



(12)发明专利

(10)授权公告号 CN 105682000 B

(45)授权公告日 2017. 11. 07

(21)申请号 201610017000.1

(56)对比文件

(22)申请日 2016.01.11

CN 105376690 A, 2016.03.02,

(65)同一申请的已公布的文献号

审查员 李萍

申请公布号 CN 105682000 A

(43)申请公布日 2016.06.15

(73)专利权人 北京时代拓灵科技有限公司

地址 100085 北京市海淀区上地信息路26号01层0014-32室

(72)发明人 张晨 孙学京 刘皓

(74)专利代理机构 北京创遇知识产权代理有限公司

公司 11577

代理人 吕学文 武媛

(51) Int. Cl.

H04S 7/00(2006.01)

G10L 19/008(2013.01)

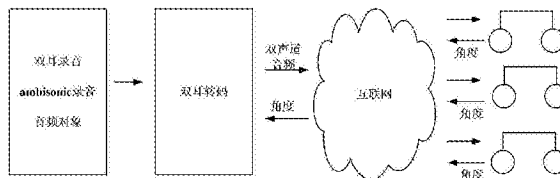
权利要求书2页 说明书8页 附图2页

(54)发明名称

一种音频处理方法和系统

(57)摘要

本发明涉及一种云端音频处理方法,服务器和系统,针对不同格式的音频信号,根据客户端的头部旋转角度,分别对所述不同格式的音频信号进行双耳转码,生成相应格式的双声道音频信号;对所述相应格式的双声道信号叠加,得到音频双耳输出虚拟环绕声信号。本发明的音频处理是在云端服务器上进行的,很好的适应了现有的基于云架构音频处理和存储,从而减少了移动终端生成虚拟环绕声质量不高、运算量大的问题。另外,针对上述在服务器上执行可能带来的延迟,本发明还包括对于角度进行平滑处理,以消除延迟。



1. 一种云端音频处理方法,其特征在于:所述音频处理方法包括以下步骤,
获取用户头部旋转的旋转角度;
获取不同格式的音频信号,根据所述旋转角度,分别对所述不同格式的音频信号进行双耳转码,生成相应格式的双声道音频信号;
对所述相应格式的双声道信号叠加,得到音频双耳输出虚拟环绕声信号;
所获取用户头部旋转的旋转角度具体为获取用户头部旋转的旋转角度,对所述旋转角度进行平滑处理。
2. 根据权利要求1所述的云端音频处理方法,其特征在于:
所述不同格式的音频信号包括双耳录音信号,Ambisonic录音信号和音频对象信号。
3. 根据权利要求2所述的云端音频处理方法,其特征在于:
对所述不同格式的音频信号进行双耳转码,生成相应格式的双耳转码音频信号具体包括:
对所述双耳录音信号,根据所述旋转角度进行插值,生成双耳录音双声道信号;
对所述Ambisonic录音信号,根据所述旋转角度对所述Ambisonic录音信号进行调整,对所述调整后的Ambisonic录音信号双耳转码生成Ambisonic录音双声道信号;
对所述音频对象信号,根据所述旋转角度对所述音频对象信号调整,对所述调整后的音频对象信号双耳转码生成音频对象双声道信号。
4. 根据权利要求3所述的云端音频处理方法,其特征在于:
如需要较高的空间精度,将音频对象信号根据旋转角度进行旋转,将旋转后的音频对象信号编码为高阶B格式音频对象信号,经双耳转码后生成高阶B格式音频对象双声道信号,与Ambisonic录音双声道信号、双耳录音双声道信号进行叠加;
如需要低复杂度低延迟,将音频对象信号编码为一阶B格式音频对象信号,与其他一阶Ambisonic录音信号叠加,然后根据旋转角度对所述叠加后的混合信号进行双耳转码,生成音频对象与Ambisonic录音信号的混合双声道信号,与所述双耳录音双声道信号进行叠加。
5. 一种云端音频处理服务器,其特征在于,所述服务器包括:
获取单元,获取用户头部旋转的旋转角度;
采集单元,采集不同格式的音频信号;
双耳转码单元,分别与所述获取单元和采集单元相连接,根据所述旋转角度,分别对所述不同格式的音频信号进行双耳转码,生成相应格式的双声道音频信号;
叠加单元,与所述双耳转码单元连接,对所述相应格式的双声道信号叠加,得到音频双耳输出虚拟环绕声信号;
所述云端音频处理服务器还包括平滑单元,分别与所述双耳转码单元和所述获取单元连接,平滑单元从获取单元接收用户头部旋转的旋转角度,对所述旋转角度进行平滑处理。
6. 根据权利要求5所述的云端音频处理服务器,其特征在于:
所述不同格式的音频信号包括双耳录音信号,Ambisonic录音信号和音频对象信号。
7. 根据权利要求6所述的云端音频处理服务器,其特征在于:
双耳转码单元对所述不同格式的音频信号进行双耳转码,生成相应格式的双耳转码音频信号具体包括:
对所述双耳录音信号,根据所述旋转角度进行插值,生成双耳录音双声道信号;

