



(21) 申请号 202011603259.7

G06N 3/08 (2023.01)

(22) 申请日 2020.12.30

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 111724807 A, 2020.09.29

申请公布号 CN 112685000 A

李卫东 等. 艺术体操多音轨音频编辑软件的功能及应用. 首都体育学院学报. 2002, 第14卷 (第2期), 第91-92、95页.

(43) 申请公布日 2021.04.20

审查员 徐淑娴

(73) 专利权人 广州酷狗计算机科技有限公司
地址 510660 广东省广州市天河区黄埔大道中315号自编1-17

(72) 发明人 张超钢

(74) 专利代理机构 北京三高永信知识产权代理有限公司 11138

专利代理师 郭翱杰

(51) Int. Cl.

G06F 3/16 (2006.01)

G06N 3/0464 (2023.01)

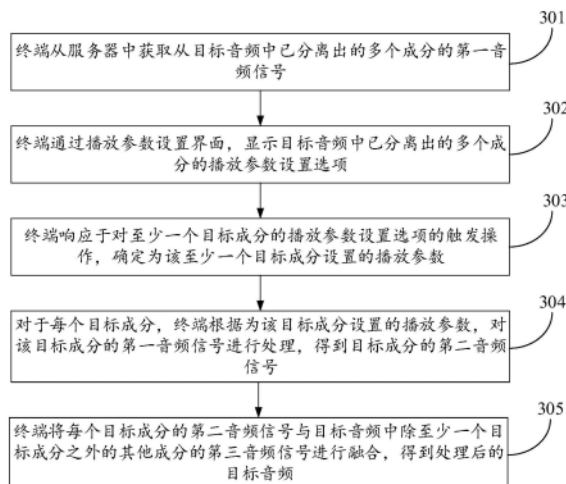
权利要求书3页 说明书19页 附图9页

(54) 发明名称

音频处理方法、装置、计算机设备及存储介质

(57) 摘要

本申请实施例公开了一种音频处理方法、装置、计算机设备及存储介质,属于计算机技术领域。该方法包括:显示目标音频中已分离出的多个成分的播放参数设置选项;响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为至少一个目标成分设置的播放参数;对于每个目标成分,根据为目标成分设置的播放参数,对目标成分的第一音频信号进行处理,得到目标成分的第二音频信号;将每个目标成分的第二音频信号与目标音频中除至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频,实现了单独对音频中的成分的音频信号进行处理,能够为同一音频中的不同成分设置不同的个性化播放效果,提高了音频处理的灵活性。



1. 一种音频处理方法,其特征在于,所述方法包括:

通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,所述成分为人声成分或者任一乐器声成分;

响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为所述至少一个目标成分设置的播放参数,所述目标成分为所述多个成分中的任一成分;

对于每个目标成分,根据为所述目标成分设置的播放参数,对所述目标成分的第一音频信号进行处理,得到所述目标成分的第二音频信号;

将所述每个目标成分的第二音频信号与所述目标音频中除所述至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频;

所述对于每个目标成分,根据为所述目标成分设置的播放参数,对所述目标成分的第一音频信号进行处理,得到所述目标成分的第二音频信号之前,所述方法还包括:

确定所述目标音频对应的第一频谱的第一实部信号和第一虚部信号,所述第一实部信号和所述第一虚部信号中包括音频信号的振幅信息和相位信息;基于所述第一实部信号和所述第一虚部信号,调用频域分离模型,从所述第一频谱的所述第一实部信号和所述第一虚部信号中分离出所述多个成分中每个成分的第二实部信号和第二虚部信号;基于所述每个成分的第二实部信号和第二虚部信号,确定所述每个成分的第六音频信号;将所述频域分离模型分离得到的每个成分的第六音频信号输入至时域分离模型,调用所述时域分离模型,对于所述每个成分,基于所述成分的时域信息,从所述成分的第六音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号;或者,

调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号;确定每个成分的第五音频信号对应的第三频谱的第三实部信号和第三虚部信号;调用所述频域分离模型,对于所述每个成分,从所述成分的第三实部信号和第三虚部信号中分离出所述成分的第四实部信号和第四虚部信号;基于所述每个成分的第四实部信号和第四虚部信号,确定所述每个成分的第一音频信号;

其中,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分。

2. 根据权利要求1所述的方法,其特征在于,所述对于每个目标成分,根据为所述目标成分设置的播放参数,对所述目标成分的第一音频信号进行处理,得到所述目标成分的第二音频信号,包括:

所述播放参数包括音量参数,对于所述每个目标成分,根据为所述目标成分设置的音量参数,调整所述目标成分的第一音频信号的振幅,得到所述目标成分的第二音频信号;或者,

所述播放参数包括音效参数,对于所述每个目标成分,根据为所述目标成分设置的音效参数,对所述目标成分的第一音频信号进行音效处理,得到所述目标成分的第二音频信号;或者,

所述播放参数包括音色参数,所述音色参数指示音频的音色,对于所述每个目标成分,获取所述目标成分对应的曲谱信息,所述曲谱信息用于表示所述目标成分的音高;根据所述曲谱信息和所述音色参数,生成所述目标成分的第二音频信号。

3. 根据权利要求1所述的方法,其特征在于,所述通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项之前,所述方法还包括:

从服务器中获取从所述目标音频中已分离出的多个成分的第一音频信号。

4.一种音频处理方法,其特征在于,所述方法包括:

