S表示到M/S表示的正向转换矩阵并且在左边乘以该正向转换矩阵的逆(其包括因子1/2)的乘积:

[0266] $$\begin{pmatrix} M_{s,c_1} \\ M_{s,c_2} \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot H_{MS} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \quad (31)$$

[0267] 其中,矩阵H_{MS}右边的示例转换矩阵基于正向转换矩阵将M/S表示中的中间通道混合内容信号限定为两个通道c_1和c_2中的两个混合内容信号之和,并且将M/S表示中的侧通道

混合内容信号限定为两个通道c_1和c_2中的两个混合内容信号之差。应当注意,在各种实施方式中,还可以使用除了表达式(31)中所示的示例转换矩阵以外的其他转换矩阵(例如,向不同的非M/S通道分配不同的权重等),以将混合内容信号从一种表示转换为不同的表示。例如,对于对话增强,其中对话不在幻象中心呈现,而是在具有不相等的权重λ_1和λ_2的两个信号之间平移。如以下表达式所示,可以将M/S转换矩阵修改成使侧信号中对话分量的能量最小:

[0268] $$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = \tfrac{1}{2} \cdot \lambda_1 \cdot \lambda_2 \cdot \begin{pmatrix} \frac{1}{\lambda_2} & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & -\frac{1}{\lambda_1} \end{pmatrix} \cdot H_{MS} \cdot \begin{pmatrix} \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & -\frac{1}{\lambda_2} \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \tag{31A}$$

[0269] 在示例实施方式中,如以下表达式所示,可以将代表M/S表示中的增强操作的矩阵H_{MS}定义为对角化(例如,厄米特矩阵等)矩阵:

[0270] $$H_{MS} = \begin{pmatrix} g \cdot p_1 + 1 & 0 \\ 0 & g \cdot p_2 + 1 \end{pmatrix} \tag{32}$$

[0271] 其中,p_1和p_2分别表示中间通道和侧通道预测参数。预测参数p_1和p_2中的每一个可以包括针对M/S表示中的相应混合内容信号的时间-频率分块的时变预测参数集合,以被用于根据混合内容信号重构语音内容。例如,如表达式(10)所示,增益参数g对应于语音增强增益G。

[0272] 在一些实施方式中,在参数通道独立增强模式下执行M/S表示中的语音增强操作。在一些实施方式中,使用中间通道信号和侧通道信号两者中的预测语音内容或者使用仅中间通道信号中的预测语音内容来执行M/S表示中的语音增强操作。出于说明的目的,如以下表达式所示,使用仅中间通道中的混合内容信号来执行M/S表示中的语音增强操作:

[0273] $$H_{MS} = \begin{pmatrix} g \cdot p_1 + 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{33}$$

[0274] 其中,预测参数p_1包括针对M/S表示的中间通道中的混合内容信号的时间-频率分块的单个预测参数集合,以被用于根据仅中间通道中的混合内容信号重构语音内容。

[0275] 基于表达式(33)中所给出的对角化矩阵H_{MS},还可以将如由表达式(31)所表示的参数增强模式下的语音增强操作进一步缩减成以下表达式,该表达式提供了表达式(30)中的矩阵H的明确示例:

[0276] $$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = \tfrac{1}{2} \cdot \begin{pmatrix} 2 + g \cdot p_1 & g \cdot p_1 \\ g \cdot p_1 & 2 + g \cdot p_1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \tag{34}$$

[0277] 在波形参数混合增强模式下,可以使用以下示例表达式在M/S表示中表示语音增强操作:

$$M_e = g_1 \cdot \begin{pmatrix} d_{c,1} \\ 0 \end{pmatrix} + \begin{pmatrix} g_2 \cdot p_1 + 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \quad (35)$$

[0278]

$$= H_d \cdot D_c + H_p \cdot M$$

[0279] 其中，m_1和m_2在混合内容信号向量M中分别表示中间通道混合内容信号(例如，非M/S通道如左前通道和右前通道等中的混合内容信号之和)和侧通道混合内容信号(例如，非M/S通道如左前通道和右前通道等中的混合内容信号之差)。信号$d_{c,1}$代表M/S表示的对话信号向量D_c中的中间通道对话波形信号(例如，表示混合内容中的对话的降低版本的编码波形等)。矩阵H_d表示基于M/S表示的中间通道中的对话信号$d_{c,1}$的M/S表示中的语音增强操作，并且可以包括在第一行第一列(1×1)处的仅一个矩阵元素。矩阵H_p表示基于使用M/S表示的中间通道的预测参数p_1重构的对话的、M/S表示中的语音增强操作。在一些实施方式中，例如，如表达式(23)和(24)中所描绘的，增益参数g_1和g_2共同(例如，在分别被应用于对话波形信号和重构对话等之后)对应于语音增强增益G。具体地，在与M/S表示的中间通道中的对话信号$d_{c,1}$有关的波形编码语音增强操作中应用参数g_1，而在与M/S表示的中间通道和侧通道中的混合内容信号m_1和m_2有关的参数编码语音增强操作中应用参数g_2。参数g_1和g_2对整体增强增益以及两种语言增强方法之间的平衡进行控制。

[0280] 在非M/S表示中，可以使用以下表达式来表示与使用表达式(35)所表示的语音增强操作相对应的语音增强操作：

$$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot H_d \cdot D_c + \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot H_p \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \quad (36)$$

[0281]

$$= \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \left(H_d \cdot D_c + H_p \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \right)$$

[0282] 其中，可以使用与非M/S表示和M/S表示之间的正向转换矩阵左乘的非M/S通道中的混合内容信号M_{c1}和M_{c2}来代替如表达式(35)中所示的M/S表示中的混合内容信号m_1和m_2。表达式(36)中的逆转换矩阵(具有因子1/2)将如表达式(35)所示的M/S表示中的语音增强混合内容信号转换回非M/S表示(例如，左前通道和右前通道等)中的语音增强混合内容信号。

[0283] 另外，可选地或替选地，在语音增强操作之后无另外的基于QMF的处理被执行的一些实施方式中，出于效率原因，在时域中的QMF合成滤波器组之后，可以执行组合基于对话信号$d_{c,1}$的语音增强内容与基于通过预测重构的对话的语音增强混合内容的语音增强操作(例如，如由H_d、H_p转换等所表示)中的一些或所有。

[0284] 可以基于以下一个或更多个预测参数生成方法中的一个来生成用于根据M/S表示的中间通道和侧通道中的一个或两个中的混合内容信号来构造/预测语音内容的预测参数，所述一个或更多个预测参数生成方法包括但不限于仅以下方法中的任意方法：如图1中所描绘的独立通道对话预测方法、如图2中所描绘的多通道对话预测方法等。在一些实施方

式中,预测参数生成方法中的至少之一可以基于MMSE、梯度下降、一个或更多个其他优化方法等。

[0285] 在一些实施方式中,可以在M/S表示中的音频节目的片段的参数编码增强数据(例如,与基于对话信号$d_{c,1}$的语音增强内容有关等)与波形编码增强(例如,与基于通过预测所重构的对话的语音增强混合内容有关等)之间使用如先前所讨论的基于“盲”时间SNR的切换方法。

[0286] 在一些实施方式中,M/S表示中的波形数据(例如,与基于对话信号$d_{c,1}$的语音增强内容有关等)和重构语音数据(例如,与基于通过预测所重构的对话的语音增强混合内容有关等)的组合(例如,由先前讨论的混和指示符指示,表达式(35)中的g_1和g_2的组合等)随时间变化,其中每个组合状态与携载波形数据和在重构语音数据时所使用的混合内容的比特流的相应片段的语音内容和其他音频内容有关。混和指示符被生成,使得由节目的相应片段中的语音内容与其他音频内容的信号特性(例如,语音内容的功率与其他音频内容的功率之比、SNR等)来确定(波形数据和重构语音数据的)当前组合状态。音频节目的片段的混和指示符可以是在图3的编码器的子系统29中针对片段所生成的混和指示符参数(或参数集合)。可以使用如先前所讨论的听觉掩蔽模型来更准确地预测对话信号向量Dc中的降低品质语音复本中的编码噪声如何被主要节目的音频混合掩蔽并且据此选择混和比率。