对所述Ambisonic录音信号,根据所述旋转角度对所述Ambisonic录音信号进行调整,对所述调整后的Ambisonic录音信号双耳转码生成Ambisonic录音双声道信号;

对所述音频对象信号,根据所述旋转角度对所述音频对象信号调整,对所述调整后的音频对象信号双耳转码生成音频对象双声道信号。

8. 根据权利要求7所述的云端音频处理服务器,其特征在于:

如需要较高的空间精度,双耳转码单元将音频对象信号根据旋转角度进行旋转,将旋转后的音频对象信号编码为高阶B格式音频对象信号,经双耳转码后生成高阶B格式音频对象双声道信号,叠加单元对双耳转码单元生成的高阶B格式音频对象双声道信号,Ambisonic录音双声道信号、双耳录音双声道信号进行叠加;

如需要低复杂度低延迟,双耳转码单元将音频对象信号编码为一阶B格式音频对象信号,与其他一阶Ambisonic录音信号叠加,然后根据旋转角度对所述叠加后的混合信号进行双耳转码,生成音频对象与Ambisonic录音信号的混合双声道信号,叠加单元对双耳转码单元生成的与所述混合双声道信号、双耳录音双声道信号进行叠加。

9. 一种音频播放系统,其特征在于:所述系统包括权利要求5-8任一所述云端音频处理服务器,以及客户端;所述客户端包括头部跟踪装置,所述头部跟踪装置抓取头部旋转角度,通过网络上传至所述云端音频处理服务器,所述云端音频处理器获取不同格式的音频信号,并根据所述旋转角度生成音频双耳输出虚拟环绕声信号后,通过所述网络传输至客户端。

10. 根据权利要求9所述的音频播放系统,其特征在于:所述系统还包括存储服务器,存储不同格式的音频信号,当用户请求下载音频文件下载时,所述云端音频处理服务器从所述存储服务器提取所述音频信号。

一种音频处理方法和系统

技术领域

[0001] 本发明涉及信号处理技术领域,特别涉及一种音频处理的方法,服务器以及系统。

背景技术

[0002] 在利用虚拟现实头戴设备(head-mounted display,HMD)向用户呈现内容时,采用虚拟3D音频技术,通过立体声耳机向用户播放音频内容,一种提高临场感的方法是跟踪用户头部动作(head tracking),对声音进行相应的处理。比如,如果原始声音被用户感知为来自正前方,当用户向左转头90度后,声音应被处理使得用户感知声音来自正右方90度。在这里虚拟现实设备可以有很多种类,比如带头部跟踪的显示设备,或者只是一部带头部跟踪传感器的立体声耳机。

[0003] 实现头部跟踪也有多种方法。比较常见的是使用多种运动传感器。运动传感器套件通常包括加速度计、陀螺仪和磁力传感器。在运动跟踪和绝对方向方面每种传感器都有自己固有的强项和弱点。因此常用做法是采用传感器“融合”(sensor fusion)将来自各传感器的信号组合在一起,产生一个更加精确的运动检测结果。

[0004] 在得到头部旋转角度后,需要对声音进行相应的变化。一种方式是将声音转到Ambisonic域,然后再通过使用旋转矩阵对信号做变换。Ambisonic信号通常是多于两个声道,而常见的媒体播放器只支持立体声两声道,这对直接播放Ambisonic或其他多声道的音频信号带来困难。

[0005] 有鉴于此,在本领域需要一种有效且高质量的虚拟环绕声生成和播放的解决方案。

发明内容

[0006] 为了克服现有技术的上述缺陷,本发明的目的在于提供一种云端音频处理方法,服务器和系统,其能有效且高质量地生成虚拟环绕声,主要用于配合虚拟现实头戴设备进行音频的立体声耳机播放,并且所述虚拟环绕声的生成是在云端服务器上进行的,很好的适应了现有的基于云架构的网络类型,由服务器执行虚拟环绕声的生成和存储,从而解决了现有客户端无法播放各种3603D audio,尤其是适用于虚拟现实应用的音频的问题。