获取目标音频,所述目标音频由多个成分组成,所述成分为人声成分或者任一乐器声成分;

获取时域分离模型和频域分离模型,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分;

确定所述目标音频对应的第一频谱的第一实部信号和第一虚部信号,所述第一实部信号和所述第一虚部信号中包括音频信号的振幅信息和相位信息;基于所述第一实部信号和所述第一虚部信号,调用所述频域分离模型,从所述第一频谱的所述第一实部信号和所述第一虚部信号中分离出所述多个成分中每个成分的第二实部信号和第二虚部信号;基于所述每个成分的第二实部信号和第二虚部信号,确定所述每个成分的第六音频信号;将所述频域分离模型分离得到的每个成分的第六音频信号输入至所述时域分离模型,调用所述时域分离模型,对于所述每个成分,基于所述成分的时域信息,从所述成分的第六音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号;或者,

调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号;确定每个成分的第五音频信号对应的第三频谱的第三实部信号和第三虚部信号;调用所述频域分离模型,对于所述每个成分,从所述成分的第三实部信号和第三虚部信号中分离出所述成分的第四实部信号和第四虚部信号;基于所述每个成分的第四实部信号和第四虚部信号,确定所述每个成分的第一音频信号。

5.一种音频处理装置,其特征在于,所述装置包括:

显示模块,用于通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,所述成分为人声成分或者任一乐器声成分;

确定模块,用于响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为所述至少一个目标成分设置的播放参数,所述目标成分为所述多个成分中的任一成分;

处理模块,用于对于每个目标成分,根据为所述目标成分设置的播放参数,对所述目标成分的第一音频信号进行处理,得到所述目标成分的第二音频信号;

融合模块,用于将所述每个目标成分的第二音频信号与所述目标音频中除所述至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频;

所述装置还包括:

分离模块,用于确定所述目标音频对应的第一频谱的第一实部信号和第一虚部信号,所述第一实部信号和所述第一虚部信号中包括音频信号的振幅信息和相位信息;基于所述第一实部信号和所述第一虚部信号,调用频域分离模型,从所述第一频谱的所述第一实部信号和所述第一虚部信号中分离出所述多个成分中每个成分的第二实部信号和第二虚部信号;基于所述每个成分的第二实部信号和第二虚部信号,确定所述每个成分的第六音频信号;将所述频域分离模型分离得到的每个成分的第六音频信号输入至时域分离模型,调用所述时域分离模型,对于所述每个成分,基于所述成分的时域信息,从所述成分的第六音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号;或者,

所述分离模块,用于调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号;确定每个成分的第五音频

信号对应的第三频谱的第三实部信号和第三虚部信号;调用所述频域分离模型,对于所述每个成分,从所述成分的第三实部信号和第三虚部信号中分离出所述成分的第四实部信号和第四虚部信号;基于所述每个成分的第四实部信号和第四虚部信号,确定所述每个成分的第一音频信号;

其中,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分。

6.一种音频处理装置,其特征在于,所述装置包括:

音频获取模块,用于获取目标音频,所述目标音频由多个成分组成,所述成分为人声成分或者任一种乐器声成分;

模型获取模块,用于获取时域分离模型和频域分离模型,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分;

分离模块,包括时域分离单元和频域分离单元,所述频域分离单元,用于确定所述目标音频对应的第一频谱的第一实部信号和第一虚部信号,所述第一实部信号和所述第一虚部信号中包括音频信号的振幅信息和相位信息;基于所述第一实部信号和所述第一虚部信号,调用所述频域分离模型,从所述第一频谱的所述第一实部信号和所述第一虚部信号中分离出所述多个成分中每个成分的第二实部信号和第二虚部信号;基于所述每个成分的第二实部信号和第二虚部信号,确定所述每个成分的第六音频信号;

用于执行以下步骤的模块:将所述频域分离模型分离得到的每个成分的第六音频信号输入至所述时域分离模型;

所述时域分离单元,用于调用所述时域分离模型,对于所述每个成分,基于所述成分的时域信息,从所述成分的第六音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号;或者,

所述时域分离单元,用于调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号;确定每个成分的第五音频信号对应的第三频谱的第三实部信号和第三虚部信号;

所述频域分离单元,用于调用所述频域分离模型,对于所述每个成分,从所述成分的第三实部信号和第三虚部信号中分离出所述成分的第四实部信号和第四虚部信号;基于所述每个成分的第四实部信号和第四虚部信号,确定所述每个成分的第一音频信号。

7.一种计算机设备,其特征在于,所述计算机设备包括处理器和存储器,所述存储器中存储有至少一条程序代码,所述至少一条程序代码由所述处理器加载并执行,以实现如权利要求1至3任一权利要求所述的音频处理方法中所执行的操作;或者,以实现如权利要求4所述的音频处理方法中所执行的操作。

8.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有至少一条程序代码,所述至少一条程序代码由处理器加载并执行,以实现如权利要求1至3任一权利要求所述的音频处理方法中所执行的操作;或者,以实现如权利要求4所述的音频处理方法中所执行的操作。

