



(12) 发明专利

(10) 授权公告号 CN 115699172 B

(45) 授权公告日 2025. 07. 08

(21) 申请号 202080101547.4

约翰娜·鲍姆加特纳-克罗内

(22) 申请日 2020.05.29

(74) 专利代理机构 中科专利商标代理有限责任

(65) 同一申请的已公布的文献号

公司 11021

申请公布号 CN 115699172 A

专利代理师 罗松梅 潘剑颖

(43) 申请公布日 2023.02.03

(51) Int.Cl.

(85) PCT国际申请进入国家阶段日

G10L 25/69 (2006.01)

2022.11.29

H04R 25/00 (2006.01)

(86) PCT国际申请的申请数据

PCT/EP2020/065035 2020.05.29

(56) 对比文件

JP 2010160246 A, 2010.07.22

(87) PCT国际申请的公布数据

W02021/239255 EN 2021.12.02

CN 102194460 A, 2011.09.21

审查员 董立静

(73) 专利权人 弗劳恩霍夫应用研究促进协会

地址 德国慕尼黑

(72) 发明人 扬·雷尼斯-霍赫穆特

权利要求书4页 说明书12页 附图3页

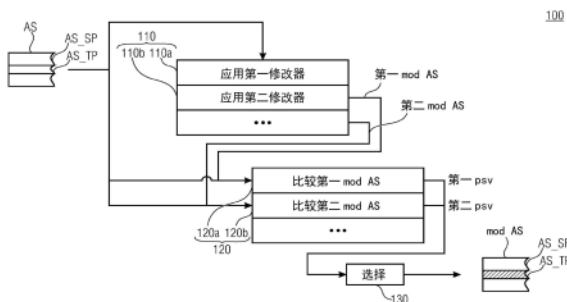
(54) 发明名称

用于处理初始音频信号的方法和装置

AS) 或第二修改音频信号 (第二MOD AS)。

(57) 摘要

一种用于处理包括目标部分 (AS_TP) 和侧边部分 (AS_SP) 的初始音频信号 (AS) 的方法 (100), 包括以下步骤: 接收初始音频信号 (AS); 通过使用第一信号修改器修改所接收的初始音频信号 (AS) 以获得第一修改 (110a) 音频信号, 并通过使用第二信号修改器修改所接收的初始音频信号 (AS) 以获得第二修改音频信号 (第二MOD AS); 将所接收的初始音频信号 (AS) 与第一修改音频信号 (第一MOD AS) 进行比较以获得第一感知相似度值 (第一PSV), 该第一PSV描述初始音频信号 (AS) 与第一修改音频信号 (第一MOD AS) 之间的感知相似度; 以及将所接收的初始音频信号 (AS) 与第二修改音频信号 (第二MOD AS) 进行比较以获得第二感知相似度值 (第二PSV), 该第二PSV描述初始音频信号 (AS) 与第二修改音频信号 (第二MOD AS) 之间的感知相似度; 以及取决于相应的第一感知相似度值或第二感知相似度值 (第二PSV) 来选择 (130) 第一修改音频信号 (第一MOD



1. 一种用于处理包括目标部分和侧边部分的初始音频信号的方法 (100), 包括以下步骤:

a. 接收所述初始音频信号;

b. 通过使用第一信号修改器来修改 (110、110a) 所接收的初始音频信号以获得第一修改音频信号;

通过使用第二信号修改器来修改 (110、110b) 所接收的初始音频信号以获得第二修改音频信号;

c. 针对评估标准评估 (120、120a) 所述第一修改音频信号以获得描述所述评估标准的满足程度的第一评估值;

针对所述评估标准评估 (120、120a) 所述第二修改音频信号以获得描述所述评估标准的满足程度的第二评估值; 以及

d. 取决于相应的第一评估值或第二评估值来选择 (130) 所述第一修改音频信号或所述第二修改音频信号; 其中, 选择的步骤是基于多个独立的第一评估值和独立的第二评估值、或者基于至少两个独立的评估标准来执行的,

其中, 所述评估标准是感知相似度, 并且其中, 步骤c包括以下子步骤:

将所接收的初始音频信号与所述第一修改音频信号进行比较 (120、120a) 以获得第一感知相似度值作为第一评估值, 所述第一感知相似度值描述所述初始音频信号与所述第一修改音频信号之间的感知相似度; 以及

将所接收的初始音频信号与所述第二修改音频信号进行比较 (120、120b) 以获得第二感知相似度值作为第二评估值, 所述第二感知相似度值描述所述初始音频信号与所述第二修改音频信号之间的感知相似度。

2. 根据权利要求1所述的方法 (100), 其中, 所述评估标准来自包括以下各项的组:

- 感知相似度, 由第一感知相似度值和第二感知相似度值描述, 所述第一感知相似度值和所述第二感知相似度值描述相应的第一修改音频信号和第二修改音频信号与所述初始音频信号之间的感知相似度;

- 语音清晰度, 以语音清晰度的计算值的形式与目标或阈值进行比较;

- 响度, 由响度值描述;

- 声音模式;

- 空间感。

3. 根据权利要求1所述的方法 (100), 其中, 所述至少两个独立的评估标准是被单独评估的, 使得针对所述第一修改音频信号描述所述至少两个独立的评估标准的满足程度的相应第一评估值和针对所述第二修改音频信号描述所述至少两个独立的评估标准的满足程度的相应第二评估值被确定, 其中, 然后所述选择是基于加权的第一评估值和加权的第二评估值来执行的。

4. 根据权利要求1所述的方法 (100), 其中, 选择所述第一修改音频信号, 其中, 所述第一感知相似度值高于所述第二感知相似度值以便指示所述第一修改音频信号的更高感知相似度; 以及

其中, 当所述第二感知相似度值高于所述第一感知相似度值以便指示所述第二修改音频信号的更高感知相似度时, 选择所述第二修改音频信号。

5. 根据权利要求1所述的方法(100),还包括以下步骤:取决于步骤d的选择来输出所述第一修改音频信号或所述第二修改音频信号。

6. 根据权利要求3所述的方法(100),其中,当相应的第一感知相似度值或第二感知相似度值低于阈值时,执行输出所述初始音频信号而不是输出所述第一修改音频信号或所述第二修改音频信号的步骤,其中,低于所述阈值,相应的第一修改音频信号或第二修改音频信号被指示为与所述初始音频信号不够相似。

7. 根据权利要求1所述的方法(100),其中,所述目标部分是所述初始音频信号的语音部分,并且所述侧边部分是所述初始音频信号的环境噪声部分。

8. 根据权利要求1所述的方法(100),其中,所述第一修改音频信号和/或所述第二修改音频信号包括:被移动到前景中的所述目标部分和被移动到背景中的所述侧边部分,和/或作为所述目标部分被移动到前景中的语音部分和作为所述侧边部分被移动到背景中的环境噪声部分。

9. 根据权利要求1所述的方法(100),其中,比较的步骤包括:通过使用感知模型、PEAQ模型、POLQA模型和/或PEMO-Q模型来提取所述第一评估值和/或所述第二评估值。

10. 根据权利要求1所述的方法(100),其中,所述第一评估值和/或所述第二评估值取决于所述第一修改音频信号或所述第二修改音频信号的物理参数、所述第一修改音频信号或所述第二修改音频信号的音量电平、所述第一修改音频信号或所述第二修改音频信号的心理声学参数、所述第一修改音频信号或所述第二修改音频信号的响度信息、所述第一修改音频信号或所述第二修改音频信号的音调信息、和/或所述第一修改音频信号或所述第二修改音频信号的感知源宽度信息。