[0007] 为了实现上述目的,本发明提供一种云端音频处理方法,所述音频处理方法包括以下步骤,获取用户头部旋转的旋转角度;获取不同格式的音频信号,根据所述旋转角度,分别对所述不同格式的音频信号进行双耳转码,生成相应格式的双声道音频信号;对所述相应格式的双声道信号叠加,得到音频双耳输出虚拟环绕声信号。

[0008] 优选地,所述不同格式的音频信号包括双耳录音信号,Ambisonic录音信号和音频对象信号。

[0009] 优选地,对所述不同格式的音频信号进行双耳转码,生成相应格式的双耳转码音频信号具体包括:

[0010] 对所述双耳录音信号,根据所述旋转角度进行插值,生成双耳录音双声道信号;

[0011] 对所述Ambisonic录音信号,根据所述旋转角度对所述Ambisonic录音信号进行调整,对所述调整后的Ambisonic录音信号双耳转码生成Ambisonic录音双声道信号;

[0012] 对所述音频对象信号,根据所述旋转角度对所述音频对象信号调整,对所述调整后的音频对象信号双耳转码生成音频对象双声道信号。

[0013] 优选地,如需要较高的空间精度,将音频对象信号根据旋转角度进行旋转,将旋转后的音频对象信号编码为高阶B格式音频对象信号,经双耳转码后生成高阶B格式音频对象双声道信号,与Ambisonic录音双声道信号、双耳录音双声道信号进行叠加;

[0014] 如需要低复杂度低延迟,将音频对象信号编码为一阶B格式音频对象信号,与其他一阶Ambisonic录音信号叠加,然后根据旋转角度对所述叠加后的混合信号进行双耳转码,生成音频对象与Ambisonic录音信号的混合双声道信号,与所述双耳录音双声道信号进行叠加。

[0015] 优选地,所获取用户头部旋转的旋转角度具体为获取用户头部旋转的旋转角度,对所述旋转角度进行平滑处理。

[0016] 本发明还提供了一种云端音频处理服务器,所述服务器包括:获取单元,获取用户头部旋转的旋转角度;采集单元,采集不同格式的音频信号;双耳转码单元,分别与所述获取单元和采集单元相连接,根据所述旋转角度,分别对所述不同格式的音频信号进行双耳转码,生成相应格式的双声道音频信号;叠加单元,与所述双耳转码单元连接,对所述相应格式的双声道信号叠加,得到音频双耳输出虚拟环绕声信号。

[0017] 优选地,所述不同格式的音频信号包括双耳录音信号,Ambisonic录音信号和音频对象信号。

[0018] 优选地,双耳转码单元对所述不同格式的音频信号进行双耳转码,生成相应格式的双耳转码音频信号具体包括:

[0019] 对所述双耳录音信号,根据所述旋转角度进行插值,生成双耳录音双声道信号;

[0020] 对所述Ambisonic录音信号,根据所述旋转角度对所述Ambisonic录音信号进行调整,对所述调整后的Ambisonic录音信号双耳转码生成Ambisonic录音双声道信号;

[0021] 对所述音频对象信号,根据所述旋转角度对所述音频对象信号调整,对所述调整后的音频对象信号双耳转码生成音频对象双声道信号。

[0022] 优选地,如需要较高的空间精度,双耳转码单元将音频对象信号根据旋转角度进行旋转,将旋转后的音频对象信号编码为高阶B格式音频对象信号,经双耳转码后生成高阶B格式音频对象双声道信号,叠加单元对双耳转码单元生成的高阶B格式音频对象双声道信号,Ambisonic录音双声道信号、双耳录音双声道信号进行叠加;

[0023] 如需要低复杂度低延迟,双耳转码单元将音频对象信号编码为一阶B格式音频对象信号,与其他一阶Ambisonic录音信号叠加,然后根据旋转角度对所述叠加后的混合信号进行双耳转码,生成音频对象与Ambisonic录音信号的混合双声道信号,叠加单元对双耳转码单元生成的与所述混合双声道信号、双耳录音双声道信号进行叠加。

[0024] 优选地,所述云端服务器还包括平滑单元,分别与所述双耳转码单元和所述获取单元连接,平滑单元从获取单元接收用户头部旋转的旋转角度,对所述旋转角度进行平滑处理。

[0025] 本发明还提供了一种音频播放系统,所述系统包括云端音频处理服务器,以及客

户端;所述客户端包括头部跟踪装置,所述头部跟踪装置抓取头部旋转角度,通过网络上传至所述云端音频处理服务器,所述云端音频处理器接收所述旋转角度,生成音频双耳输出虚拟环绕声信号后,通过所述网络传输至客户端。