音频处理方法、装置、计算机设备及存储介质

技术领域

[0001] 本申请实施例涉及计算机技术领域,特别涉及一种音频处理方法、装置、计算机设备及存储介质。

背景技术

[0002] 随着计算机技术的不断发展,音频处理类软件得到了广泛的应用,与人们的生活越来越密切。例如,该音频处理类软件具有播放音频的功能,人们可以在休闲时间收听该音频处理类软件播放的音频;又如,该音频处理类软件还具有为音频添加音效的功能,人们可以为音频添加混响、均衡等音效。

[0003] 但是上述对音频进行处理的方式,只能对音频整体进行统一处理,因此音频处理的方式比较单一,音频处理的灵活性较差。

发明内容

[0004] 本申请实施例提供了一种音频处理方法、装置、计算机设备及存储介质,提高了音频处理的灵活性。所述技术方案如下:

[0005] 一方面,提供了一种音频处理方法,所述方法包括:

[0006] 通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,所述成分为人声成分或者任一乐器声成分;

[0007] 响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为所述至少一个目标成分设置的播放参数,所述目标成分为所述多个成分中的任一成分;

[0008] 对于每个目标成分,根据为所述目标成分设置的播放参数,对所述目标成分的第一音频信号进行处理,得到所述目标成分的第二音频信号;

[0009] 将所述每个目标成分的第二音频信号与所述目标音频中除所述至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频。

[0010] 另一方面,提供了一种音频处理方法,所述方法包括:

[0011] 获取目标音频,所述目标音频由多个成分组成,所述成分为人声成分或者任一乐器声成分;

[0012] 获取时域分离模型和频域分离模型,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分;

[0013] 调用所述时域分离模型和所述频域分离模型,从所述目标音频的第四音频信号中分离出至少一个成分中每个成分的第一音频信号。

[0014] 另一方面,提供了一种音频处理装置,所述装置包括:

[0015] 显示模块,用于通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,所述成分为人声成分或者任一乐器声成分;

[0016] 确定模块,用于响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为所述至少一个目标成分设置的播放参数,所述目标成分为所述多个成分中的任一成分

分；

[0017] 处理模块,用于对于每个目标成分,根据为所述目标成分设置的播放参数,对所述目标成分的第一音频信号进行处理,得到所述目标成分的第二音频信号；

[0018] 融合模块,用于将所述每个目标成分的第二音频信号与所述目标音频中除所述至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频。

[0019] 在一种可能实现方式中,所述播放参数包括音量参数,所述处理模块,用于对于所述每个目标成分,根据为所述目标成分设置的音量参数,调整所述目标成分的第一音频信号的振幅,得到所述目标成分的第二音频信号；或者,

[0020] 所述播放参数包括音效参数,所述处理模块,用于对于所述每个目标成分,根据为所述目标成分设置的音效参数,对所述目标成分的第一音频信号进行音效处理,得到所述目标成分的第二音频信号；或者,

[0021] 所述播放参数包括音色参数,所述音色参数指示音频的音色,所述处理模块,用于对于所述每个目标成分,获取所述目标成分对应的曲谱信息,所述曲谱信息用于表示所述目标成分的音高；根据所述曲谱信息和所述音色参数,生成所述目标成分的第二音频信号。

[0022] 在一种可能实现方式中,所述装置还包括：

[0023] 获取模块,用于从服务器中获取从所述目标音频中已分离出的多个成分的第一音频信号。

[0024] 在一种可能实现方式中,所述装置还包括：

[0025] 分离模块,用于调用时域分离模型和频域分离模型,从所述目标音频的第四音频信号中分离出所述多个成分中每个成分的第一音频信号；或者,

[0026] 所述分离模块,用于确定所述目标音频对应的第一频谱的第一实部信号和第一虚部信号；基于所述第一实部信号和所述第一虚部信号,调用所述频域分离模型,从所述第一频谱的所述第一实部信号和所述第一虚部信号中分离出所述多个成分中每个成分的第二实部信号和第二虚部信号；基于所述每个成分的第二实部信号和第二虚部信号,确定所述每个成分的第一音频信号。

[0027] 在一种可能实现方式中,所述分离模块,包括：

[0028] 时域分离单元,用于调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号；

[0029] 频域分离单元,用于调用所述频域分离模型,对于所述每个成分,基于所述成分的频域信息,从所述成分的第五音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号；

[0030] 其中,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分。

[0031] 在一种可能实现方式中,所述分离模块,包括：

[0032] 频域分离单元,用于调用所述频域分离模型,基于所述目标音频的频域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第六音频信号；

[0033] 时域分离单元,用于调用所述时域分离模型,对于所述每个成分,基于所述成分的时域信息,从所述成分的第六音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号；

[0034] 其中,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分。

[0035] 在一种可能实现方式中,所述分离模块,包括:

[0036] 频域分离单元,用于调用所述频域分离模型,基于所述目标音频的频域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第六音频信号;

[0037] 时域分离单元,用于调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号;

[0038] 融合单元,用于对于所述每个成分,将所述成分的第五音频信号与所述成分的第六音频信号进行融合处理,得到所述成分的第一音频信号。

[0039] 在一种可能实现方式中,所述装置还包括:

[0040] 获取模块,用于获取样本数据,所述样本数据包括样本音频以及所述样本音频的多个成分中每个成分的样本音频信号;

[0041] 所述分离模块,用于调用所述频域分离模型,基于所述样本音频的频域信息,从所述样本音频的样本音频信号中分离出所述多个成分中每个成分的第一预测音频信号;