11. 根据权利要求1所述的方法(100),其中,所述第一信号修改器和/或所述第二信号修改器被配置为执行SNR增加、动态压缩、所述初始音频信号的SNR增加、和/或所述初始音频信号的动态压缩;和/或

其中,如果所述初始音频信号包括单独的目标部分和单独的侧边部分,则修改的步骤包括:增加所述目标部分、增加对所述目标部分的频率加权、动态压缩所述目标部分、减少所述侧边部分、减少对所述侧边部分的频率加权;和/或

其中,如果所述初始音频信号包括组合的目标部分和侧边部分,则修改包括:执行对所述目标部分和所述侧边部分的分离。

12. 根据权利要求1所述的方法(100),其中,选择(130)的步骤是考虑到以下因素中的一个或多个而执行的:

- 听力受损者的听力损失等级;
- 个人听力表现;
- 个人频率相关听力表现;
- 个人偏好;
- 关于信号修改率的个人偏好。

13. 根据权利要求1所述的方法(100),其中,修改(110)和/或比较(120)的步骤是考虑到以下因素中的一个或多个而执行的:

- 听力受损者的听力损失等级;
- 个人听力表现;

个人频率相关听力表现；

个人偏好；

关于信号修改率的个人偏好。

14. 根据权利要求1所述的方法(100), 其中, 所述方法还包括以下步骤: 接收关于定义个人偏好的优化目标的信息; 其中, 所述评估标准取决于所述优化目标; 或者其中, 修改和/或评估和/或选择的步骤取决于所述优化目标; 或者其中, 对用于选择步骤的描述独立的评估标准的独立的第一评估值和第二评估值的加权取决于所述优化目标。

15. 根据权利要求1所述的方法(100), 其中, 比较(120)的步骤是针对以下内容执行的: 整个初始音频信号与整个第一修改音频信号和第二修改音频信号; 和/或

单独音频信号的目标部分与所述第一修改音频信号和所述第二修改音频信号的相应目标部分; 和/或

所述初始音频信号的侧边部分与第一修改音频部分和第二修改音频部分的侧边部分。

16. 根据权利要求1所述的方法(100), 其中, 所述初始音频信号包括多个时间帧, 并且其中, 针对每个时间帧重复步骤a至d; 和/或

其中, 针对所述初始音频信号的场景的时间部分或时间帧重复步骤a至d。

17. 根据权利要求1所述的方法(100), 其中, 对包括多个时间帧的所述初始音频信号的适配是针对需要所述适配的时间帧和其他时间帧来执行的, 以便保持感知连续性, 或者其中, 对包括多个时间帧的所述初始音频信号的适配是针对需要所述适配的时间帧并且以插值方式针对其他时间帧执行的, 以便保持感知连续性; 和/或

其中, 执行对第一后续时间帧和第二后续时间帧的适配, 使得形成所述第一后续时间帧和所述第二后续时间帧之间的过渡以便保持感知连续性。

18. 根据权利要求1所述的方法(100), 其中, 所述方法(100)还包括以下初始步骤:

分析(21)初始音频部分以便确定语音部分;

将所述语音部分与环境噪声部分进行比较, 以便评估所述初始音频信号的语音清晰度; 以及

如果指示所述语音清晰度的值低于阈值, 则激活所述第一信号修改器和/或所述第二信号修改器以进行修改的步骤。

19. 一种计算机可读存储介质, 其上存储有具有程序代码的计算机程序, 所述程序代码用于当在计算机上运行时执行根据权利要求1所述的方法。

20. 一种用于处理包括目标部分和侧边部分的初始音频信号的装置, 所述装置包括:

接口, 用于接收所述初始音频信号;

第一信号修改器(11)和第二信号修改器(11), 所述第一信号修改器(11)用于修改(110)所接收的初始音频信号以获得第一修改音频信号, 所述第二信号修改器(11)用于修改所接收的初始音频信号以获得第二修改音频信号;

评估器, 用于针对评估标准评估(120、120a)所述第一修改音频信号以获得描述所述评估标准的满足程度的第一评估值, 并且针对所述评估标准评估(120、120a)所述第二修改音频信号以获得描述所述评估标准的满足程度的第二评估值; 以及

选择器(13), 用于取决于相应的第一评估值或第二评估值来选择(130)所述第一修改音频信号或所述第二修改音频信号; 其中, 选择的步骤是基于多个独立的第一评估值和独

立的第二评估值、或者基于至少两个独立的评估标准来执行的；

其中,所述评估标准是感知相似度,并且其中,所述评估器在评估时被配置用于:将所接收的初始音频信号与所述第一修改音频信号进行比较(120、120a)以获得第一感知相似度值作为第一评估值,所述第一感知相似度值描述所述初始音频信号与所述第一修改音频信号之间的感知相似度;以及将所接收的初始音频信号与所述第二修改音频信号进行比较(120、120b)以获得第二感知相似度值作为第二评估值,所述第二感知相似度值描述所述初始音频信号与所述第二修改音频信号之间的感知相似度。

用于处理初始音频信号的方法和装置

技术领域

[0001] 本发明的实施例涉及一种用于处理初始音频信号(如录音或原始数据)的方法和对应的装置。优选实施例涉及一种用于改进语音清晰度和用于收听广播音频材料的方法(方式和算法)。

背景技术

[0002] 在制作并广播音频媒体和视听媒体(例如,电影、电视、广播、播客、YouTube视频)时,无法始终确保最终混音中的足够高的语音清晰度,例如由于添加过多的背景声音(音乐、音效、录音中的噪声等)。

[0003] 这对具有听力损伤的人来说尤其成问题,但提高语音清晰度对听力正常的人或非母语听众也有好处。

[0004] 在制作音频媒体和视听媒体时的基本问题是背景信号(音乐、音效、氛围)构成了制作中重要的声音美学部分,即不能将背景信号视为应该尽可能被消除的“干扰噪声”。因此,所有旨在针对该应用提高语音清晰度或减少收听努力的方法应该附加地考虑仅尽可能少地改变最初预期的声音特性,以考虑声音制作的高质量要求和创造性方面。然而,目前,不存在用于确保良好清晰度与保持声音场景/录音之间的最佳折衷的技术方法或工具。

[0005] 然而,存在基本上可以提高音频媒体和视听媒体的语音清晰度(或减少收听努力)的不同技术方法:

[0006] 一个解决方案是让专业音响工程师手动制作备选音频混音,使得最终用户可以在原始混音和具有改进的语音清晰度的混音之间自由选择。例如通过采用听力损失模拟并确保预期的混音也适用于具有目标听力损失的听众,可以产生具有改进的清晰度的混音[1]。然而,这种手动过程成本非常高,并且不适用于大部分制作的音频/视听媒体。

[0007] 作为提供自动信号增强的备选解决方案,存在用于减少或消除不期望的信号部分(例如,干扰噪声)的不同方法,然而,这些方法不同于本发明的技术方法:

[0008] 通过用于混合信号的干扰降噪方法改进语音清晰度:这种方法旨在处理包括目标信号(例如,语音)以及干扰信号(例如,背景噪声)二者的混合信号,使得在目标信号理想地保持原样的同时消除尽可能多的干扰噪声(例如,根据[2]的方法)。由于这些方法必须估计混合信号中目标和干扰噪声分量的相应部分,因此这些方法总是基于对信号分量的物理特性的假设。这种算法例如用于助听器和移动电话,属于现有技术并且会不断得到进一步发展。

[0009] 在过去几年中,越来越多地提出了旨在分离混合信号中的不同源的基于机器学习(神经网络)的方法。基于大量数据,这些方法针对特定问题进行训练(例如,在混音中将若干说话者分离[3]),并且基本上可以用于从视听媒体中的氛围/音乐中提取对话,因此为具有改进SNR的重新混音提供了基础。在[4]中,已经提出了用于让用户选择自己调整语音与背景的比率的这种方法。