[0287] 图3的编码器20的子系统28可以被配置成将与M/S语音增强操作有关的混和指示符包括在比特流中以作为要从编码器20输出的M/S语音增强元数据的一部分。可以根据与对话信号D_c中的编码伪声有关的缩放因子$g_{max}(t)$等来生成(例如,在图7的编码器的子系统13中)与M/S语音增强操作有关的混和指示符。缩放因子$g_{max}(t)$可以由图7编码器的子系统14生成。图7编码器的子系统13可以被配置成将混和指示符包括在要从图7编码器输出的比特流中。另外,可选地或替选地,子系统13可以将由子系统14所生成的缩放因子$g_{max}(t)$包括在要从图7编码器输出的比特流中。

[0288] 在一些实施方式中,由图7的操作10所生成的未增强音频混合A(t)表示参考音频通道配置中的混合内容信号向量(例如,其时间片段等)。由图7的元件12所生成的参数编码增强参数p(t)表示用于关于混合内容信号向量的每个片段执行M/S表示中的参数编码语音增强的M/S语音增强元数据中的至少一部分。在一些实施方式中,由图7的编码器15所生成的降低品质语音复本s’(t)表示M/S表示(例如,关于中间通道对话信号、侧通道对话信号等)中的对话信号向量。

[0289] 在一些实施方式中,图7的元件14生成缩放因子$g_{max}(t)$并且将其提供至编码元件13。在一些实施方式中,元件13针对音频节目的每个片段生成指示参考音频通道配置中的(例如,未增强等)混合内容信号向量、M/S语音增强元数据、如果可应用则有M/S表示中的对话信号向量、以及如果可应用则有缩放因子$g_{max}(t)$的编码音频比特流,该编码音频比特流可以被发送至或者以其他方式被递送至接收器。

[0290] 当将非M/S表示中的未增强音频信号与M/S语音增强元数据一起递送(例如,发送)至接收器时,接收器可以转换M/S表示中的未增强音频信号的每个片段并且针对片段执行由M/S语音增强元数据所指示的M/S语音增强操作。如果要在混合语音增强模式下或在波形编码增强模式下对片段执行语音增强操作,则可以向节目的片段的M/S表示中的对话信号向量提供非M/S表示中的未增强混合内容信号向量。如果可应用,则接收并解析比特流的接

收器可以被配置成:响应于缩放因子$g_{max}(t)$来生成混和指示符并且确定表达式(35)中的增益参数g_1和g_2。

[0291] 在一些实施方式中,在元件13的编码输出已经被递送至的接收器中,至少部分地在M/S表示中执行语音增强操作。在一个示例中,可以至少部分地基于根据由接收器接收的比特流所解析的混和指示符对未增强混合内容信号的每个片段应用与增强的预定(例如,所要求的)总量相对应的表达式(35)中的增益参数g_1和g_2。在另一个示例中,可以至少部分地基于从根据由接收器接收的比特流所解析的片段的缩放因子$g_{max}(t)$所确定的混和指示符对未增强的混合内容信号的每个片段应用与增强的预定(例如,所要求的)总量相对应的表达式(35)中的增益参数g_1和g_2。

[0292] 在一些实施方式中,图3的编码器20的元件23被配置成响应于从级21和22输出的数据,生成包括M/S语音增强元数据的参数数据(例如,根据中间通道和/或侧通道中的混合内容等重构对话/语音内容的预测参数)。在一些实施方式中,图3的编码器20的混和指示符生成元件29被配置成响应于从级21和22输出的数据来生成确定参数语音增强内容(例如,使用增益参数g_1等)和基于波形的语音增强内容(例如,使用增益参数g_1等)的组合的混和标识符“BI”。

[0293] 在对图3实施方式的变型中,在编码器中没有生成用于M/S混合语音增强的混和指示符(以及该混和指示符没有包括在从编码器输出的比特流中),而是替代地响应于从编码器输出的比特流(该比特流包括M/S通道中的波形数据和M/S语音增强元数据)来(例如,在对接收器40的变型中)生成用于M/S混合语音增强的混和指示符。

[0294] 解码器40被耦接和配置(例如,被编程)为:从子系统30接收编码音频信号(例如,通过从子系统30中的存储装置读取或取回指示编码音频信号的数据,或者接收已经被子系统30发送的编码音频信号);根据编码音频信号对指示参考音频通道配置中的混合(语音与非语音)内容信号向量的数据进行解码;以及至少部分地在M/S表示中对参考音频通道配置中的解码混合内容执行语音增强操作。解码器40可以被配置成生成和输出(例如,至呈现系统等)指示语音增强混合内容的语音增强的解码音频信号。

[0295] 在一些实施方式中,图4至图6中所描绘的呈现系统中的一些或全部可以被配置成:呈现通过M/S语音增强操作生成的语音增强混合内容,所述M/S语音增强操作中的至少一些是在M/S表示中所执行的操作。图6A示出了被配置成执行如表达式(35)中所表示的语音增强操作的示例呈现系统。

[0296] 图6A的呈现系统可以被配置成:响应于确定在参数语音增强操作中所使用的至少一个增益参数(例如,表达式(35)中的g_2等)是非零的(例如,在混合增强模式下、在参数增强模式下等)来执行参数语音增强操作。例如,根据这样的确定,图6A的子系统68A可以被配置成:对非M/S通道上分布的混合内容信号向量(“混合音频(T/F)”)执行转换以生成M/S通道上分布的相应混合内容信号向量。若适当的话,该转换可以使用正向转换矩阵。可以应用用于参数增强操作的预测参数(例如,p_1、p_2等)、增益参数(例如,表达式(35)中的g_2等),以根据M/S通道的混合内容信号向量来预测语音内容并且增强所预测的语音内容。

[0297] 图6A的呈现系统可以被配置成:响应于确定波形编码语音增强操作中所使用的至少一个增益参数(例如,表达式(35)中的g_1等)是非零的(例如,在混合增强模式下、在波形编码增强模式下等)来执行波形编码语音增强操作。例如,根据这样的确定,图6A的呈现系

统可以被配置成从所接收的编码音频信号接收/提取M/S通道上分布的对话信号向量(例如,关于混合内容信号向量中存在的语音内容的降低版本)。可以应用用于波形编码增强操作的增益参数(例如,表达式(35)中的g_1等)以增强由M/S通道的对话信号向量所表示的语音内容。用户可定义的增强增益(G)可以用于使用可以或不可以存在于比特流中的混和参数来导出增益参数g_1和g_2。在一些实施方式中,可以从所接收的编码音频信号中的元数据中提取要与用户可定义的增强增益(G)一起使用以导出增益参数g_1和g_2的混和参数。在一些其他实施方式中,可以不从所接收的编码音频信号中的元数据提取这样的混和参数,而是可以由接收方编码器基于所接收的编码音频信号中的音频内容来导出这样的混和参数。

[0298] 在一些实施方式中,M/S表示中的参数增强语音内容和波形编码增强语音内容的组合被设定(assert)或被输入至图6A的子系统64A。图6的子系统64A可以被配置成:对M/S通道上分布的增强语音内容的组合执行转换以生成非M/S通道上分布的增强语音内容信号向量。若适当的话,该转换可以使用逆转换矩阵。可以将非M/S通道的增强语音内容信号向量与分布在非M/S通道上的混合内容信号向量(“混合音频(T/F)”)进行组合以生成语音增强的混合内容信号向量。

[0299] 在一些实施方式中,(例如,从图3的编码器20等输出的)编码音频信号的语法支持M/S标记从上游音频编码器(例如,图3的编码器20等)到下游音频解码器(例如,图3的解码器40等)的传输。当接收方音频解码器(例如,图3的解码器40等)至少部分地使用与M/S标记一起被传输的M/S控制数据、控制参数等来执行语音增强操作时,M/S标记由音频编码器呈现/设置(例如,图3的编码器20中的元件23等)。例如,当M/S标记被设置时,在根据语言增强算法(例如,独立通道对话预测、多通道对话预测、基于波形的波形参数混合等)中的一个或更多个来使用如与M/S标记一起所接收的M/S控制数据、控制参数等来应用M/S语音增强操作之前,接收方音频解码器(例如,图3的解码器40等)可以首先将非M/S通道中的立体声信号(例如,来自左通道和右通道等)转换成M/S表示的中间通道和侧通道。在接收方音频解码器(例如,图3的解码器40等)中,在执行M/S语言增强操作之后,可以将M/S表示中的语音增强信号转换回非M/S通道。