[0042] 所述分离模块,还用于调用所述时域分离模型,基于所述样本音频的时域信息,从所述样本音频的样本音频信号中分离出所述每个成分的第二预测音频信号;

[0043] 所述分离模块,还用于对于所述每个成分,将所述成分的第一预测音频信号与所述成分的第二预测音频信号进行融合处理,得到所述成分的第三预测音频信号;

[0044] 训练模块,用于根据所述每个成分的第三预测音频信号与所述样本数据中对应的样本音频信号之间的差异,对所述频域分离模型和所述时域分离模型进行训练。

[0045] 另一方面,提供了一种音频处理装置,所述装置包括:

[0046] 音频获取模块,用于获取目标音频,所述目标音频由多个成分组成,所述成分为人声成分或者任一乐器声成分;

[0047] 模型获取模块,用于获取时域分离模型和频域分离模型,所述时域分离模型和所述频域分离模型用于从音频中获取相同类型的成分;

[0048] 分离模块,用于调用所述时域分离模型和所述频域分离模型,从所述目标音频的第四音频信号中分离出至少一个成分中每个成分的第一音频信号。

[0049] 在一种可能实现方式中,所述分离模块,包括:

[0050] 时域分离单元,用于调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号;

[0051] 频域分离单元,用于调用所述频域分离模型,对于所述每个成分,基于所述成分的频域信息,从所述成分的第五音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号。

[0052] 在一种可能实现方式中,所述分离模块,包括:

[0053] 频域分离单元,用于调用所述频域分离模型,基于所述目标音频的频域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第六音频信号;

[0054] 时域分离单元,用于调用所述时域分离模型,对于所述每个成分,基于所述成分的时域信息,从所述成分的第六音频信号中分离出所述成分的第一音频信号,得到所述每个成分的第一音频信号。

[0055] 在一种可能实现方式中,所述频域分离单元,用于确定所述目标音频对应的第一频谱的第一实部信号和第一虚部信号;调用所述频域分离模型,从所述第一频谱的第一实部信号和第一虚部信号中分离出所述每个成分的第二实部信号和第二虚部信号;基于所述每个成分的第二实部信号和所述第二虚部信号,确定所述每个成分的第六音频信号。

[0056] 在一种可能实现方式中,所述分离模块,包括:

[0057] 频域分离单元,用于调用所述频域分离模型,基于所述目标音频的频域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第六音频信号;

[0058] 时域分离单元,用于调用所述时域分离模型,基于所述目标音频的时域信息,从所述目标音频的第四音频信号中分离出所述每个成分的第五音频信号;

[0059] 融合单元,用于对于所述每个成分,将所述成分的第五音频信号与所述成分的第六音频信号进行融合处理,得到所述成分的第一音频信号。

[0060] 在一种可能实现方式中,所述装置还包括:

[0061] 样本获取模块,用于获取样本数据,所述样本数据包括样本音频以及所述样本音频的至少一个成分中每个成分的样本音频信号;

[0062] 所述分离模块,用于调用所述频域分离模型,基于所述样本音频的频域信息,从所述样本音频的样本音频信号中分离出所述至少一个成分中每个成分的第一预测音频信号;

[0063] 所述分离模块,用于调用所述时域分离模型,基于所述样本音频的时域信息,从所述样本音频的样本音频信号中分离出所述每个成分的第二预测音频信号;

[0064] 融合模块,用于对于所述每个成分,将所述成分的第一预测音频信号与所述成分的第二预测音频信号进行融合处理,得到所述成分的第三预测音频信号;

[0065] 训练模块,用于根据所述每个成分的第三预测音频信号与所述样本数据中对应的样本音频信号之间的差异,对所述频域分离模型和所述时域分离模型进行训练。

[0066] 另一方面,提供了一种计算机设备,所述计算机设备包括处理器和存储器,所述存储器中存储有至少一条程序代码,所述至少一条程序代码由所述处理器加载并执行,以实现如上述方面所述的音频处理方法中所执行的操作。

[0067] 另一方面,提供了一种计算机可读存储介质,所述计算机可读存储介质中存储有至少一条程序代码,所述至少一条程序代码由处理器加载并执行,以实现如上述方面所述的音频处理方法中所执行的操作。

[0068] 再一方面,提供了一种计算机程序,所述计算机程序中存储有至少一条程序代码,所述至少一条程序代码由处理器加载并执行,以实现上述实施例的音频处理方法中所执行的操作。

[0069] 本申请实施例提供的音频处理方法、装置、设备及介质,在对音频进行处理时,能够通过播放参数设置界面为音频中的一个或者多个成分设置播放参数,对于每个成分,采用为该成分设置的播放参数对该成分的音频信号进行处理,从而实现了单独对音频中的成分的音频信号进行处理,因此,通过上述方法能够为同一音频中的不同成分设置不同的个性化播放效果,提高了音频处理的灵活性。

附图说明

[0070] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使

用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请实施例的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