[0010] 通过预处理语音信号改进语音清晰度:在一些应用中,目标信号(例如,语音)与其

他信号部分是分离的;因此,目标信号不是如上所述的混合信号,并且该方法不需要对与目标和干扰噪声相对应的信号分量进行任何估计。例如,火车站公告就是这种情况。同时,在信号处理层面上,干扰噪声无法受到影响,即无法消除或降低干扰噪声(例如,过往火车的噪声干扰车站公告的清晰度)。对于这样的应用场景,存在以下方法:自适应地预处理目标信号,使得目标信号的清晰度在当前存在的干扰噪声中是最佳的或改进的(例如,[5]的方法)。这种方法例如使用对目标信号的带通滤波、频率相关放大、时间延迟和/或动态压缩,并且在未(显著地)修改背景噪声/氛围时将基本上也适用于视听媒体。

[0011] 将目标和背景噪声编码为单独的音频对象:此外,存在以下方法:当编码和发送音频信号时,参数化地编码关于目标信号的信息,使得可以在接收器处进行解码期间单独地调整目标信号的能量。增加目标对象(例如,语音)相对于其他音频对象(例如,氛围)的能量可以导致改进语音清晰度[11]。

[0012] 对混合信号中的语音信号的检测和电平适配:在此之上,存在以下技术系统:识别混合信号中的语音通道并修改这些通道,目的是获得改进的语音清晰度,例如提高这些通道的音量。取决于修改的类型,仅当混合信号中同时不存在其他干扰噪声时,这才会改进语音清晰度[12]。

[0013] 降低主要不包括语音的通道:在以一个通道(通常是中心)包括大部分语音信息而其他通道(例如,左/右)主要包括背景噪声的方式混合的多通道音频信号中,一个技术解决方案包括将非语音通道衰减固定的增益(例如,6dB),并以这种方式改进信噪比(例如,声音检索系统(SRS)对话清晰度或环绕声解码器的适配的缩混规则)。

[0014] 在这种方法中,可能发生:已经非常低并且实际上对语音清晰度没有不利影响的背景噪声部分也被衰减。这可能降低整体声音美学印象,因为不再能够感知音响工程师所预期的氛围。为了防止这种情况,US 8,577,676 B2描述了一种方法,其中,非语音通道仅降低到语音清晰度的度量达到特定阈值的效果,但不会更多。此外,US 8,577,676 B2公开了一种方法,其中,计算多个频率相关衰减,每个具有语音清晰度的度量达到特定阈值的效果。然后,从多个选项中选择使背景噪声的响度最大化的选项。这是基于这尽可能最好地保持原始声音特性的假设。

[0015] 基于此,US 2016/0071527 A1描述了一种方法,其中,当与一般假设相反,非语音通道还包括相关语音信息并因此降低可能不利于清晰度时,非语音声道不被降低或不被降低太多。该文献还包括一种方法,其中,计算多个频率相关衰减并选择最大化背景噪声的响度的衰减(再次基于这尽可能最好地保持原始声音特性的假设)。

[0016] 两份美国专利文献都在其独立权利要求中描述了并非本文描述的发明所需的非常具体的方法(例如,用语音出现的概率缩放降低因子)。因此,本发明可以在不使用US 8,577,676 B2和US 2016/0071527 A1所公开的技术的情况下实现。

[0017] US 8,195,454 B2描述了一种通过使用语音活动检测(VAD)来检测音频信号中出现语音的部分的方法。然后,针对这些部分修改一个或若干参数(例如,动态范围控制、动态均衡、频谱锐化、移频、语音提取、降噪或其他语音增强动作),使得语音清晰度的度量(例如,语音清晰度指数(SII)[6])被最大化或提高到期望阈值以上。这里,可以考虑听力损失或收听者的偏好或收听环境中的噪声。

[0018] US 8,271,276 B1描述了语音片段的响度或电平适配,其中,放大因子取决于前面

的时间片段。这与本文描述的发明的核心无关,并且仅当本文描述的发明取决于前面的片段而简单地改变被识别为语音的片段的响度或电平时才会变得相关。不包括除了放大语音片段之外的音频信号的适配,例如源分离、降低背景噪声、频谱变化、动态压缩。因此,US 8,271,276 B1中公开的步骤也不是不利的。

[0019] 本发明的目的是提供一种实现在(语音)清晰度与保持声音场景之间的改进折衷的构思。

[0020] 通过独立权利要求的内容来实现该目的。

[0021] 本发明的实施例提供了一种用于处理包括目标部分(例如,语音部分)和侧边部分(例如,环境噪声)的初始音频信号的方法。该方法包括以下四个步骤:

[0022] 1.接收初始音频信号;

[0023] 2.通过使用第一信号修改器修改所接收的初始音频信号以获得第一修改音频信号,并且通过使用第二信号修改器修改所接收的初始音频信号以获得第二修改音频信号;

[0024] 3.针对评估标准评估第一修改音频信号以获得描述评估标准的满足程度的第一评估值,并且针对评估标准评估第二修改音频信号以获得描述评估标准的满足程度的第二评估值;

[0025] 4.取决于相应的第一评估值或第二评估值来选择第一修改音频信号或第二修改音频信号。

[0026] 根据实施例,评估标准可以是包括感知相似度、语音清晰度、响度、声音模式和空间感的组中的一个或多个。注意,根据实施例,选择的步骤可以基于描述独立评估标准的多个独立的第一评估值和第二评估值来执行。评估标准并且特别是选择的步骤可以取决于所谓的优化目标。因此,根据实施例,该方法包括以下步骤:接收关于定义个人偏好的优化目标的信息;其中,评估标准取决于优化目标;或者其中,修改和/或评估和/或选择的步骤取决于优化目标;或者其中,对用于选择的步骤的描述独立评估标准的独立的第一评估值和第二评估值的加权取决于优化目标。

[0027] 例如,如果优化目标是两个元素(例如,初始音频信号和修改音频信号之间的最佳语音清晰度和可容忍的感知相似度)的组合,则可以执行针对选择的加权。例如,可以分别对语音清晰度和感知相似度这两个标准进行评估,使得确定评估标准的相应评估值,其中,然后基于加权评估值来执行选择。加权取决于优化目标,反之亦然,可以通过个人偏好来设置。

[0028] 根据实施例,适配、评估和选择的步骤可以通过使用神经网络/人工智能来执行。

[0029] 根据优选实施例,假设语音清晰度是通过两个或更多个使用的修改器以足够的方式来改进的。从另一角度表达,这意味着仅考虑能够足够高地改进语音清晰度或输出语音的清晰度足够的信号的修改器。在下一步骤中,在不同修改的信号之间进行选择。对于该选择,感知相似度用作评估标准,从而步骤3和4(参见上述方法)可以被如下执行:

[0030] 3.将所接收的初始音频信号与第一修改音频信号进行比较以获得第一感知相似度值,该第一感知相似度值描述初始音频信号与第一修改音频信号之间的感知相似度;以及将所接收的初始音频信号与第二修改音频信号进行比较以获得第二感知相似度值,该第一感知相似度值描述初始音频信号与第二修改音频信号之间的感知相似度;以及