[0300] 在一些实施方式中,由如本文中所描述的音频编码器(例如,图3的编码器20、图3的编码器20的元件23等)生成的语音增强元数据可以携载指示针对一个或更多个不同类型的语音增强操作的语音增强控制数据、控制参数等的一个或更多个集合的存在的一个或更多个特定标记。针对一个或更多个不同类型的语音增强操作的语音增强控制数据、控制参数等的一个或更多个集合可以但不限于仅包括作为M/S语音增强元数据的M/S控制数据、控制参数等的集合。语音增强元数据还可以包括指示对于要被语音增强的音频内容而言优选哪种类型的语音增强操作(例如,M/S语音增强操作、非M/S语音增强操作等)的优选标记。可以将语音增强元数据作为在包括针对非M/S参考音频通道配置编码的混合音频内容的编码音频信号中所递送的元数据的一部分递送至下游解码器(例如,图3的解码器40等)。在一些实施方式中,仅M/S语音增强元数据而不是非M/S语音增强元数据被包括在编码音频信号中。

[0301] 另外,可选地或替选地,音频解码器(例如,图3的40等)可以被配置成基于一个或更多个因素来确定并执行特定类型的语音增强操作(例如,M/S语音增强、非M/S语音增强等)。这些因素可以包括但不限于仅下述中的一个或更多个:指定对特定用户选择类型的语

音增强操作的偏好的用户输入;指定对系统选择类型的语音增强操作的偏好的用户输入;由音频解码器操作的特定音频通道配置的能力;用于特定类型的语音增强操作的语音增强元数据的可用性;针对一种类型的语音增强操作的任意编码器生成的优选标记等。在一些实施方式中,音频解码器可以实现一个或更多个优先规则,如果这些因素之间冲突,则可以请求进一步的用户输入等以确定特定类型的语音增强操作。

[0302] 7.示例处理流程

[0303] 图8A和图8B示出了示例处理流程。在一些实施方式中,媒体处理系统中的一个或更多个计算装置或单元可以执行该处理流程。

[0304] 图8A示出了可以由如本文中所描述的音频编码器(例如,图3的编码器20)实现的示例处理流程。在图8A的块802中,音频编码器接收在参考音频通道表示中具有语音内容与非语音音频内容的混合的混合音频内容,该混合音频内容被分布在参考音频通道表示的多个音频通道中。

[0305] 在块804中,音频编码器将参考音频通道表示的多个音频通道中的一个或更多个非中间/侧(M/S)通道上分布的混合音频内容的一个或更多个部分转换成M/S音频通道表示的一个或更多个M/S通道上分布的M/S音频通道表示中的一个或更多个转换混合音频内容部分。

[0306] 在块806中,音频编码器确定针对M/S音频通道表示中的一个或更多个转换混合音频内容部分的M/S语音增强元数据。

[0307] 在块808中,音频编码器生成音频信号,该音频信号包括参考音频通道表示中的混合音频内容、以及M/S音频通道表示中的一个或更多个转换混合音频内容部分的M/S语音增强元数据。

[0308] 在实施方式中,音频编码器还被配置成执行:生成M/S音频通道表示中的与混合音频内容分立的语音内容的版本;以及输出使用M/S音频通道表示中的语音内容的版本所编码的音频信号。

[0309] 在实施方式中,音频编码器还被配置成执行:生成混和指示数据,该混和指示数据使得接收方音频解码器能够使用基于M/S音频通道表示中的语音内容的版本的波形编码语音增强与基于M/S音频通道表示中的语音内容的重构版本的参数语音增强的特定量组合来对混合音频内容应用语音增强;以及输出使用混和指示数据所编码的音频信号。

[0310] 在实施方式中,音频编码器还配置成阻止将M/S音频通道表示中的一个或更多个转换混合音频内容部分编码为音频信号的一部分。

[0311] 图8B示出了可以由如本文中所描述的音频解码器(例如,图3的解码器40)来实现的示例处理流程。在图8B的块822中,音频解码器接收包括参考音频通道表示中的混合音频内容以及中间/侧(M/S)语音增强元数据的音频信号。

[0312] 在图8B的块824中,音频解码器将参考音频通道表示的多个音频通道中的一个、两个或更多个非M/S通道上分布的混合音频内容的一个或更多个部分转换成M/S音频通道表示的一个或更多个M/S通道上分布的M/S音频通道表示中的一个或更多个转换混合音频内容部分。

[0313] 在图8B的块826中,音频解码器基于M/S语音增强元数据对M/S音频通道表示中的一个或更多个转换混合音频内容部分执行一个或更多个M/S语音增强操作,以生成M/S表示

中的一个或更多个增强语音内容部分。

[0314] 在图8B的块828中,音频解码器将M/S音频通道表示中的一个或更多个转换混合音频内容部分与M/S表示中的一个或更多个增强语音内容进行组合,以生成M/S表示中的一个或更多个语音增强混合音频内容部分。

[0315] 在实施方式中,音频解码器还被配置成将M/S表示中的一个或更多个语音增强混合音频内容部分逆转换成参考音频通道表示中的一个或更多个语音增强混合音频内容部分。

[0316] 在实施方式中,音频解码器还被配置成执行:从音频信号中提取M/S音频通道表示中的与混合音频内容分立的语音内容的版本;以及基于M/S语音增强元数据对M/S音频通道表示中的语音内容的版本的一个或更多个部分来执行一个或更多个语音增强操作,以生成M/S音频通道表示中的一个或更多个第二增强语音内容部分。

[0317] 在实施方式中,音频解码器还被配置成执行:确定用于语音增强的混和指示数据;以及基于用于语音增强的混和指示数据,生成基于M/S音频通道表示中的语音内容的版本的波形编码语音增强与基于M/S音频通道表示中的语音内容的重构版本的参数语音增强的特定量组合。

[0318] 在实施方式中,至少部分地基于针对M/S音频通道表示中的一个或更多个转换混合音频内容部分的一个或更多个SNR值来生成混和指示数据。一个或更多个SNR值表示下述功率比中的一个或更多个功率比:M/S音频通道表示中的一个或更多个转换混合音频内容部分的语音内容与非语音音频内容的功率比;或者M/S音频通道表示中的一个或更多个转换混合音频内容部分的语音内容与总音频内容的功率比。

[0319] 在实施方式中,使用以下听觉掩蔽模型来确定基于M/S音频通道表示中的语音内容的版本的波形编码语音增强与基于M/S音频通道表示中的语音内容的重构版本的参数语音增强的特定量组合,在该听觉掩蔽模型中,基于M/S音频通道表示中的语音内容的版本的波形编码语音增强表示波形编码语音增强与参数语音增强的多个组合中的、确保输出语音增强的音频节目中的编码噪声不听起来令人讨厌的最大相对语音增强量。

[0320] 在实施方式中,M/S语音增强元数据的至少一部分使得接收方音频解码器能够根据参考音频通道表示中的混合音频内容来重构M/S表示中的语音内容的版本。

[0321] 在实施方式中,M/S语音增强元数据包括与M/S音频通道表示中的波形编码语音增强操作或者M/S音频通道中的参数语音增强操作中的一个或更多个有关的元数据。

[0322] 在实施方式中,参考音频通道表示包括与环绕扬声器有关的音频通道。在实施方式中,参考音频通道表示的一个或更多个非M/S通道包括中央通道、左通道、或者右通道中的一个或更多个,而M/S音频通道表示的一个或更多个M/S通道包括中间通道或侧通道中的一个或更多个。

[0323] 在实施方式中,M/S语音增强元数据包括与M/S音频通道表示的中间通道有关的单个语音增强元数据的集合。在实施方式中,M/S语音增强元数据表示编码在音频信号中的全部音频元数据的一部分。在实施方式中,编码在音频信号中的音频元数据包括指示M/S语音增强元数据的存在的数据字段。在实施方式中,音频信号是音视频信号的一部分。

[0324] 在实施方式中,包括处理器的设备被配置成执行如本文中所描述的方法中任意一种方法。

[0325] 在实施方式中,一种非暂态计算机可读存储介质,其包括以下软件指令:所述软件指令当由一个或更多个处理器执行时使得执行如本文中所描述的方法中的任一方法。注意,虽然本文中讨论了单独的实施方式,但是可以将本文中所讨论的实施方式的任意组合和/或部分实施方式进行组合以形成另外的实施方式。