- [0071] 图1是本申请实施例提供的一种实施环境的结构示意图。
- [0072] 图2是本申请实施例提供的一种音频处理方法的流程图。
- [0073] 图3是本申请实施例提供的一种音频处理方法的流程图。
- [0074] 图4是本申请实施例提供的一种播放参数设置界面的示意图。
- [0075] 图5是本申请实施例提供的一种播放参数设置界面的示意图。
- [0076] 图6是本申请实施例提供的一种播放参数设置界面的示意图。
- [0077] 图7是本申请实施例提供的一种音频处理方法的流程图。
- [0078] 图8是本申请实施例提供的一种音频处理方法的流程图。
- [0079] 图9是本申请实施例提供的一种音频处理装置的结构示意图。
- [0080] 图10是本申请实施例提供的另一种音频处理装置的结构示意图。
- [0081] 图11是本申请实施例提供的一种音频处理装置的结构示意图。
- [0082] 图12是本申请实施例提供的另一种音频处理装置的结构示意图。
- [0083] 图13是本申请实施例提供的一种终端的结构示意图。
- [0084] 图14是本申请实施例提供的一种服务器的结构示意图。

具体实施方式

[0085] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合附图对本申请实施方式作进一步地详细描述。

[0086] 可以理解,本申请所使用的术语“第一”、“第二”、“第三”、“第四”、“第五”、“第六”等可在本文中用于描述各种概念,但除非特别说明,这些概念不受这些术语限制。这些术语仅用于将一个概念与另一个概念区分。举例来说,在不脱离本申请的范围的情况下,可以将第一成分称为第二成分,将第二成分称为第一成分。

[0087] 本申请所使用的术语“每个”、“多个”、“至少一个”、“任一”等,至少一个包括一个、两个或两个以上,多个包括两个或两个以上,而每个是指对应的多个中的每一个,任一是指多个中的任意一个。举例来说,多个成分包括3个成分,而每个是指这3个成分中的每一个成分,任一是指这3个成分中的任意一个,可以是第一个,可以是第二个,也可以是第三个。

[0088] 本申请实施例提供的音频处理方法,应用于计算机设备。在一种可能实现方式中,计算机设备为终端,例如,手机、平板电脑、计算机等。在另一种可能实现方式中,计算机设备为服务器,该服务器为一台服务器,或者由若干服务器组成的服务器集群,或者是一个云计算服务中心。在另一种可能实现方式中,计算机设备包括终端和服务器。

[0089] 图1是本申请实施例提供的一种实施环境的示意图。参见图1,该实施环境包括至少一个终端101和服务器102。终端101和服务器102之间通过无线或者有线网络连接。

[0090] 终端101上安装由服务器102提供服务的目标应用,终端101能够通过该目标应用实现例如数据传输、消息交互等功能。可选地,目标应用为终端101操作系统中的目标应用,或者为第三方提供的目标应用。例如,目标应用为音频处理应用,该音频处理应用具有播放音频的功能,当然,该音频处理应用还能够具有其他功能,例如,录制功能、直播功能、分享

功能、音效调整功能等。可选地,服务器102为该目标应用的后台服务器或者为提供云计算以及云存储等服务的云服务器。

[0091] 可选地,终端101向服务器发送为音频中至少一个目标成分设置的播放参数,由服务器102基于接收的播放参数,对该音频中的至少一个目标成分的音频信号进行个性化处理,得到处理后的音频,将处理后的音频发送至终端101,终端101播放该处理后的音频。

[0092] 可选地,终端101从服务器中获取音频的多个成分的音频信号,对该多个成分中的至少一个目标成分的音频信号进行个性化处理,得到处理后的音频。可选地,终端101从服务器中获取音频,从该音频中分离出多个成分的音频信号,对该多个成分中的至少一个目标成分的音频信号进行个性化处理,得到处理后的音频。需要说明的是,在一种可能实现方式中,终端101得到处理后的音频之后,播放该处理后的音频。

[0093] 其中,音频通常是由人声和乐器声组成,音频中的成分是指组成该音频的人声成分、乐器声成分等。

[0094] 本申请实施例提供的音频处理方法,能够应用于音频处理场景中:

[0095] 例如,应用于音频播放场景中。

[0096] 在播放音频的过程中,采用本申请实施例提供的音频处理方法,用户可以在终端界面中对音频中的成分进行个性化设置,例如,增大人声的音量、为伴奏中的贝斯声添加音效、将伴奏中的鼓声替换为非洲鼓鼓声等。

[0097] 例如,应用于歌曲录制场景中。

[0098] 在用户录制歌曲时,终端播放原歌曲,使得用户根据播放的原歌曲进行演唱,终端录制用户的歌声,将用户的歌声与该原歌曲合成一个新的音频,采用本申请实施例提供的音频处理方法,能够改变原歌曲中的人声音量,通过降低原歌曲中的人声音量,将原歌曲中的人声与用户的歌声混合在一起,达到了将原歌曲中的人声作为和声的效果。

[0099] 需要说明的是,本申请实施例仅是以音频播放场景和歌曲录制场景为例,对音频处理场景进行示例性说明,并不对音频处理场景造成限制,本申请实施例提供的音频处理方法还能够应用于其他任一音频处理场景中。

[0100] 图2是本申请实施例提供的一种音频处理方法的流程图。本申请实施例的执行主体为计算机设备。参见图2,该方法包括:

[0101] 201、通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,该成分为人声成分或者任一乐器声成分。

[0102] 其中,播放参数设置界面是用于设置音频的播放效果的界面,该播放参数设置界面中包括至少一个播放参数设置选项,以供用户调整音频的播放参数,从而改变音频的播放效果。目标音频为计算机设备中的任一音频,例如,任一歌曲等。