[0031] 4.取决于相应的第一感知相似度值或第二感知相似度值来选择第一修改音频信

号或第二修改音频信号。

发明内容

[0032] 根据本发明的实施例,当第一感知相似度值高于第二感知相似度值(高的第一感知相似度值指示第一修改音频信号的更高感知相似度)时,选择第一修改音频信号;反之亦然,当第二感知相似度值高于第一感知相似度值(高的第二感知相似度值指示第二修改音频信号的更高感知相似度)时,选择第二修改音频信号。根据另外的实施例,代替感知相似度值,可以使用另一值,如响度值。

[0033] 可以根据另外的实施例通过在步骤2之后并在步骤3之前的针对另一优化标准(例如,针对语音清晰度)评估第一修改信号和第二修改信号的附加步骤来增强具有基于感知相似度值的比较步骤3和选择步骤4的这种适配方法。如上所述,在这种情况下可以不考虑一些修改信号,因为例如当语音清晰度太低时未(充分地)满足该第一评估标准。备选地,可以在选择未加权或加权的步骤期间考虑所有评估标准。该加权可以由用户来选择。

[0034] 根据实施例,该方法还包括以下步骤:取决于选择来输出第一修改音频信号或第二修改音频信号。

[0035] 本发明实施例提供一种方法,其中,目标部分是初始音频信号的语音部分,并且侧边部分是音频信号的环境噪声部分。

[0036] 本发明的实施例基于定义不同的语音清晰度选项关于它们的改进效果而变化,这取决于多个影响因素,例如,取决于输入音频流或输入音频场景。在一个音频流中,最佳语音清晰度算法也可以因场景而异。因此,本发明的实施例分析音频信号的不同修改,特别是关于初始音频信号与修改音频信号之间的感知相似度,以便选择具有最高感知相似度的修改器/修改音频信号。该系统/构思首次使整体声音在感知上仅在必要时改变,但要尽可能少地改变,以满足两个要求,即改进初始信号的语音清晰度(或减少收听努力),同时尽可能少地影响声音美学分量。与非自动方法相比,这表示工作量和成本的显著减少,并且相对于迄今为止仅作为边界条件用于改进清晰度的方法,这表示显著的附加值。由于保持该声音美学表示用户接受度的重要组成部分,迄今为止在自动化方法中尚未考虑到这一点。

[0037] 根据实施例,当相应的第一感知相似度值或第二感知相似度值低于阈值时,执行输出初始音频信号而不是输出第一修改音频信号或第二修改音频信号的步骤。“低于”指示修改信号与初始音频信号不够相似。这是有利的,因为该系统能够针对语音清晰度或收听努力自动检查混音,并且同时它确保整体声音以高效方式在感知上改变。

[0038] 本发明的实施例提供了一种方法,其中,比较的步骤包括:通过使用(感知)模型(如PEAQ模型、POLQA模型和/或PEMO-Q模型[8]、[9]、[10])来提取第一感知相似度值和/或第二感知相似度值。注意,PEAQ、POLQA和PEMOQ是被训练以输出两个音频信号的感知相似度的特定模型。根据实施例,处理的程度由另一模型控制。

[0039] 注意,根据实施例,第一感知相似度值和/或第二感知相似度值取决于第一修改音频信号或第二修改音频信号的物理参数、第一修改音频信号或第二修改音频信号的音量电平、第一修改音频信号或第二修改音频信号的心理声学参数、第一修改音频信号或第二修改音频信号的响度信息、第一修改音频信号或第二修改音频信号的音调信息、和/或第一修改音频信号或第二修改音频信号的感知源宽度信息。

[0040] 本发明的实施例提供了一种方法,其中,第一信号修改器和/或第二信号修改器被配置为执行(例如,初始音频信号的)SNR增加、(例如,初始音频信号的)动态压缩;和/或其中,如果初始音频信号包括单独的目标部分和单独的侧边部分,则修改的步骤包括:增加目标部分、增加对目标部分的频率加权、动态压缩目标部分、减少侧边部分、减少对侧边部分的频率加权;备选地,如果初始音频信号包括组合的目标部分和侧边部分,则修改包括:执行对目标部分和侧边部分的分离。通常,这意味着本发明的实施例提供了一种方法,其中,第一修改音频信号和/或第二修改音频信号包括:被移动到前景中的目标部分和被移动到背景中的侧边部分,和/或作为目标部分被移动到前景中的语音部分和作为侧边部分被移动到背景中的环境噪声部分。

[0041] 根据实施例,考虑到一个或多个另外的因素(如听力受损者的听力损失等级、个人听力表现)来执行选择的步骤;个人频率相关听力表现;个人偏好;关于信号修改率的个人偏好。类似地,根据实施例,考虑到一个或多个因素(如听力受损者的听力损失等级、个人听力表现)来执行修改和/或比较的步骤;个人频率相关听力表现;个人偏好;关于信号修改率的个人偏好。因此,选择、修改和/或比较也可以考虑个人听力或个人偏好。

[0042] 根据实施例,用于控制处理的模型可以例如针对听力损失或个人偏好来配置。

[0043] 根据实施例,比较的步骤是针对以下内容执行的:整个初始音频信号以及整个第一修改音频信号和第二修改音频信号,或者单独音频信号的目标部分与第一修改音频信号和第二修改音频信号的相应目标部分进行比较,或初始音频信号的侧边部分与第一修改音频部分和第二修改音频部分的侧边部分进行比较。

[0044] 本发明的实施例提供了一种方法,其中,该方法还包括以下初始步骤:分析初始音频部分以确定语音部分;将语音部分与环境噪声部分进行比较以评估初始音频信号的语音清晰度,并且如果指示语音清晰度的值低于阈值,则激活第一信号修改器和/或第二信号修改器以进行修改的步骤。因此,仅在出现语音的通道处进行处理是有利的。这里,针对该语音部分生成修改混音,其中,混音旨在满足或最大化特定感知度量。

[0045] 本发明的实施例提供了一种方法,其中,初始音频信号包括多个时间帧或场景,其中,针对每个时间帧或场景重复基本步骤。

[0046] 根据实施例,可以使用第一修改器来适配第一时间帧,其中,针对第二时间帧选择另一修改器。为了确保感知连续性,可以插入时间帧或两个时间帧的适配部分之间的过渡。例如,第一时间帧的结束和后续时间帧的开始针对其适配性能进行适配。例如,可以应用两种适配方法之间的一种插值。根据另外的实施例,可以针对所有或多个后续时间帧使用相同的修改器以便实现感知连续性。根据另外的实施例,即使例如从清晰度性能的角度来看不需要适配,也可以执行对时间帧的适配。然而,这能够确保相应时间帧之间的感知相似性。

[0047] 本发明的实施例提供了一种具有程序代码的计算机程序,该程序代码用于当在计算机上运行时执行上述方法。

[0048] 本发明的另一实施例提供了一种用于处理初始音频信号的装置。该装置包括:用于接收初始音频信号的接口;用于处理初始音频信号以获得相应修改音频信号的相应修改器、用于执行对相应修改音频信号的评估的评估器、以及用于取决于相应的第一评估值或第二评估值来选择第一修改音频信号或第二修改音频信号的选择器。

附图说明

[0049] 进一步的细节通过从属权利要求的主题来限定。下面,将参考附图详细讨论本发明的实施例。这里,

[0050] 图1示意性地示出了根据基本实施例的用于处理音频信号以改进目标部分(如音频信号的语音部分)的再现质量的方法序列;

[0051] 图2示出了说明增强实施例的示意性流程图;以及

[0052] 图3示出了根据实施例的用于处理音频信号的解码器的示意性框图。

具体实施方式

[0053] 下面,将参考附图随后讨论本发明的实施例,其中,相同的附图标记被提供给具有相同或相似功能的对象。

[0054] 图1示出了说明包括三个步骤/步骤组110、120和130的方法100的示意性流程图。方法100的目的是能够处理初始音频信号AS并且可以具有输出修改音频信号MOD AS的结果。使用虚拟语气,因为输出音频信号MOD AS的可能结果可以是不需要处理音频信号AS。然后,音频信号和修改音频信号是相同的。