[0326] 在一些实施方式中,方法或设备可以涉及接收混合音频内容。混合音频内容至少包括中间通道混合内容信号和侧通道混合内容信号。中间通道信号表示参考音频通道表示的两个通道的加权和或非加权和。侧通道信号表示参考音频通道表示的两个通道的加权差或非加权差。音频解码器将中间通道信号和侧通道信号解码成左通道信号和右通道信号。解码包括基于语音增强元数据进行解码。语音增强元数据包括指示在解码期间要对中间通道信号和侧通道信号执行的至少一种类型的语音增强操作的优选标记。语音增强元数据还指示用于中间通道信号的第一类型的语音增强和中间通道信号的第二类型的语音增强。生成音频信号,其中,该音频信号包括针对混合音频内容的解码的中间通道信号和侧通道信号的一个或更多部分的左通道信号和右通道信号。语音增强元数据可以包括与波形编码语音增强操作或者参数语音增强操作中的一个或更多个有关的元数据。混合音频内容可以包括参考音频通道表示,该参考音频通道表示包括与环绕扬声器有关的音频通道。语音增强元数据可以包括与中间通道信号有关的单个语音增强元数据的集合。语音增强元数据可以表示混合音频内容的全部音频元数据的一部分。编码在混合音频内容中的音频元数据可以包括指示语音增强元数据的存在的数据字段。混合音频内容可以是音视频信号的一部分。

[0327] 8.实现机构——硬件概述

[0328] 根据一种实施方式,本文中描述的技术由一个或多个专用计算设备来实现。专用计算设备可以是硬连线的以执行技术,或者可以包括诸如永久地被编程成执行技术的一个或多个专用集成电路(ASIC)或现场可编程门阵列(FPGA)的数字电子设备,或者可以包括被编程成根据固件、存储器、其他存储装置或其组合中的程序指令执行技术的一个或多个通用硬件处理器。这样的专用计算设备还可以将定制的硬连线逻辑、ASIC或FPGA与定制的编程进行组合以实现技术。专用计算设备可以是台式计算机系统、便携式计算机系统、手持式设备、连网设备或合并硬连线和/或程序逻辑以实现技术的任何其他设备。

[0329] 例如,图9是图示了可以在其上实现本发明的实施方式的计算机系统900的框图。计算机系统900包括用于传送信息的总线902或其他通信机构,以及用于处理信息的与总线902耦接的硬件处理器904。硬件处理器904例如可以是通用微处理器。

[0330] 计算机系统900还包括用于存储要由处理器904执行的信息和指令的、与总线902耦接的诸如随机存取存储器(RAM)或其他动态存储设备的主存储器906。主存储器906还可以用于在执行要由处理器904执行的指令期间存储临时变量或其他中间信息。当这样的指令被存储在处理器904能够访问的非暂态存储介质中时,这样的指令使计算机系统900成为专用机器,该专用机器是专用于执行指令中指定的操作的设备。

[0331] 计算机系统900还包括用于存储处理器904的静态信息和指令的、与总线902耦接的只读存储器(ROM)908或其他静态存储设备。诸如磁盘或光盘的存储设备910被设置并且耦接至总线902以存储信息和指令。

[0332] 计算机系统900可以经由总线902耦接至诸如液晶显示器(LCD)的显示器912,以向计算机用户显示信息。包括字母数字和其他键的输入设备914耦接至总线902,以向处理器

904传送信息和命令选择。另一类型的用户输入设备是用于向处理器904传送方向信息和命令选择并且用于控制显示器912上的光标运动诸如鼠标、跟踪球或光标方向键的光标控件916。该输入设备通常具有在两个轴,第一轴(例如,x)和第二轴(例如,y)上的两个自由度,这允许设备指定平面中的位置。

[0333] 计算机系统900可以使用与计算机系统结合致使或编程计算机系统900成为专用机器的设备特定硬连线逻辑、一个或多个ASIC或FPGA、固件和/或程序逻辑,来实现本文中描述的技术。根据一个实施方式,计算机系统900可以响应于处理器904执行主存储器906中包括的一个或多个指令的一个或多个序列来执行本文中的技术。这样的指令可以从诸如存储设备910的另一存储介质被读入主存储器906中。主存储器906中包括的指令序列的执行使处理器904执行本文中描述的处理步骤。在替选实施方式中,可以使用硬连线电路代替软件指令,或者可以将硬连线电路与软件指令结合使用。

[0334] 如本文中使用的术语“存储介质”指代存储使机器能够以特定方式进行操作的数据和/或指令的任意非暂态介质。这样的存储介质可以包括非易失性介质和/或易失性介质。非易失性介质包括例如诸如存储设备910的光盘或磁盘。易失性介质包括诸如主存储器906的动态存储器。存储介质的常见形式包括例如软盘、软磁盘、硬盘、固态驱动器、磁带或任何其他磁数据存储介质、CD-ROM、任何其他光数据存储介质、具有孔图案的任何物理介质、RAM、PROM和EPROM、闪速EPROM、NVRAM、任何其他存储器芯片或盒式磁带。

[0335] 存储介质与传输介质不同,但是可以与传输介质结合使用。传输介质参与在存储介质之间传输信息。例如,传输介质包括同轴线缆、铜线和光纤,包括具有总线902的引线。传输介质还能够采用诸如在无线电波和红外线数据通信期间生成的那些声波或光波的声波或光波的形式。

[0336] 各种形式的介质可以涉及:向处理器904传送一个或多个指令的一个或多个序列以用于执行。例如,最初可以将指令携载在远程计算机的磁盘或固态驱动器上。远程计算机能够将指令加载至其动态存储器中并且使用调制解调器在电话线路上发送指令。计算机系统900本地的调制解调器能够接收电话线路上的数据并且使用红外线发送器将数据转换成红外线信号。红外线检测器能够接收红外线信号中携载的数据,并且适当的电路可以将数据放置在总线902上。总线902将数据携载至主存储器906,处理器904从该主存储器取回指令并执行指令。在处理器904执行之前或之后,由主存储器906接收的指令可以可选地存储在存储设备910上。

[0337] 计算机系统900还包括与总线902耦接的通信接口918。通信接口918提供耦接至与本地网络922连接的网络链路920的双向数据通信。例如,通信接口918可以是综合业务数字网(ISDN)卡、有线调制解调器、卫星调制解调器或向相应类型的电话线路提供数据通信连接的调制解调器。作为另一示例,通信接口918可以是提供至兼容LAN的数据通信连接的局域网(LAN)卡。还可以实现无线链路。在任何这样的实现中,通信接口918发送并接收携载表示各种类型的信息的数字数据流的电信号、电磁信号或光信号。

[0338] 网络链路920通常通过一个或多个网络向其他数据设备提供数据通信。例如,网络链路920可以通过本地网络922向由因特网服务提供商(ISP)926操作的数据设备或主计算机924提供连接。ISP 926进而通过现在通常称为“因特网”928的全球分组数据通信网络提供数据通信服务。本地网络922和因特网928都使用携载数字数据流的电信号、电磁信号或

光信号。向计算机系统900携载数字数据或从计算机系统900携载数字数据的通过各种网络的信号以及网络链路920上和通过通信接口918的信号是传输介质的示例形式。

[0339] 计算机系统900可以通过网络、网络链路920和通信接口918发送消息并且接收数据,包括程序代码。在因特网示例中,服务器930可以通过因特网928、ISP 926、本地网络922和通信接口918来传输应用程序的请求代码。

[0340] 当代码被接收和/或存储在存储设备910或其他非易失性存储设备中以供稍后执行时,所接收的代码可以由处理器904执行。

[0341] 9.等同方案、扩展方案、替代方案和其他方案

[0342] 在前面的说明中,已经参考可以根据实现而变化的许多特定细节描述了本发明的实施方式。因此,本发明是什么以及本发明的申请人所期望的唯一且排他的指示是以这样的权利要求提出的特定形式而从本申请提出的权利要求组,包括任何后续改正。针对在这样的权利要求中包括的术语,本文中明确阐述的任何定义应约束如在权利要求中使用的这样的术语的含义。因此,权利要求中未明确记载的限制、要素、特性、特征、优点或属性不应以任何方式对这样的权利要求的范围进行限制。因此,说明书和附图应被视为说明性意义而不是限制性意义。