[0103] 音频通常是由人声和乐器声组成,音频中的成分是指组成该音频的人声成分、乐器声成分等。例如,音频包括的成分为人声成分和伴奏成分,其中,伴奏成分是指除音频中除人声成分之外的其余多个乐器声成分。又如,音频包括的成分为人声成分、鼓声成分、贝斯声成分、剩余伴奏成分,其中,剩余伴奏成分是指音频中除人声成分、鼓声成分和贝斯声成分之外的其他组成成分。

[0104] 202、响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为至少一个目标成分设置的播放参数。

[0105] 其中,目标成分是目标音频的多个成分中的任一成分。播放参数是用于控制音频播放效果的参数,例如,音量参数、音效参数、音色参数等。目标成分的播放参数即是用于控制该目标成分的播放效果的参数,需要说明的是,在本申请实施例中,目标成分的播放参数仅是用于控制该目标成分的播放效果,而不会控制其他成分的播放效果。

[0106] 203、对于每个目标成分,根据为目标成分设置的播放参数,对目标成分的第一音频信号进行处理,得到目标成分的第二音频信号。

[0107] 其中,根据为目标成分设置的播放参数,对目标成分的第一音频信号进行处理之后,目标成分的播放效果发生改变,从而实现了通过设置目标成分的播放参数,改变目标成分的播放效果。

[0108] 204、将每个目标成分的第二音频信号与目标音频中除至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频。

[0109] 其中,处理后的目标音频中的成分与目标音频中的成分相同,但是处理后的目标音频中目标成分的音频信号与目标音频中目标成分的音频信号并不相同,处理后的目标音频中的目标成分具有与步骤202中设置的播放参数对应的播放效果。

[0110] 本申请实施例提供的音频处理方法,在对音频进行处理时,能够通过播放参数设置界面为音频中的一个或者多个成分设置播放参数,对于每个成分,采用为该成分设置的播放参数对该成分的音频信号进行处理,从而实现了单独对音频中的成分的音频信号进行处理,因此,通过上述方法能够为同一音频中的不同成分设置不同的个性化播放效果,提高了音频处理的灵活性。

[0111] 图3是本申请实施例提供的一种音频处理方法的流程图。在本申请实施例中仅是以执行主体为终端为例进行示例性说明,参见图3,该方法包括:

[0112] 301、终端从服务器中获取从目标音频中已分离出的多个成分的第一音频信号。

[0113] 终端安装有目标应用,服务器是为该目标应用提供服务的服务器。该目标应用为音频处理应用,终端能够从服务器中获取音频,对该音频进行处理,或者进行播放等。

[0114] 在本申请实施例中,服务器中存储有多个音频,以及从每个音频中已分离出的多个成分的第一音频信号,或者,服务器中仅存储有从每个音频中已分离出的多个成分的第一音频信号。因此,终端能够直接从服务器中获取目标音频的多个成分的第一音频信号,而无需对目标音频进行分离处理。

[0115] 在一种可能实现方式中,终端从服务器中获取从目标音频中已分离出的多个成分的第一音频信号,包括:终端向服务器发送音频获取请求,该音频获取请求携带目标音频的音频标识;服务器接收该音频获取请求,基于该目标音频的音频标识,将从目标音频中分离出的多个成分的第一音频信号发送给终端,或者,将目标音频以及从目标音频中分离出的多个成分的第一音频信号发送给终端。

[0116] 其中,音频标识可以为音频的名称、音频的作者、音频的序号等,本申请实施例对音频标识不做限定。

[0117] 例如,用户通过终端的歌曲播放应用播放歌曲A时,终端向服务器发送歌曲获取请求,该歌曲获取请求携带歌曲A的歌曲名,服务器根据该歌曲A的歌曲名,获取歌曲A的人声成分和每个乐器声成分,将人声成分和每个乐器声成分发送给终端。其中,若歌曲A是由人声、钢琴声、架子鼓声和贝斯声组成,则每个乐器声成分是指钢琴声、架子鼓声和贝斯声。

[0118] 需要说明的是,本申请实施例仅是以服务器中存储有从音频中已分离出的多个成分的第一音频信号为例,对终端获取目标音频中已分离出的多个成分的第一音频信号的过程进行示例性说明,而在另一实施例中,服务器中仅存储有多个音频,终端从服务器中获取到音频之后,对获取的音频进行分离处理,从该获取的音频中分离出多个成分的第一音频信号,其中,从音频中分离出多个成分的第一音频信号的过程可以参见图7和图8所示的实施例,在此不再一一赘述。

[0119] 302、终端通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项。

[0120] 其中,播放参数设置界面是用于设置音频的播放效果的界面,该播放参数设置界面中包括至少一个播放参数设置选项。可选地,播放参数设置选项包括音量设置选项、音效设置选项、或者音色设置选项中的至少一种。

[0121] 可选地,播放参数设置界面中显示有每个成分对应的一个播放参数设置选项,例如,如图3所示,播放参数设置界面中包括人声成分的音量设置选项、鼓声成分的音量设置选项、贝斯声成分的音量设置选项和其他伴奏的音量设置选项。通过该播放参数设置界面能够目标音频中多种成分的音量。