[0055] 三个基本步骤110和120被解释为步骤组,因为这里声纳(sonar)步骤110a、110b等和120a彼此并行或顺序地执行。

[0056] 在步骤组110内,通过使用不同的修改器/处理方法来单独地处理音频信号AS。这里,示出了应用第一修改器和第二修改器的由附图标记110a、110b标记的两个示例性步骤。这两个步骤可以彼此并行或顺序地执行,并且执行对音频信号AS的处理。音频信号例如可以是包括一个音轨的音频信号,其中,该音轨包括两个信号部分。例如,该音轨可以包括语音信号部分(目标部分)和环境噪声信号部分(侧边部分)。这两个部分由附图标记AS_TP和AS_SP标记。在该实施例中,假设AS_TP应该从音频信号AS中提取或者在音频信号AS内标识,以便放大该信号部分AS_TP从而增加语音清晰度。该过程可以针对仅具有包括两个部分AS_SP和AS_TP的一个音轨的音频信号进行,而无需分离包括多个音轨(例如,针对AS_SP一个音轨,并且针对AS_TP一个音轨)的音频AS。

[0057] 如上所述,存在音频信号AS的多个可能的修改,其例如通过放大AS_TP部分或通过减少AS_SP部分都能够改进语音清晰度。其他示例是降低非语音通道、动态范围控制、动态均衡、频谱锐化、频移、语音提取、降噪或现有技术上下文中讨论的其他语音增强动作。这些修改的效率取决于多个因素,例如,取决于记录本身、AS的格式(例如,仅具有一个音轨的格式或具有多个音轨的格式)或取决于多个其他因素。为了实现最佳语音清晰度,至少两个信号修改应用于信号AS。在第一步骤110a内,通过使用第一修改器来修改所接收的初始音频信号AS以获得第一修改音频信号第一MOD AS。独立于步骤110a,通过使用第二修改器来执行对所接收的初始音频信号AS的第二修改以获得第二修改音频信号第二MOD AS。例如,第一修改器可以基于动态范围控制,其中,第二修改器可以基于频谱整形。当然,其他修改器(例如,基于动态均衡、频率重传、语音提取、降噪或语音增强动作、或这种修改器的组合)也可以用于代替第一修改器和/或第二修改器或作为第三修改器(未示出)。所有方法可以导致不同所得的修改音频信号第一MOD AS和第二MOD AS,其可以在语音清晰度方面和与初始音频信号AS的相似度方面不同。这两个参数或这两个参数中的至少一个在下一步骤120

内进行评估。

[0058] 详细地,在步骤120a中,将第一修改音频信号第一MOD AS与原始音频信号AS进行比较,以便找出相似度。类似地,在步骤120b中,将第二修改音频信号第二MOD AS与初始音频信号AS进行比较。为了比较,执行步骤120的实体直接接收音频信号AS和第一MOD AS/第二MOD AS。该比较的结果分别是第一感知相似度值和第二感知相似度值。两个值由附图标记第一PSV和第二PSV标记。两个值描述相应的第一修改音频信号第一MOD AS/第二修改音频信号第二MOD AS与初始音频信号AS之间的感知相似度。在语音清晰度的改进是足够的假设下,选择具有指示更高相似度的第一PSV/第二PSV的第一修改音频信号或第二修改音频信号。这是通过选择130的步骤来执行的。

[0059] 根据实施例,选择的结果可以被输出/转发,使得方法100能够输出与原始信号具有最高相似度的相应修改音频信号第一MOD AS或第二MOD AS。可以看出,修改音频信号MOD AS仍然包括AS_SP'和AS_TP'两个部分。如AS_SP'和AS_TP'内的(')所示,两个部分AS_SP'和AS_TP'中的两者或至少一个被修改。例如,可以增加AS_TP'的放大率。

[0060] 根据另一实施例,可以在步骤120内执行增强的评估。这里,然后进一步证明由第一修改器或第二修改器执行的修改(参见步骤110a和110b)是否足够并改进语音清晰度。例如,可以分析,其中,AS_TP'与AS_SP'的比值大于AS_TP与AS_SP的比值。

[0061] 上述实施例从假设该方法100的目的是具有改进的语音清晰度的MOD AS开始。根据另外的实施例,修改的目的可以不同。例如,部分AS_TP可以是另一部分,通常是应该在整个修改信号MOD AS内被强调的目标部分。这可以通过强调/放大AS_TP'和/或通过修改AS_SP'来进行。

[0062] 此外,已经在感知相似度的上下文中讨论了图1的上述实施例。应当注意,该方法可以更普遍地用于其他评估标准。图1从评估标准是感知相似度的假设开始。然而,根据另外的实施例,也可以附加地代替使用另一评估标准。例如,语音清晰度可以用作评估标准。在这种情况下,代替步骤120a进行对第一修改音频信号第一MOD AS的评估,其中,在步骤120b中执行对第二修改音频信号第二MOD AS的评估。评估120a和120b的这两个步骤的结果是相应的第一评估值和第二评估值。之后,基于相应的评估值来执行步骤130。

[0063] 另一评估标准可以是响度或听觉空间感等。

[0064] 参考图2,下面将讨论具有增强特征的另外的实施例。

[0065] 图2示出了能够处理包括两个部分AS_TP(语音S)和AS_SP(环境噪声N)的音频信号AS的示意性流程图。这里,信号修改器11用于处理信号AS,使得选择实体13可以输出修改信号模式AS。在该实施例中,修改器执行不同的修改1、2、...、M。这些修改基于多个不同的模型,从而生成三个修改信号第一MOD AS、第二MOD AS和M MOD AS。对于每个信号第一MOD AS、第二MOD AS和M MOD AS,示出了两个部分S1'、N1',S2'、N2'和SN'、NNM'。第一MOD AS、第二MOD AS和M MOD AS的输出信号由评估器12针对其与初始信号AS的感知相似度进行评估。因此,一个或多个评估器阶段12接收信号AS和相应的修改信号第一MOD AS、第二MOD AS和M MOD AS。该评估12的输出是相应的修改信号第一MOD AS、第二MOD AS和M MOD AS以及相应的相似度信息。基于该相似度信息,定位阶段13决定要输出的调制信号MOD AS。

[0066] 根据实施例,信号AS可以由分析器21分析以确定语音是否存在。在初始音频信号AS内不存在要修改的语音或信号的情况下,该决定步骤由21s标记。初始/原始音频信号AS

用作信号,即未经修改(参见N-MOD AS)。

[0067] 在存在语音的情况下,第二分析器22分析是否需要改进语音清晰度。该决定点由附图标记22s标记。在不需要修改的情况下,原始信号AS用作要输出的信号(参见N-MOD AS)。在建议修改的情况下,启用信号修改器11。

[0068] 基于该结构,可以改进音频和视听媒体中的语音清晰度。这里,要处理的混音可以是完成的混音,或者可以包括单独的音轨或声音对象(例如,对话、音乐、混响、效果)。在第一步骤中,针对语音的存在来分析信号(参见附图标记21、21s)。例如基于[7]中呈现的混合信号方法,语音活动通道将根据例如以语音清晰度(例如,SII)或收听努力的计算值的形式物理或心理声学参数进行进一步分析(参见附图标记22、22s)。基于该评估,通过将参数与目标或阈值进行比较,决定语音清晰度是否足够或是否需要声音适配。如果不需要适配,则混音照常进行或保持原始混音AS。如果需要适配,将应用修改音轨或不同音轨以便获得期望的清晰度的算法。至此,该方法类似于US 8,195,454 B2和US 8,271,276 B1中公开的方法,但不限于相应权利要求1中所述的细节。