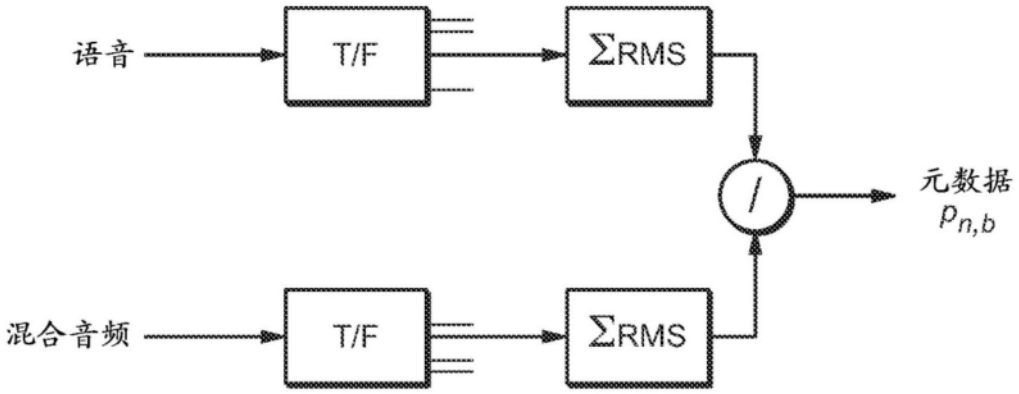


图1

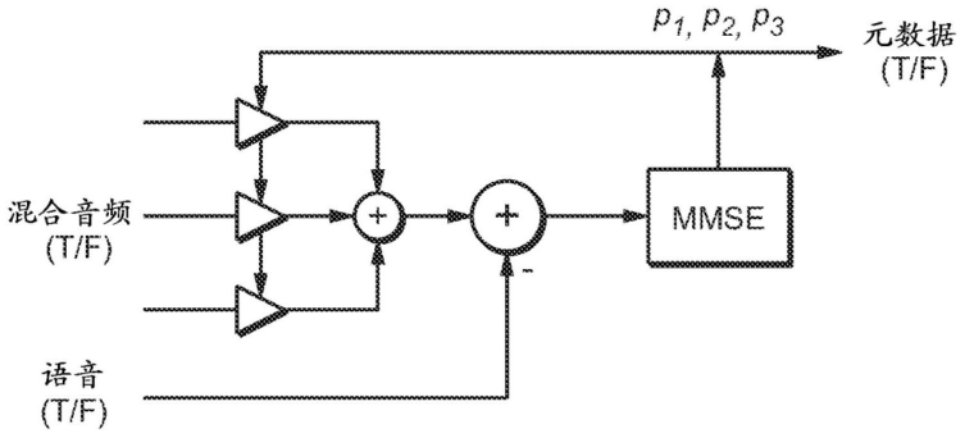


图2

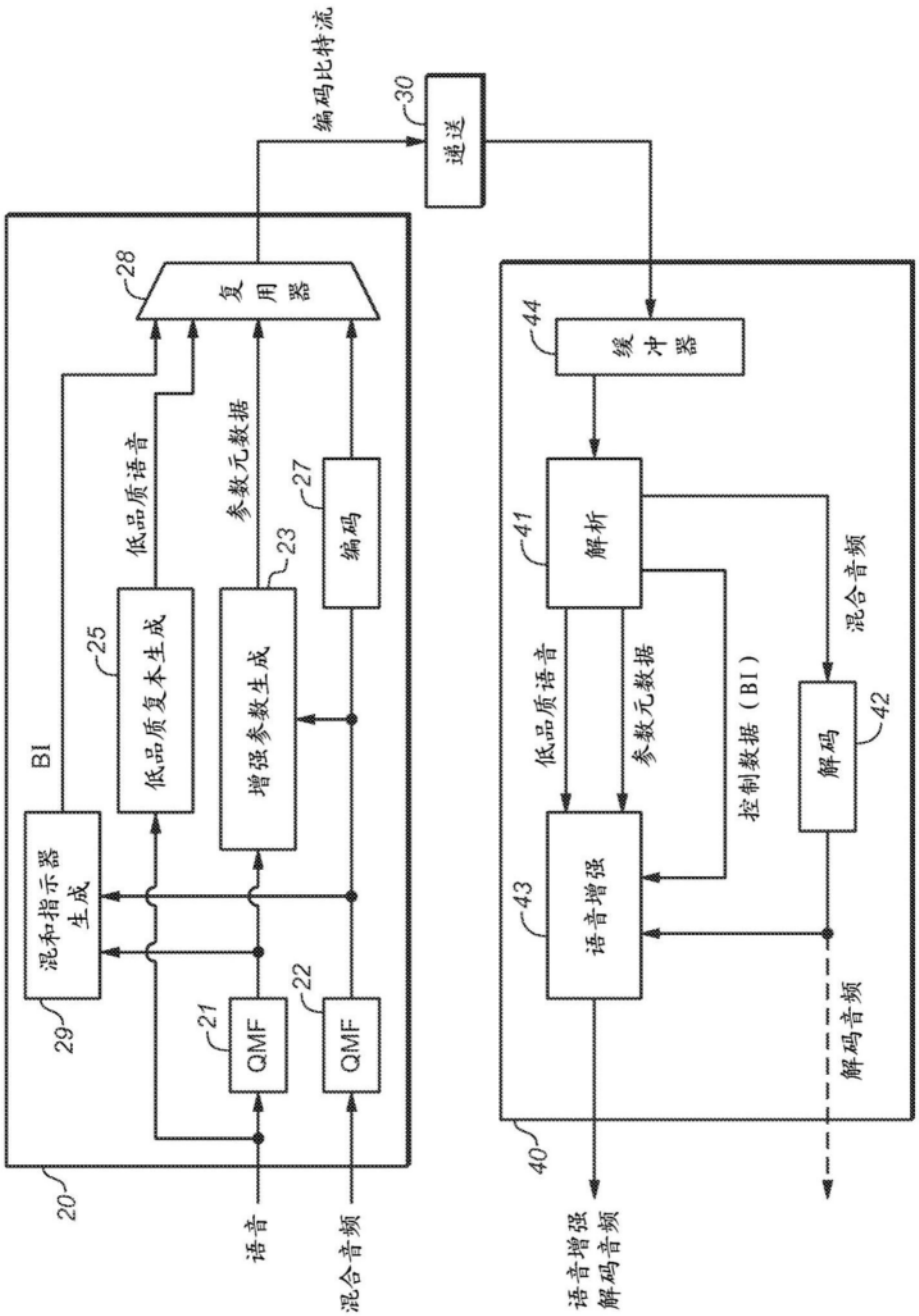

编码比特流
30
递送
28
复用器
BI
低品质语音
参数元数据
27
编码
23
25
低品质复本生成
增强参数生成
混和指示器生成
29
21
QMF
22
QMF
20
语音
混合音频
44
缓冲器
41
解析
低品质语音
参数元数据
控制数据（BI）
混合音频
42
解码
43
语音增强
解码音频
40
语音增强解码音频