[0026] 根据本发明的云端音频处理方法,服务器和系统,有效且高质量地生成虚拟环绕声,主要用于配合虚拟现实头戴设备进行音频的立体声耳机播放,并且所述虚拟环绕声的生成是在云端服务器上进行的,很好的适应了现有的基于云架构的网络类型,由云端服务器执行音频处理和存储,从而解决了现有客户端无法播放各种3603D audio,尤其是适用于虚拟现实应用的音频的问题。

[0027] 采用本发明的云端音频处理技术,在多人语音通讯中会大大提升临场感,用户可以随意转头来关注某一方向的声音,更加逼近现实中的多人交谈场景。特别在使用流媒体的场景中,通过实时调整空间声,音频的方位,可以提升用户的音频体验。如果辅助虚拟现实视频内容,则会更好的提升用户体验。

附图说明

[0028] 图1是本发明的云端音频处理方法一个实施例的原理框图;

[0029] 图2a-c是本发明的云端音频处理方法另一个实施例的原理框图;

[0030] 图3是本发明的音频处理服务器的一个实施例的结构示意图;

[0031] 图4是本发明的音频处理系统的另一个实施例的结构示意图;

具体实施方式

[0032] 实施例一:如图1所示,一种对音频对象处理包括如下处理步骤:

[0033] 通过头部跟踪装置获取用户头部旋转角度;

[0034] 根据所述旋转角度,将音频对象编码到高阶(优选为2阶或3阶)Ambisonic B-格式信号;

[0035] 将所述Ambisonic B-格式信号转换成虚拟扬声器阵列信号;以一个一阶B-格式信号 $[W_1 \ X_1 \ Y_1 \ Z_1]^T$ 为例,转换成虚拟扬声器阵列信号 $[L_1 \ L_2 \ \dots \ L_N]^T$ 的过程就是进行下列运算:

$$[0036] \quad \begin{bmatrix} L_1 \\ L_2 \\ \dots \\ L_N \end{bmatrix} = \begin{bmatrix} G_{w1} & G_{x1} & G_{y1} & G_{z1} \\ G_{w2} & G_{x2} & G_{y2} & G_{z2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ G_{wN} & G_{xN} & G_{yN} & G_{zN} \end{bmatrix} \begin{bmatrix} W_1 \\ X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = G \begin{bmatrix} W_1 \\ X_1 \\ Y_1 \\ Z_1 \end{bmatrix}。$$

[0037] 其中,N为虚拟扬声器拓扑结构中包括的虚拟扬声器的数目。上式中所用的G矩阵为ambisonic解码矩阵,可以通过求伪逆矩阵来得出。

[0038] 对音频对象的所述虚拟扬声器阵列信号基于双耳房间脉冲响应(BRIR)进行双耳转码(通常是3维,即包含高度信息),得到音频对象的双耳输出虚拟环绕声信号。具体是:从虚拟扬声器信号转到耳机信号对应的二路立体声BRIR矩阵,将该二路立体声矩阵和虚拟扬声器阵列信号进行矩阵乘法,得到虚拟环绕声。

[0039] BRIR矩阵为 $\begin{bmatrix} B_{1L} & B_{2L} & \cdots & B_{NL} \\ B_{1R} & B_{2R} & \cdots & B_{NR} \end{bmatrix}$ ，则虚拟环绕声为

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} B_{1L} & B_{2L} & \cdots & B_{NL} \\ B_{1R} & B_{2R} & \cdots & B_{NR} \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \cdots \\ L_N \end{bmatrix} = \begin{bmatrix} F_{WL} & F_{XL} & F_{YL} & F_{ZL} \\ F_{WR} & F_{XR} & F_{YR} & F_{ZR} \end{bmatrix} \begin{bmatrix} W_1 \\ X_1 \\ Y_1 \\ Z_1 \end{bmatrix}。所述音频信号可以为一个或多个。$$

[0040] 所述双耳房间脉冲响应优选为离线生成，可以采用真实测量或由专门的软件生成，因此不必像现有技术下采用在线生成方式时需要存储大量的BRIR，减少了内存消耗。

[0041] 将音频对象编码到Ambisonic B-格式信号时，水平方向阶数优选大于或等于垂直方向阶数，例如，水平方向编码优选为3阶Ambisonic B-格式信号时，垂直方向编码优选为2阶或1阶Ambisonic B-格式信号，分别用H3V2、H3V1表示。由于人对高度感知低于平面角度的分辨率，因此采用以上适当在某个特定方向上降低阶数的方法，减少了运算量，但又不明显降低用户对声音的感知效果。