[0069] 根据实施例,这意味着:对超过非语音通道的响度的最大化的声音降低方法的基于模型的选择13(例如,在US 8,577,676 B2和US 2016/0071527 A1中描述的)是用该构思执行的。对于选择,应用了另一模型阶段12,其基于物理和/或心理声学参数模拟原始混音AS与以不同方式(第一MOD AS、第二MOD AS、M MOD AS)修改的混音之间的感知相似度。这里,原始混音AS以及不同类型的修改混音第一MOD AS、第二MOD AS、M MOD AS用作另一模型阶段12的输入。

[0070] 为了获得尽可能最好地保持声音场景的目标,可以选择用于声音适配的方法(参见附图标记13),该方法通过在感知上最不明显的信号修改获得期望的清晰度。

[0071] 根据实施例,可以以工具(instrumental)方式测量感知相似度并可以在本文中使用的可能模型是例如PEAQ[8]、POLQA[9]或Pemo-Q[10]。此外或附加地,另外的物理(例如,电平)或心理声学度量(例如,响度、音调、感知源宽度)可以用于评估感知相似度。

[0072] 音频流通常包括沿时域布置的不同场景。因此,根据实施例,可以在音轨AS中的不同时间发生不同的声音适配,以便具有最小的侵入性感知效果。如果例如语音AS_TP和背景噪声AS_TP已经具有明显不同的频谱,则简单的SNR适配可以是最佳解决方案,因为简单的SNR适配可以尽可能最好地保持背景噪声的真实性。如果另外的说话者叠加目标语音,则其他方法(例如,动态压缩)可能更适合实现优化目标。

[0073] 根据另外的实施例,该基于模型的选择可以在计算中例如以听力图、个体响度函数的形式或以输入个人声音偏好的形式考虑音频材料的未来收听者的可能的听力损伤。因此,不仅针对具有正常听力能力的人而且针对具有特定形式听力损伤(例如,年龄相关的听力损失)的人确保语音清晰度,并且还考虑原始版本和处理版本之间的感知相似度可以各不相同。

[0074] 注意,模型对语音清晰度和感知相似度的分析以及相应的信号处理可以针对整个混音或仅针对混音的部分(单独场景、单独对话)进行,或者可以在沿整个混音的短时间窗口中进行,使得可以针对每个窗口做出是否必须进行声音适配的决定。

[0075] 下面,将示例性地讨论这种过程的示例:

[0076] i. 无声音适配:如果对收听模型的分析表明确保足够高的语音清晰度,则将不进

行进一步的声音适配。备选地,执行以下适配以避免不同场景之间的感知差异。也可以执行无处理与下面选择的处理之间的“插值”。两个模式可以在不同的时间帧/场景上实现感知连续性。

[0077] 对于对话和背景噪声的分离音轨,可以进行以下步骤:

[0078] ii. 适配声音信号:例如通过提高电平、通过频率加权和/或单通道或多通道动态压缩,仅处理语音信号的音轨以改进语音清晰度。

[0079] iii. 适配干扰噪声:例如通过降低电平、通过频率加权和/或单通道或多通道动态压缩,处理不包括语音的一个或若干个音轨以改进语音清晰度。然而,由于声音美学的原因,完全地消除背景噪声将导致改进的语音清晰度的简单情况是不切实际的,因为音乐、效果等的设计也是创意声音设计的重要部分。

[0080] iv. 适配所有音轨:语音信号的音轨和其他音轨中的一个或若干个都通过上述方法进行处理以改进语音清晰度。

[0081] 注意,对于适配,可以使用例如使用神经网络的人工智能。在已经混合的音频信号(即,对话和背景噪声的非分离音轨)中,例如,当预先使用源分离方法时,也可以执行步骤ii至iv,该源分离方法将混音分离为语音和一个或若干个背景噪声。然后,改进语音清晰度可以例如包括:以改进的SNR重新混合分离的信号,或者通过频率加权或单通道或多通道动态压缩来修改语音信号和/或背景噪声或部分背景噪声。这里,将再次选择既如期望地改进语音清晰度又同时尽可能最好地保持原始声音的声音适配。可以在没有用于检测语音活动的任何明确阶段的情况下应用于源分离的方法。

[0082] 注意,根据实施例,可以通过使用人工智能/神经网络来执行对相应处理的选择。例如,如果存在用于选择的多于一个因素(例如,感知值和响度值或描述与个人收听偏好匹配的值),则可以使用该人工智能/神经网络。

[0083] 上面已经讨论了可以执行对场景的适配(即使这不是必需的)以在不同时间帧/场景上保持感知连续性。根据另一变型,可以选择对多个或所有场景的适配。此外,应当注意,在不同场景之间,可以集成不同适配场景或适配场景和非适配场景之间的一种过渡以保持感知连续性。

[0084] 根据实施例,基于感知相似度的评估和优化(参见附图标记12)可以涉及目标语言、背景噪声或语音和背景噪声的混合。可以存在例如针对处理语音信号、处理背景噪声或处理混音与相应原始信号的感知相似度的不同阈值,使得可以不超过对相应信号的信号修改的特定程度。另一边界条件可以是:背景噪声(例如,音乐)可以相对于之前或之后的时间点感知上没有改变太大,因为否则当例如在语音存在的时刻,感知的连续性将受到干扰,音乐将降低太多或在其频率内容中将被改变,或者演员的语音在电影过程期间可以不会改变太多。也可以基于上述模型来检查这种边界条件。

[0085] 这可能具有以下效果:在不过多干扰语音和/或背景噪声的感知相似度的情况下,可能无法获得期望的清晰度改进。这里(可能是可配置的)决定阶段可以决定要获得哪个目标,或者是否以及如何找到折衷。

[0086] 这里,处理可以迭代地进行,即,可以在声音适配之后再次检查收听模型,以验证期望的语音清晰度和与原始语音的感知相似度已经获得。

[0087] 处理可以在音频材料的整个持续时间内或仅在音频材料的持续时间的部分(例

如,场景、对话)内(取决于收听模型的计算)进行。

[0088] 实施例可用于所有音频和视听媒体(电影、广播、播客、一般的音频渲染)。可能的商业应用是例如:

[0089] i. 基于互联网的服务,其中,客户加载他的音频材料,激活自动语音清晰度改进,以及下载经处理信号。基于互联网的服务可以通过对声音适配方法和声音适配程度的客户特定选择来扩展。这种服务已经存在,但没有使用关于语音清晰度的声音适配的收听模型(参见上文2.(V.))。

[0090] ii. 用于声音制作工具的软件解决方案,例如,集成在数字音频工作站(DAW)中以启用对归档或当前制作的混音的校正。

[0091] iii. 测试算法,标识音频材料中的不与期望的语音清晰度相对应的通道,并可能向用户提供建议的声音适配修改以供选择。

[0092] iv. 软件和/或硬件,集成在广播链的收听者端处的终端设备中,例如条形音箱、耳机、电视设备或接收流式音频内容的设备。

[0093] 在图1的上下文中讨论的方法或在图2的上下文中讨论的构思可以通过使用处理器来实现。该处理器由图3示出。

[0094] 图3示出了两个阶段信号修改器11和评估器/选择器12和13中的处理器10。修改器从接口接收音频信号,并基于不同的模型执行修改,以便获得修改音频信号MOD AS。评估器/选择器12从接口接收音频信号,并基于不同的模型执行修改,以便获得修改音频信号MOD AS。评估器/选择器12、13评估相似度,并基于该信息选择具有最高相似度或具有高相似度和改进的语音清晰度(其是足够的)的信号,以输出MOD AS。