[0122] 可选地,播放参数设置界面中显示有每个成分对应的多个播放参数设置选项。例如,如图4所示,播放参数设置界面中显示有人声成分的音量设置选项和音效设置选项、鼓声成分的音量设置选项和音效设置选项、贝斯声成分的音量设置选项和音效设置选项和其他伴奏的音量设置选项和音效设置选项。其中,音效设置选项为一个或者多个选项。例如音效设置选项包括混响选项、舒缓选项、摇滚选项等;或者,该音效设置选项用于触发显示音效设置界面,该音效设置界面包括混响选项、舒缓选项、摇滚选项等多个音效选项。

[0123] 可选地,播放参数设置界面中显示有一个成分对应的一个或者多个播放参数设置选项,也就是说,播放参数设置界面每次仅能显示一种成分的播放参数设置选项。在一种可能实现方式中,播放参数设置界面中包括成分选择选项,该成分选择选项用于指示显示哪种成分的播放参数设置选项,或者,用于指示当前播放参数设置选项指示的播放参数对应哪种成分。例如,如图5所示,播放参数设置界面中包括人声选项、鼓声选项、贝斯声选项、其他伴奏选项以及至少一个播放参数设置选项,当人声选项处于选中状态时,通过对该至少一个播放参数设置选项进行触发操作,为人声成分设置播放参数;当贝斯声选项处于选中状态时,通过对该至少一个播放参数设置选项进行触发操作,为贝斯声成分设置播放参数。

[0124] 需要说明的是,不同音频的成分组成可能不同,因此,从音频中分离出的成分可能也不同。例如,从音频A中分离出的成分为人声成分和伴奏成分;从音频B中分离出的成分为人声成分、鼓声成分和其他伴奏成分,因此,上述终端通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,包括:获取目标音频中的每个成分的成分标识,根据获取的成分标识,通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,这样,能够保证播放参数设置界面显示的成分与目标音频的成分相对应。

[0125] 303、终端响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为该至少一个目标成分设置的播放参数。

[0126] 其中,目标成分为从目标音频中分离出的多个成分中的任一成分,这里,目标成分

是指被修改播放参数的成分。

[0127] 可选地,播放参数设置选项为音量调节选项,那么播放参数为音量参数;可选地,播放参数设置选项为音效调节选项,那么播放参数为音效参数,该音效参数为音效名称、对音频信号的调整参数等;可选地,播放参数设置选项为音色调节选项,那么播放参数为音色参数,该音色参数用于标识将音频的音色调整为目标音色,例如,该音色参数为目标音色的音色标识。

[0128] 可选地,每一种音效都对应至少一个对音频信号的调整参数,例如音频信号的频率调整参数、相位调整参数等。

[0129] 需要说明的是,为每个目标成分设置的播放参数可以是一个,也可以是多个,例如,为目标成分设置的播放参数包括音量参数和音效参数;并且,为每个目标成分设置的播放参数可以相同,也可以不同。

[0130] 304、对于每个目标成分,终端根据为该目标成分设置的播放参数,对该目标成分的第一音频信号进行处理,得到目标成分的第二音频信号。

[0131] 其中,播放参数的类型不同,对目标成分的处理方式也不同,本申请实施例分别与播放参数为音量参数、音效参数和音色参数为例,对目标成分的处理过程进行示例性说明,在一些其他实施例中,播放参数还可以是其他类型的参数,本申请实施例对播放参数的类型不做限定,对根据播放参数对音频进行处理的过程也不做限定。

[0132] 在一种可能实现方式中,播放参数包括音量参数,对于每个目标成分,根据为目标成分设置的播放参数,对目标成分的第一音频信号进行处理,得到目标成分的第二音频信号,包括:对于每个目标成分,根据为目标成分设置的音量参数,调整目标成分的第一音频信号的振幅,得到目标成分的第二音频信号。

[0133] 其中,音频的播放音量是由音频信号的振幅确定的,第一音频信号与第二音频信号之间仅是振幅不同,而频率、相位等信息是相同的,因此,在调整第一音频信号的振幅之后,仅改变了目标成分的音量,而不会改变目标成分的音色、播放速度等。后续,可以基于目标成分的第二音频信号进行播放。例如,音频A包括人声成分、贝斯声成分和鼓声成分,音频A的音量为10,现将人声成分的音量调整为20,而并未对贝斯成分和鼓声成分的音量进行调整,那么终端会调整人声成分的音频信号的振幅,得到人声成分调整后的音频信号,将人声成分调整后的音频信号与贝斯声成分的音频信号以及鼓声成分的音频信号进行融合,得到处理后的目标音频,处理后的目标音频中人声成分的音量为20、贝斯声成分和鼓声成分的音量为10,那么播放时,处理后的目标音频中人声的声音较大。

[0134] 在一种可能实现方式中,播放参数包括音效参数,对于每个目标成分,根据为目标成分设置的播放参数,对目标成分的第一音频信号进行处理,得到目标成分的第二音频信号,包括:对于每个目标成分,根据为目标成分设置的音效参数,对目标成分的第一音频信号进行音效处理,得到目标成分的第二音频信号。

[0135] 其中,音效参数是指示音频的音色。可选地,音效参数包括音量参数、播放速度参数、频率调整参数、相位调整参数等。例如,舒缓音效包括音量减小参数、播放速度减小参数、频率减小参数等,其中,减小参数用于指示减小的振幅。

[0136] 可选地,音效参数包括音量参数、播放速度参数和频率调整参数,对目标成分的第一音频信号进行音效处理,得到目标成分的第二音频信号,包括:对第一音频信号的振幅、