[0042] 对声场信号即环境声进行处理包括如下步骤：

[0043] 将环境声转换成环境声的双耳输出虚拟环绕声信号，再将所述音频对象（此时的音频对象主要是指环境声之外的声音内容）和所述环境声各自的双耳输出虚拟环绕声信号对应混音并双耳输出。图1所示为该方法的一个实施例的原理框图。其中，所述将环境声（即图1中的声场信号）转换成环境声的双耳输出虚拟环绕声信号优选包括如下步骤：

[0044] 获取环境声的1阶Ambisonic B-格式信号；

[0045] 根据所述旋转角度，将环境声的所述Ambisonic B-格式信号旋转得到旋转后的Ambisonic B-格式信号；具体来说，是根据所述旋转角度生成旋转矩阵，再根据所述旋转矩阵，对环境声的所述Ambisonic B-格式信号（即待调整信号）进行旋转。所谓旋转，即将旋转矩阵与待调整信号矩阵相乘，旋转不改变音频信号矩阵分量的大小，只改变分量的方向。旋转矩阵的阶数与音频信号矩阵相适应。例如，当待调整信号矩阵为 $[W_2 \ X_2 \ Y_2]^T$ 时，旋转矩阵为

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}；当待调整信号矩阵为 $[W_2 \ X_2 \ Y_2 \ Z_2]^T$ 时，旋转矩阵为 $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}。$$$

[0046] 将环境声的所述旋转后的Ambisonic B-格式信号转换成虚拟扬声器阵列信号；对环境声的所述虚拟扬声器阵列信号基于头相关变换函数(HRTF)进行双耳转码（通常是2维，即不包含高度信息），得到环境声的双耳输出虚拟环绕声信号。HRTF在时间域所对应的名称是HRIR(Head Related Impulse Response)。

[0047] 需要指出的是对于音频对象或环境声都可以根据需要使用BRIR或HRIR进行滤波。由于BRIR通常包含房间模型和一组描述声音方位的HRIR/HRTF组成，所以如果输入信号已带有房间或环境的信息则使用HRIR就可满足需求。

[0048] 所述生成虚拟环绕声的方法在实施运算时优选基于以下假定：虚拟扬声器阵列具有左右对称性，用户在房间的中轴线上，用户对应的所述双耳房间脉冲响应和头相关变换

函数也具有左右对称性。基于该假设,可以利用高阶Ambisonic B-格式对称性优化方法,显著减少运算量,提高运算效率。

[0049] 下面描述了如何将音频对象编码到ambisonic域。

[0050] 将音频对象编码到一阶ambisonic信号:

$$[0051] \quad W = \frac{1}{k} \sum_{i=1}^k s_i \left[\frac{1}{\sqrt{2}} \right];$$

$$[0052] \quad X = \frac{1}{k} \sum_{i=1}^k s_i [\cos\theta_i, \cos\phi_i];$$

$$[0053] \quad Y = \frac{1}{k} \sum_{i=1}^k s_i [\sin\theta_i, \cos\phi_i];$$

$$[0054] \quad Z = \frac{1}{k} \sum_{i=1}^k s_i [\sin\phi_i];$$

[0055] s_i 是第*i*个音频对象, $i=1..k$, k 是音频对象的个数。 θ_i 是平面上的角度(方位角), ϕ_i 是垂直方向上的角度。 W 声道信号表示全方向声波, X 声道信号、 Y 声道信号和 Z 声道信号分别表示沿空间三个互相垂直取向X、Y、Z的声波。

$$[0056] \quad \text{一阶Ambisonic B-格式信号表示为} \begin{bmatrix} W_1 \\ X_1 \\ Y_1 \\ Z_1 \end{bmatrix}。$$

[0057] 同理,将音频对象编码到2阶或3阶Ambisonic B-格式信号优选依照下表定义进行:

[0058]

阶数	信道	定义
0	W	1
1	X	$\cos\theta\cos\phi$
	Y	$\sin\theta\cos\phi$
	Z	$\sin\phi$
2	R	$(3\sin^2\phi - 1)/2$
	S	$(\sqrt{3}/2)\cos\theta\sin(2\phi)$
	T	$(\sqrt{3}/2)\sin\theta\sin(2\phi)$
	U	$(\sqrt{3}/2)\cos(2\theta)\cos^2\phi$
	V	$(\sqrt{3}/2)\sin(2\theta)\cos^2\phi$
3	K	$\sin\phi(5\sin^2\phi - 3)/2$
	L	$(\sqrt{5}/2)\cos\theta\cos\phi(5\sin^2\phi - 1)$
	M	$(\sqrt{5}/2)\sin\theta\cos\phi(5\sin^2\phi - 1)$
	N	$(\sqrt{15}/2)\cos(2\theta)\sin\phi\cos^2\phi$
	O	$(\sqrt{15}/2)\sin(2\theta)\sin\phi\cos^2\phi$
	P	$(\sqrt{3}/2)\cos(3\theta)\cos^3\phi$
	Q	$(\sqrt{3}/2)\sin(3\theta)\cos^3\phi$