[0095] 当然,这两个阶段11、12和13可以由一个处理器来实现。

[0096] 虽然已经在装置的上下文中描述了一些方面,但是将清楚的是,这些方面还表示对应方法的描述,其中,块或设备对应于方法步骤或方法步骤的特征。类似地,在方法步骤上下文中描述的方面也表示对相应块或项或者相应装置的特征的描述。可以由(或使用)硬件设备(诸如,微处理器、可编程计算机或电子电路)来执行一些或全部方法步骤。在一些实施例中,可以由这种装置来执行最重要方法步骤中的某一个或多个方法步骤。

[0097] 新颖的编码音频信号可以存储在数字存储介质上,或者可以在诸如无线传输介质或有线传输介质(例如,互联网)等的传输介质上传输。

[0098] 取决于某些实现要求,可以在硬件中或在软件中实现本发明的实施例。可以使用其上存储有电子可读控制信号的数字存储介质(例如,软盘、DVD、蓝光、CD、ROM、PROM、EPROM、EEPROM或闪存)来执行实现,该电子可读控制信号与可编程计算机系统协作(或者能够与之协作)从而执行相应方法。因此,数字存储介质可以是计算机可读的。

[0099] 根据本发明的一些实施例包括具有电子可读控制信号的数据载体,其能够与可编程计算机系统协作以便执行本文所述的方法之一。

[0100] 通常,本发明的实施例可以实现为具有程序代码的计算机程序产品,程序代码可操作以在计算机程序产品在计算机上运行时执行方法之一。程序代码可以例如存储在机器可读载体上。

[0101] 其他实施例包括存储在机器可读载体上的计算机程序,该计算机程序用于执行本文所述的方法之一。

[0102] 换言之,本发明方法的实施例因此是具有程序代码的计算机程序,该程序代码用于在计算机程序在计算机上运行时执行本文所述的方法之一。

[0103] 因此,本发明方法的另一实施例是其上记录有计算机程序的数据载体(或者数字存储介质或计算机可读介质),该计算机程序用于执行本文所述的方法之一。数据载体、数字存储介质或记录介质通常是有形的和/或非瞬时性的。

[0104] 因此,本发明方法的另一实施例是表示计算机程序的数据流或信号序列,该计算机程序用于执行本文所述的方法之一。数据流或信号序列可以例如被配置为经由数据通信连接(例如,经由互联网)传输。

[0105] 另一实施例包括处理装置,例如,计算机或可编程逻辑器件,所述处理装置被配置为或适于执行本文所述的方法之一。

[0106] 另一实施例包括其上安装有计算机程序的计算机,该计算机程序用于执行本文所述的方法之一。

[0107] 根据本发明的另一实施例包括被配置为向接收机(例如,以电子方式或以光学方式)传送计算机程序的装置或系统,该计算机程序用于执行本文所述的方法之一。接收机可以是例如计算机、移动设备、存储设备等。装置或系统可以例如包括用于向接收机传送计算机程序的文件服务器。

[0108] 在一些实施例中,可编程逻辑器件(例如,现场可编程门阵列)可以用于执行本文所述的方法的功能中的一些或全部。在一些实施例中,现场可编程门阵列可以与微处理器协作以执行本文所述的方法之一。通常,方法优选地由任意硬件装置来执行。

[0109] 上述实施例对于本发明的原理仅是说明性的。应当理解的是,本文所述的布置和细节的修改和变形对于本领域其他技术人员将是显而易见的。因此,旨在仅由所附专利权利要求的范围来限制而不是由借助对本文的实施例的描述和解释所给出的具体细节来限制。

[0110] 参考文献

[0111] [1] Simon, C. and Fassio, G. (2012). Optimierung audiovisueller Medien für Hörgeschädigte. In: Fortschritte der Akustik - DAGA 2012, Darmstadt, March 2012.

[0112] [2] Ephraim, Y. und Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics Speech and Signal Processing, 32(6):1109-1121.

[0113] [3] Kolbæk, M., Yu, D., Tan, Z-H., & Jensen, J. (2017). Multitalker Speech Separation With Utterance-Level Permutation Invariant Training of Deep Recurrent Neural Networks. IEEE Transactions on Audio, Speech and Language Processing, 25(10), 1901-1913. <https://doi.org/10.1109/TASLP.2017.2726762>

[0114] [4] Jouni, P., Torcoli, M., Uhle, C., Herre, J., Disch, S., Fuchs, H. (2019). Source Separation for Enabling Dialogue Enhancement in Object-based Broadcast with MPEG-H. JAES 67, 510-521. <https://doi.org/10.17743/jaes.2019.0032>

[0115] [5] Sauert, B. and Vary, P. (2012). Near end listening enhancement in

the presence of bandpass noises. In: Proc. der ITG-Fachtagung Sprachkommunikation, Braunschweig, September 2012.

[0116] [6] ANSI S3.5 (1997). Methods for calculation of speech intelligibility index.

[0117] [7] Huber, R., Pusch, A., Moritz, N., Rannies, J., Schepker, H., Meyer, B.T. (2018). Objective Assessment of a Speech Enhancement Scheme with an Automatic Speech Recognition-Based System. ITG-Fachbericht 282: Speech Communication, 10. - 12. October 2018 in Oldenburg, 86-90.

[0118] [8] ITU-R Recommendation BS.1387: Method for objective measurements of perceived audio quality (PEAQ)

[0119] [9] ITU-T Recommendation P.863: Perceptual objective listening quality assessment

[0120] [10] Huber, R. und Kollmeier, B. (2006). PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. IEEE Transactions on Audio, Speech, and Language Processing 14(6), 1902-1911

[0121] [11] NetMix player of Fraunhofer IIS, <http://www.iis.fraunhofer.de/de/bf/amm/forschudentw/forschaudiomulti/dialogenhanc.html>

[0122] [12] <https://auphonic.com/>.