图3

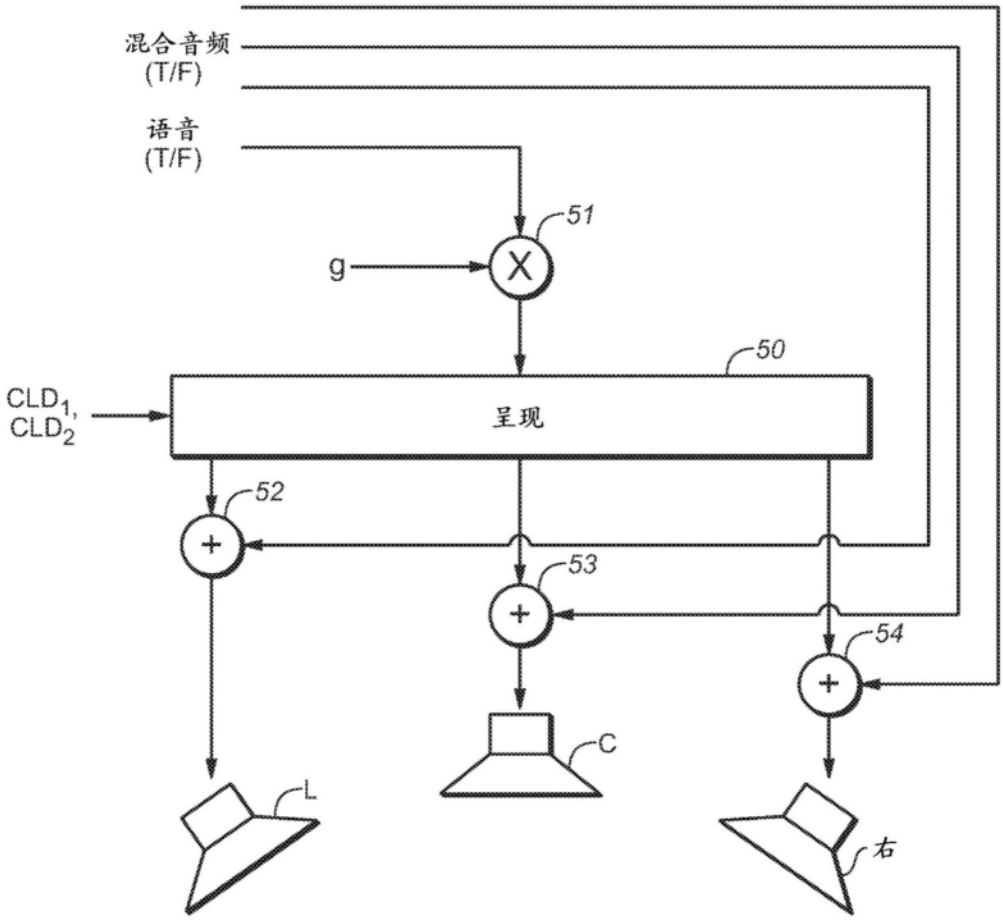


图4

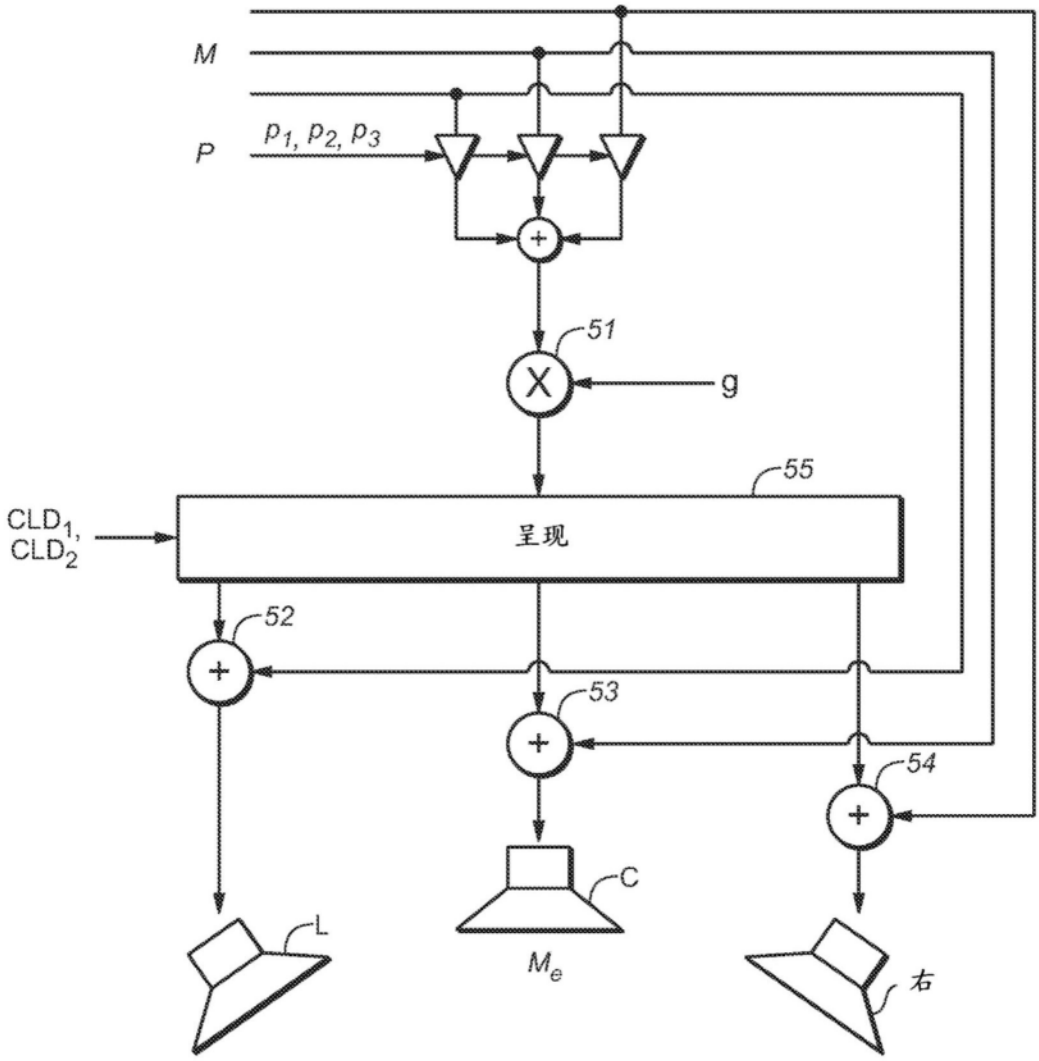


图5

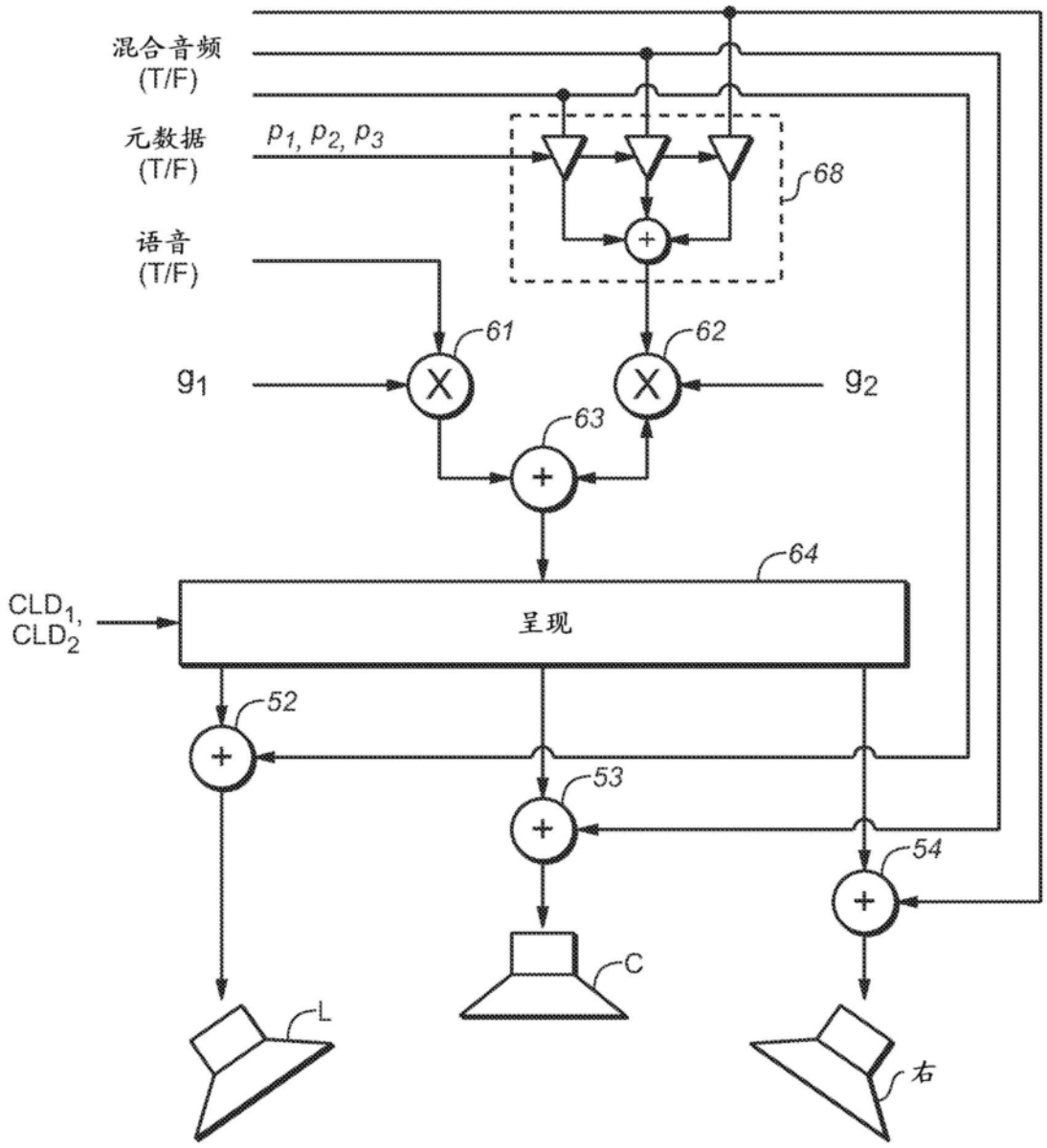

混合音频
(T/F)
元数据
(T/F)
p_1, p_2, p_3
68
语音
(T/F)
61
62
g_1
g_2
63
64
CLD_1,
CLD_2
呈现
52
53
54
C
L
右