[0059] 上表中的三角函数对于方位角 θ 是偶函数的,则相应Ambisonic B-格式信号的相应分量是左右对称的,如果上表中的三角函数对于方位角 θ 是奇函数,则相应Ambisonic B-格式信号的相应分量是左右相反的。以一阶Ambisonic B-格式信号为例,从物理意义和坐标来看,w,x,z不分左右,所以如果听着的位置左右对称,并且假定相应的HRTF系数也近似左右对称,那么w,x,z对应的双耳输出的分量对于输出的左右通道是相同的。而y对于左右正好反向。所以y对应的双耳输出的分量对于左右通道是相反的。对于具有对称性的分量,可以采用快速算法,即运算过程中的对称性优化,可进一步降低运算量。

[0060] 另外,由于服务器对音频文件的处理可能存延迟,采取的解决方案是获取用户头部旋转的旋转角度,对所述旋转角度进行平滑处理。因此小的角度变化可不做新的旋转方向处理,有效解决了服务器处理的时延问题。

[0061] 实施例二:

[0062] 图2a-c描述基于云端多路音频传输用来提升沉浸式体验效果的实施例。需要注意的是本发明涵盖两种应用场景(1)音频实时通讯(会议场景),如图2b所示;(2)音频下载,如图2c所示;

[0063] 针对两种场景,输入有三种形式:单独的音频对象,声场输入(wxy形式),双耳录音信号。

[0064] 如图2b所示,针对音频下载场景:

[0065] 存储服务器存储有双耳录音信号,Ambisonic录音信号(声场信号),和/或音频对象,双耳转码服务器从存储服务器获取上述信号,在双耳转码服务器端将音频对象转成Ambisonic信号,例如,一阶水平方向B格式信号,即wxy,并与其他wxy信号(声场信号)相加。双耳转码服务器根据客户端头部跟踪装置传来的角度通过使用旋转矩阵对wxy信号进行旋转,将wxy信号转成双声道,再与双耳录音双声道信号叠加生成音频下载文件。通常需要压缩来降低传输带宽。然后客户端下载压缩后的双声道音频。这种做法会更加高效,但缺点是如果音频对象只用一阶B格式,空间定位的分辨率会有所下降,但是基于云服务的优选做法如果双耳化过程放置在客户端,则客户端从服务器下载wxy信号,则旋转操作无需经过服务器。

[0066] 如需要较高的空间精度,双耳转码服务器先将音频对象根据旋转角度进行旋转,将旋转后的音频对象信号编码为高阶B格式(例如三阶),与其他B格式信号在双声道域叠加:经双耳转码后生成高阶B格式音频对象双声道信号,与Ambisonic录音双声道信号、双耳录音双声道信号进行叠加生成音频文件。

[0067] 在这里我们要注意的是头部跟踪只是一种形式,不排除其他动作参数,如挥手等。本发明同样适用。

[0068] 如图2c所示,针对音频实时通讯(会议场景):

[0069] 双耳转码服务器直接获取双耳录音麦克风阵列,Ambisonic麦克风阵列,单独声源或音频对象,在双耳转码服务器端将双耳录音麦克风阵列,Ambisonic麦克风阵列,单独声源或音频对象执行上述相似的处理过程。

[0070] 实施例三:

[0071] 如图3所示,一种云端音频处理服务器,获取单元,获取客户端中的头部跟踪装置传送的用户头部旋转的旋转角度;采集单元,分别采集双耳录音信号,Ambisonic录音信号,音频对象;双耳转码单元,分别与所述获取单元和采集单元相连接,根据所述旋转角度,分别对所述不同格式的音频信号进行双耳转码,其中对于双耳录音信号,根据所述旋转角度进行插值,生成双耳录音双声道信号;而如需要较高的空间精度的情况下,双耳转码单元将音频对象信号根据旋转角度进行旋转,将旋转后的音频对象信号编码为高阶B格式音频对象信号,经双耳转码后生成高阶B格式音频对象双声道信号,叠加单元对双耳转码单元生成的高阶B格式音频对象双声道信号,Ambisonic录音双声道信号、双耳录音双声道信号进行叠加;

[0072] 如需要低复杂度低延迟,双耳转码单元将音频对象信号编码为一阶B格式音频对象信号,与其他一阶Ambisonic录音信号叠加,然后根据旋转角度对所述叠加后的混合信号进行双耳转码,生成音频对象与Ambisonic录音信号的混合双声道信号,叠加单元对双耳转码单元生成的与所述混合双声道信号、双耳录音双声道信号进行叠加,得到音频双耳输出虚拟环绕声信号。