100

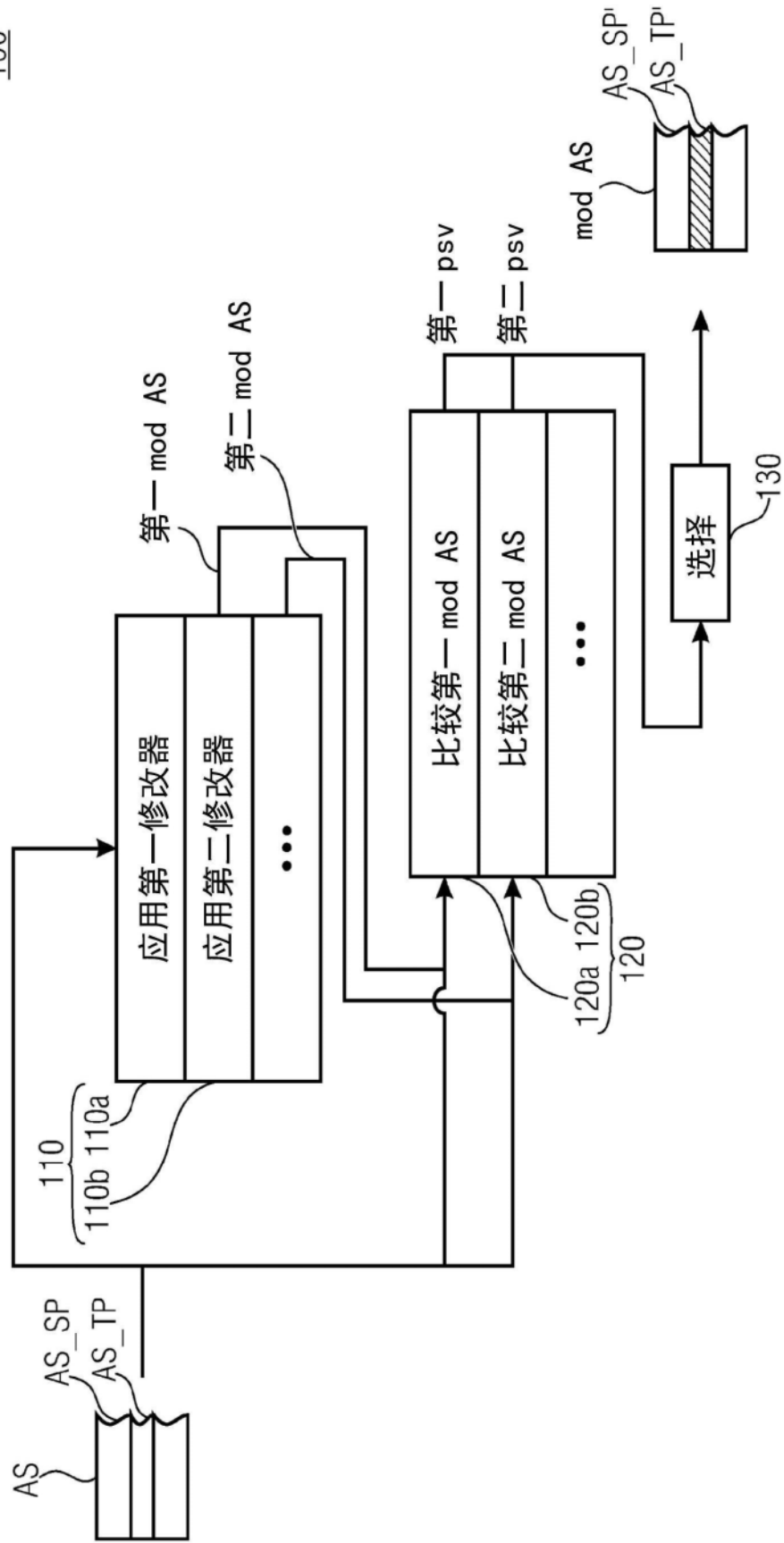


图1

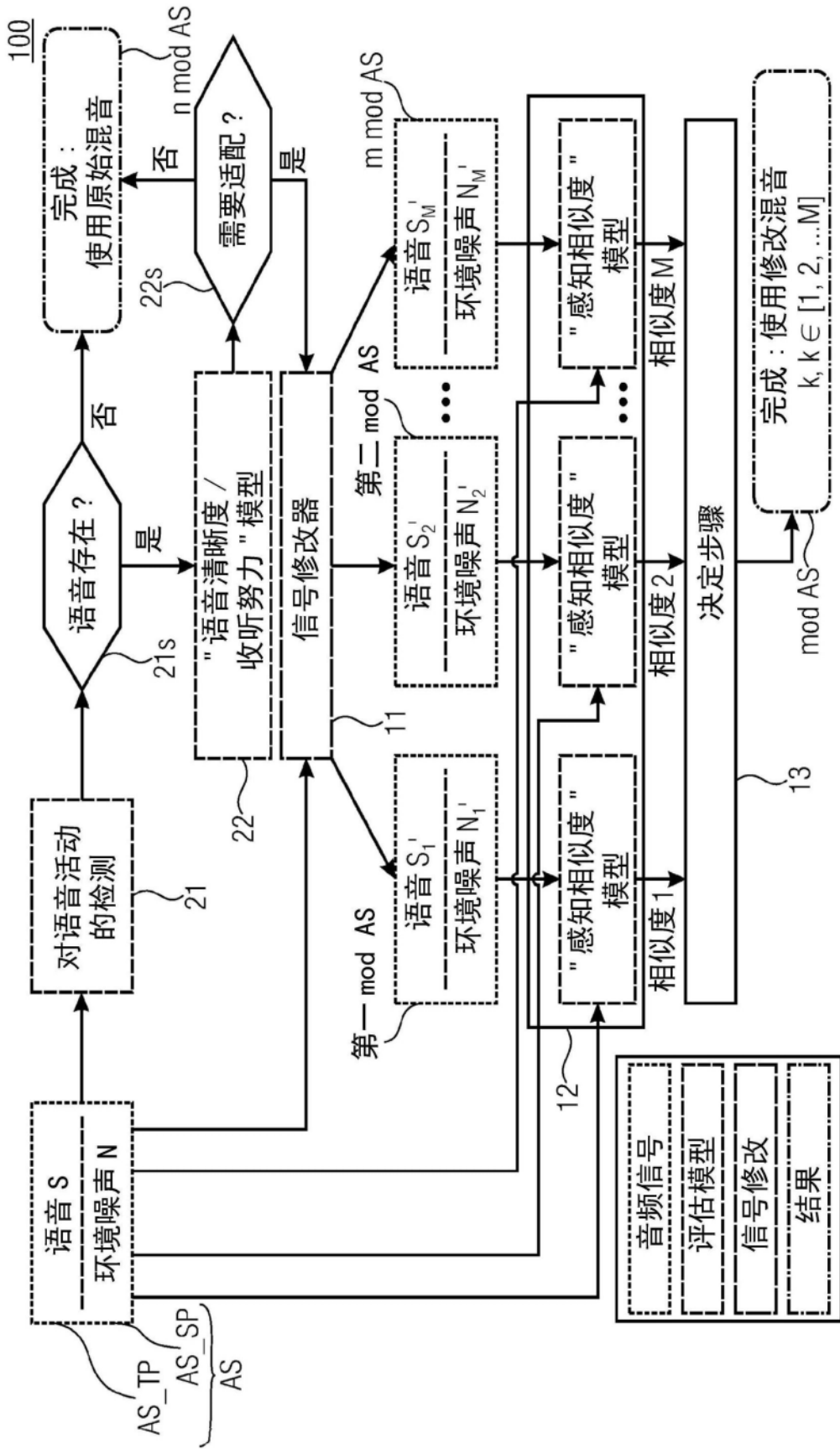


图2

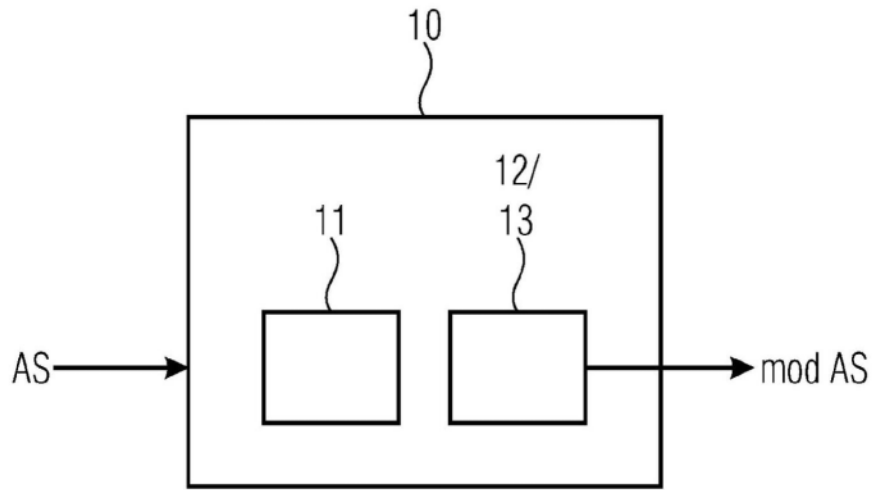


图3