图6

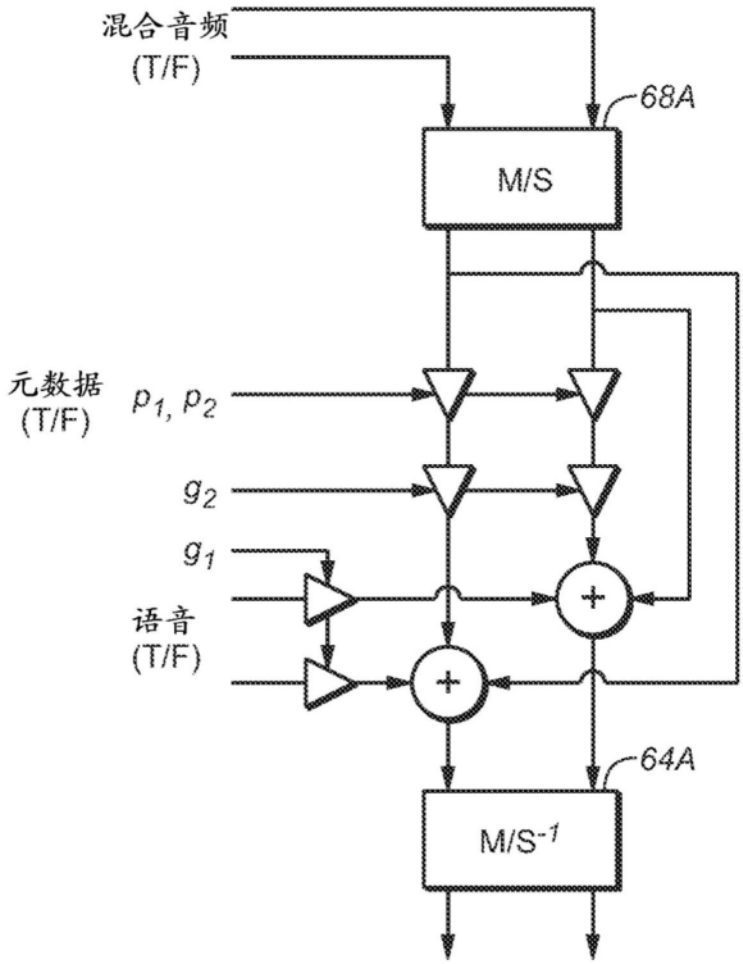


图6A

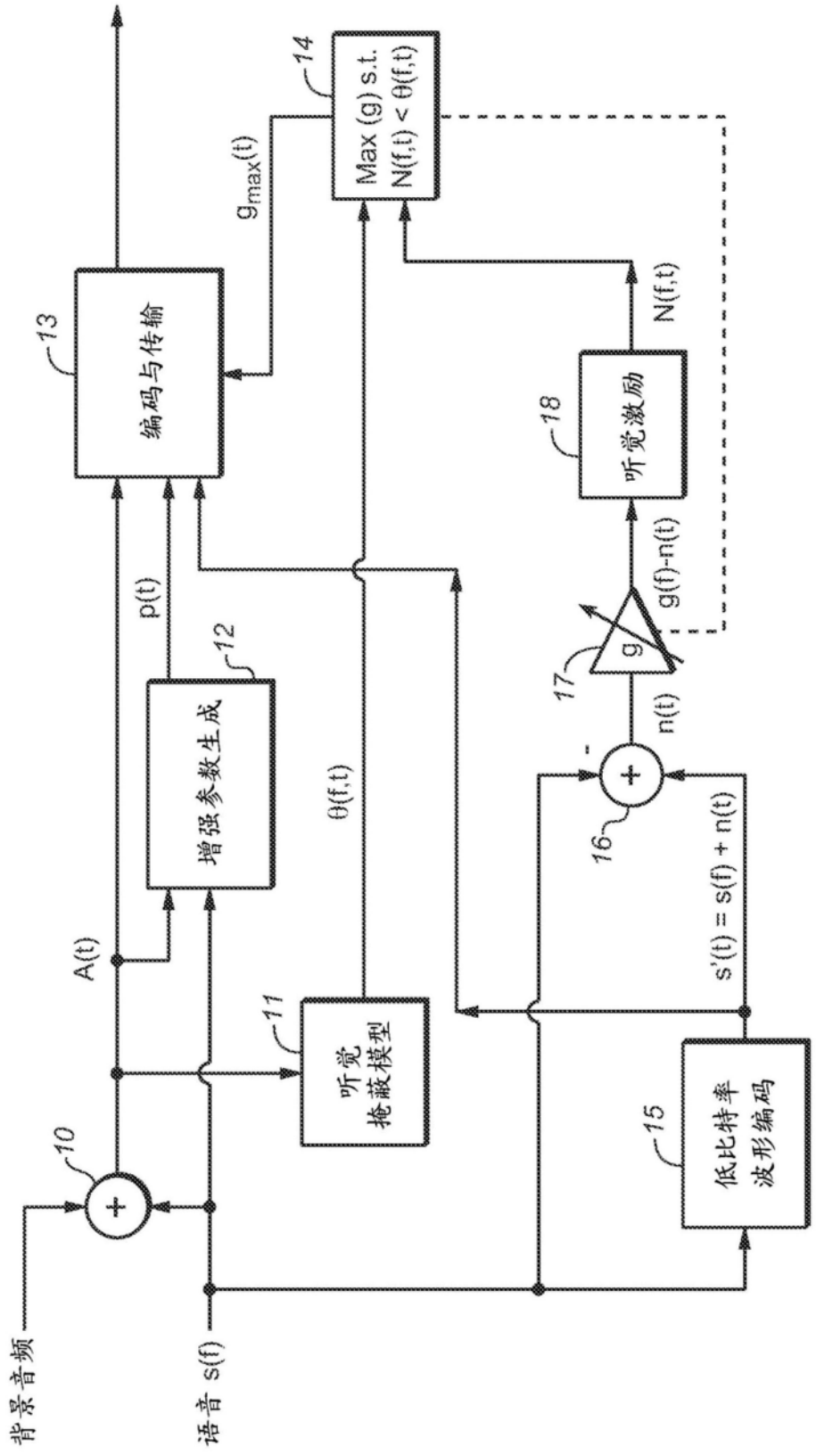

背景音频
语音 s(f)
10
A(t)
p(t)
13
编码与传输
$g_{max}(t)$
14
Max (g) s.t.
N(f,t) < θ(f,t)
12
增强参数生成
θ(f,t)
11
听觉
掩蔽模型
15
低比特率
波形编码
s'(t) = s(f) + n(t)
16
n(t)
17
g
g(f)-n(t)
18
听觉激励
N(f,t)