[0073] 本实施例利用云端服务器来解决支持头部跟踪的多声道音频传输和播放的问题。

[0074] 实施例四:

[0075] 如图4所示,本发明的一种音频处理系统主要包含客户端,存储服务器,云端音频处理服务器;客户端包括头部跟踪模块,存储服务器端存有多声道音频文件,以特定方式存

放。客户端头部跟踪模块获取用户头部动作如头部旋转角度,将参数经互联网上传到服务器端的一台或多台云端音频处理服务器,对多声道音频文件进行相应处理:云端音频处理服务器从存储服务器提取不同格式的音频信号,并根据接收的旋转角度生成音频双耳输出虚拟环绕声信号,将经过双耳转码后的音频文件通过所述网络传输至客户端。

[0076] 客户端下载上述处理后的音频文件,优选的,以双声道立体声格式播放。

[0077] 以上结合附图详细描述了本发明的优选实施方式,但是,本发明并不限于上述实施方式中的具体细节,在本发明的技术构思范围内,可以对本发明的技术方案进行多种简单变型,这些简单变型均属于本发明的保护范围。

[0078] 另外需要说明的是,在上述具体实施方式中所描述的各个具体技术特征,在不矛盾的情况下,可以通过任何合适的方式进行组合。为了避免不必要的重复,本发明对各种可能的组合方式不再另行说明。