图7

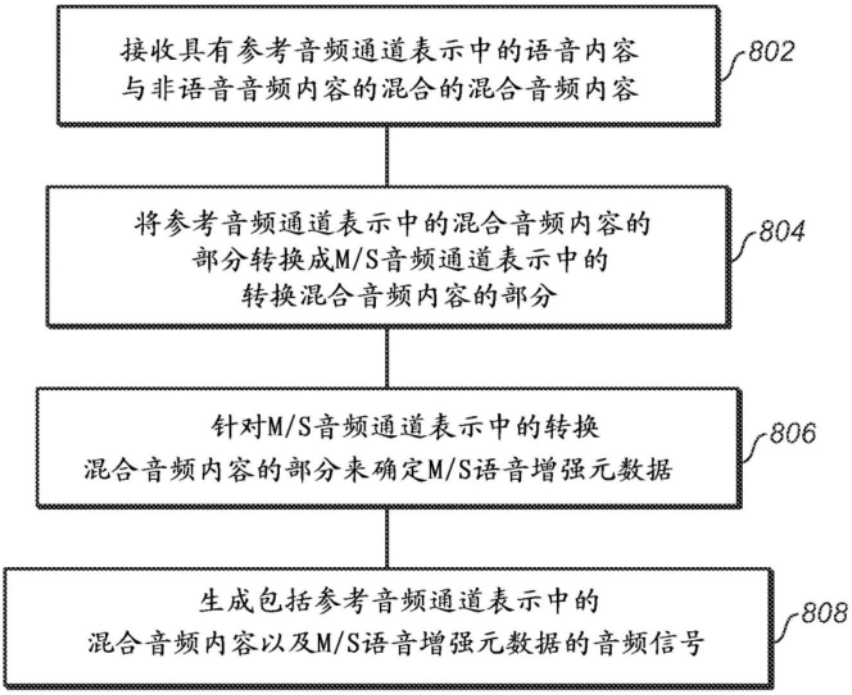

接收具有参考音频通道表示中的语音内容与非语音音频内容的混合的混合音频内容
802
将参考音频通道表示中的混合音频内容的部分转换成M/S音频通道表示中的转换混合音频内容的部分
804
针对M/S音频通道表示中的转换混合音频内容的部分来确定M/S语音增强元数据
806
生成包括参考音频通道表示中的混合音频内容以及M/S语音增强元数据的音频信号
808

图8A

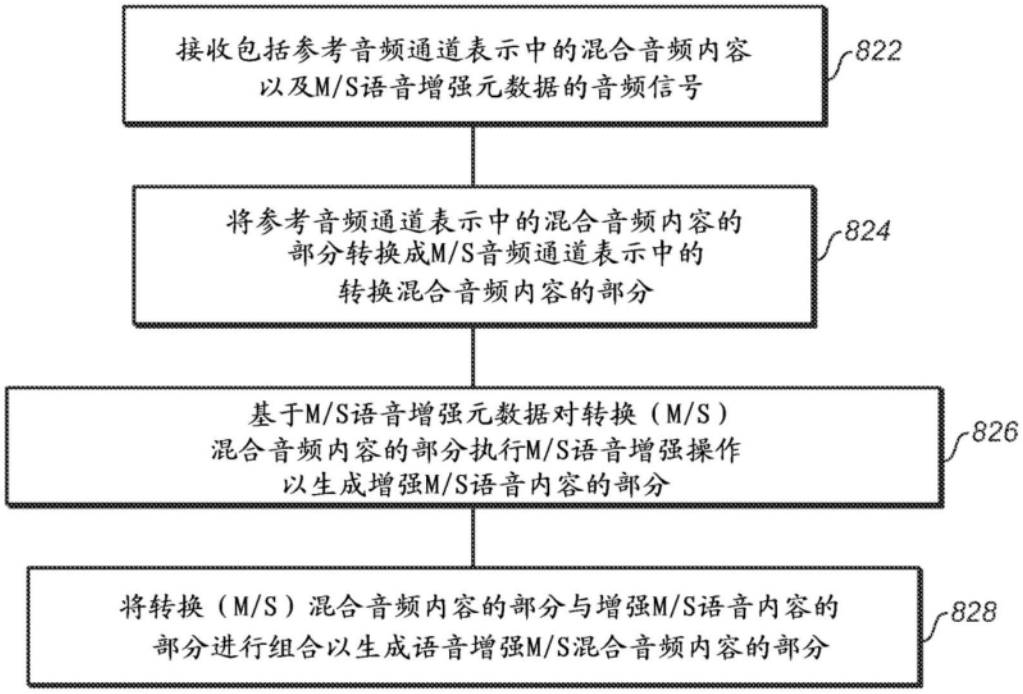

接收包括参考音频通道表示中的混合音频内容以及M/S语音增强元数据的音频信号
822
将参考音频通道表示中的混合音频内容的部分转换成M/S音频通道表示中的转换混合音频内容的部分
824
基于M/S语音增强元数据对转换（M/S）混合音频内容的部分执行M/S语音增强操作以生成增强M/S语音内容的部分
826
将转换（M/S）混合音频内容的部分与增强M/S语音内容的部分进行组合以生成语音增强M/S混合音频内容的部分
828

图8B

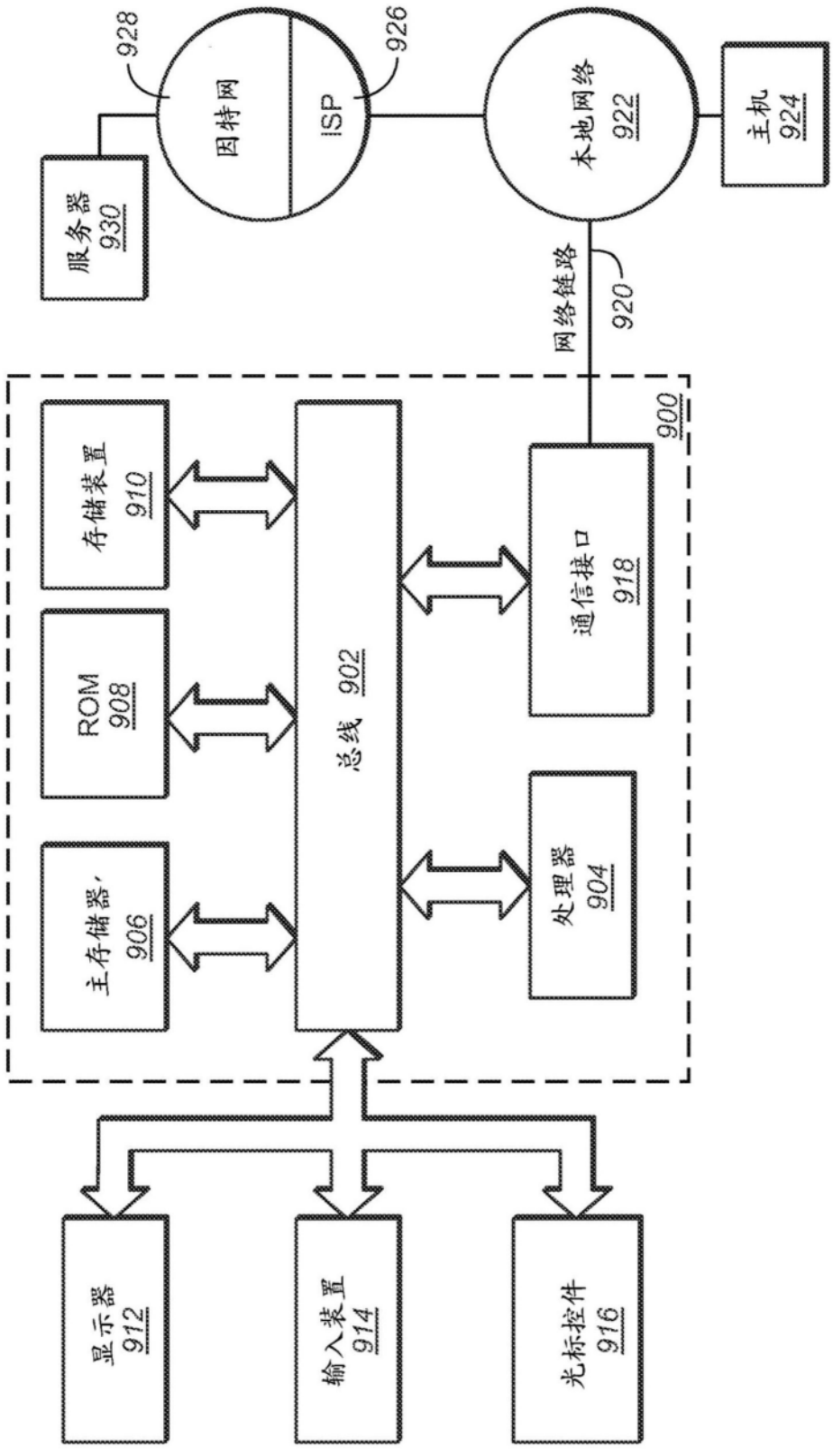

928
因特网
ISP
926
本地网络
922
主机
924
服务器
930
网络链路
920
900
存储装置
910
ROM
908
主存储器，
906
总线 902
通信接口
918
处理器
904
显示器
912
输入装置
914
光标控件
916

图9