[0079] 此外,本发明的各种不同的实施方式之间也可以进行任意组合,只要其不违背本发明的思想,其同样应当视为本发明所公开的内容。

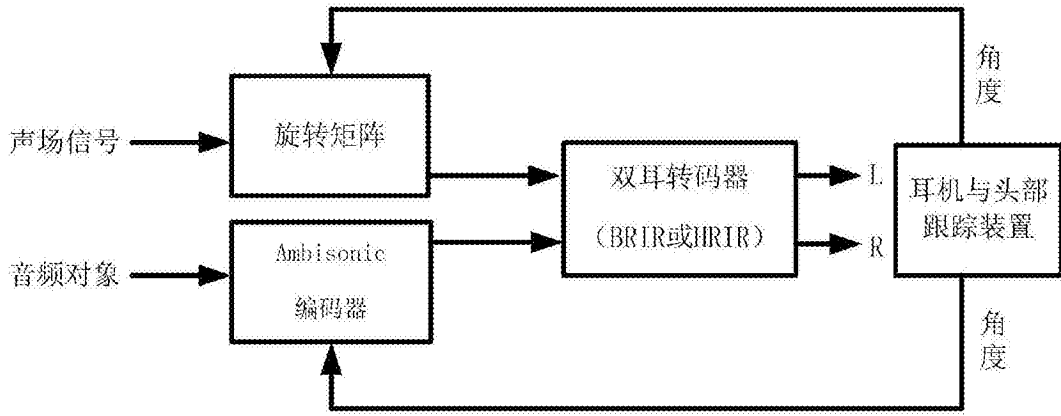


图1

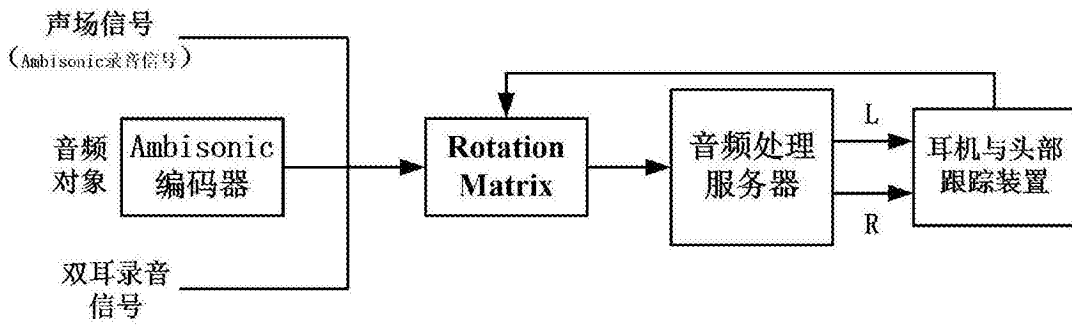


图2a

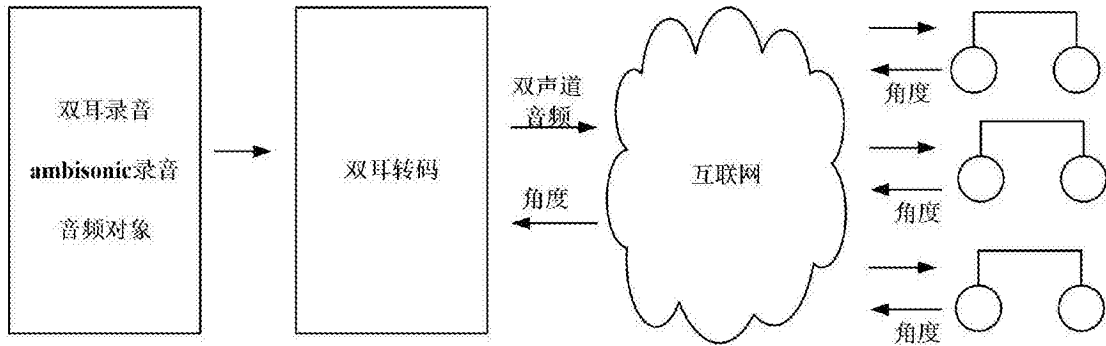


图2b

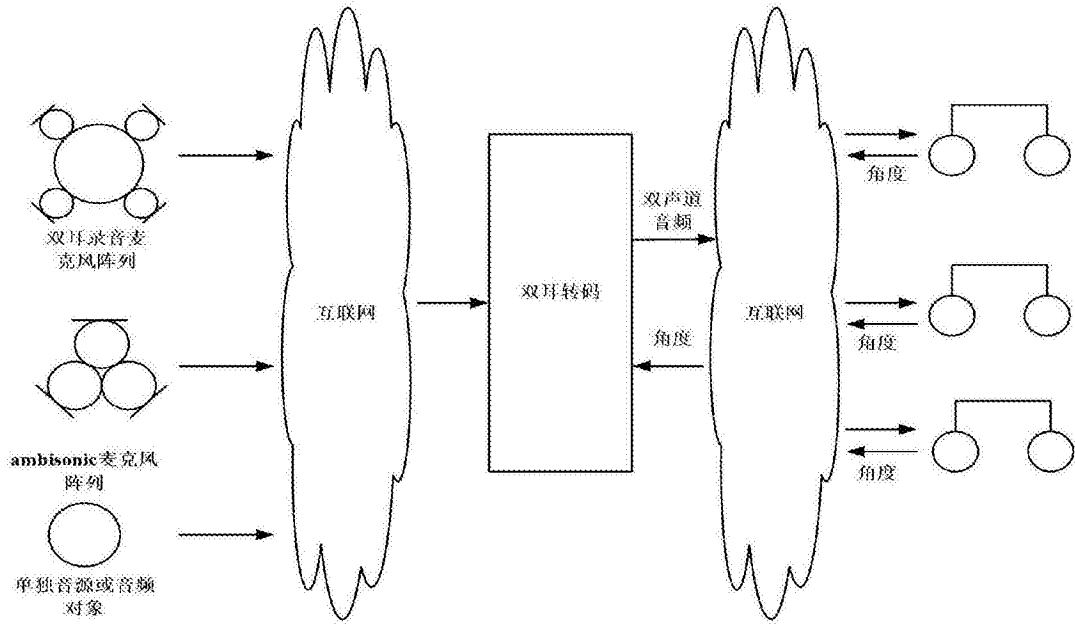


图2c

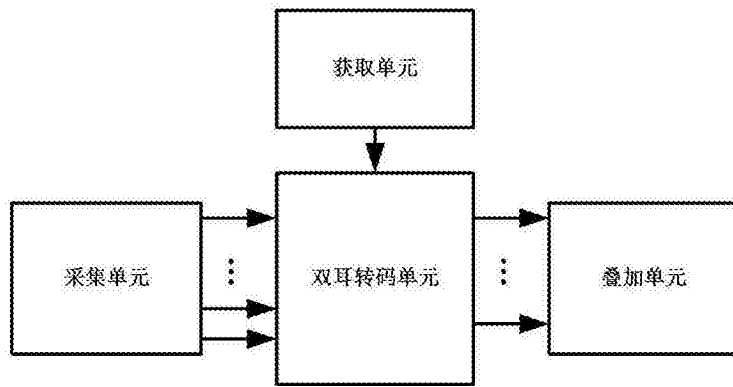


图3

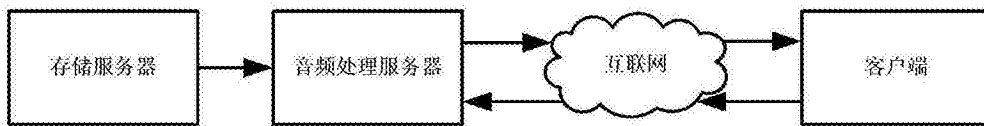


图4