时间和频率进行调整,得到第二音频信号。

[0137] 在一种可能实现方式中,播放参数包括音色参数,音色参数用于表示将音频的音色调整为目标音色,对于每个目标成分,根据为目标成分设置的播放参数,对目标成分的第一音频信号进行处理,得到目标成分的第二音频信号,包括:对于至少一个成分中的每个成分,获取成分对应的曲谱信息,曲谱信息用于表示成分的音高;根据曲谱信息和音色参数,生成具有目标音色的第二音频信号,目标成分由生成的第二音频信号组成。

[0138] 其中,曲谱信息包括至少一个音频,以及每个音频持续的时长。例如,目标成分为架子鼓声成分,音色参数为非洲鼓的标识,对目标音频中架子鼓声成分进行分析,得到架子鼓声成分的曲谱信息,根据该曲谱信息和音色参数,生成非洲鼓声成分,该非洲鼓声成分与该架子鼓声成分的曲谱信息相同,因此,可以将非洲鼓声成分加入目标音频中代替原来的架子鼓声成分。

[0139] 由于音符指示音频的音高,而音高是由音频信号的频率确定的,因此,在一种可能实现方式中,获取成分对应的曲谱信息,包括:根据该成分的第一音频信号的频率,确定该成分对应的音符;根据该成分在频率上的持续时长,确定对应音频的持续时长;根据该成分对应的多个音符以及每个音符的持续时长,生成该成分对应的曲谱信息。

[0140] 例如,成分A的音频信号在第0至1.5秒的频率为B,则第0至1.5秒的音频信号对应的音符为C,且该音符C的持续时长为1.5秒;成分A的音频信号在第1.5至2.5秒的频率为D,则第1.5至2.5秒的音频信号对应的音符为E,且该音符E的持续时长为1秒;成分A的音频信号在第2.5至4秒的频率为F,则第2.5至4秒的音频信号对应的音符为G,且该音符G的持续时长为1.5秒;得到的成分A的曲谱信息中音符依次为C、E和G,且每个音频的持续时长依次为1.5秒、1秒和1.5秒。

[0141] 音色是由音频信号的波形确定,例如,音频信号的谐波幅值、相位偏移等。因此,能够通过生成相应的音频信号,来得到具有目标音色的音频。在一种可能实现方式中,根据曲谱信息和音色参数,生成具有目标音色的第二音频信号,目标成分由生成的第二音频信号组成,包括:将曲谱信息与音色参数输入音频信号合成器,获取音频信号合成器输出的音频信号,音频信号合成器用于根据输入的曲谱信息和指定的音色合成音频信号。

[0142] 可选地,音频信号合成器中存储有多种音色对应的波形特征,将曲谱信息与音色参数输入音频信号合成器,获取音频信号合成器输出的音频信号,包括:将曲谱信息与音色参数输入音频信号合成器,音频信号合成器根据该音色参数确定与该音色参数对应的波形特征,根据该波形特征以及曲谱信息指示的音频信号的频率以及每个频率的持续时长,合成对应的音频信号。

[0143] 在另一种可能实现方式中,终端或者服务器中存储有多种乐器的乐器素材,每个乐器素材为一种乐器对应的音频,可以通过改变乐器素材的音高和节奏,使得该乐器素材变能够替换该目标成分。可选地,根据音色参数,获取音色为目标音色的目标音频素材;按照曲谱信息对目标音频素材进行调整,得到调整后的目标音频素材,将该目标音频素材作为目标成分。

[0144] 其中,按照曲谱信息对目标音频素材进行调整,得到调整后的目标音频素材是指:按照曲谱信息,调整目标音频素材的音高和节奏,使得调整后的目标音频素材对应的曲谱信息与目标成分对应的曲谱信息相同。

[0145] 其中,调整目标音频素材的音高是指:调整目标音频素材的音频信号的频率。调整目标音频素材的节奏是指:调整目标音频素材的音频信号中每个频率的持续时长。

[0146] 305、终端将每个目标成分的第二音频信号与目标音频中除至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频。

[0147] 其中,将多个成分进行融合,与从音频中分离出多个成分为逆过程。通过将每个目标成分的第二音频信号与目标音频中除至少一个目标成分之外的其他成分的第三音频信号进行融合,使得得到的处理后的目标音频中成分完整。

[0148] 可选地,终端将每个目标成分的第二音频信号与目标音频中除至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频,包括:终端将每个目标成分的第二音频信号与目标音频中除至少一个目标成分之外的其他成分的第三音频信号进行叠加,得到第八音频信号,该处理后的目标音频由第八音频信号组成。

[0149] 需要说明的是,在得到处理后的目标音频之后,可以播放该处理后的目标音频,或者将该处理后的目标音频进行存储等,本申请实施例对处理后的目标音频的后续处理方式不做限定。

[0150] 本申请实施例提供的音频处理方法,在对音频进行处理时,能够通过播放参数设置界面为音频中的一个或者多个成分设置播放参数,对于每个成分,采用为该成分设置的播放参数对该成分的音频信号进行处理,从而实现了单独对音频中的成分的音频信号进行处理,因此,通过上述方法能够为同一音频中的不同成分设置不同的个性化播放效果,提高了音频处理的灵活性。

[0151] 并且,终端能够直接从服务器中获取从音频中分离出的多个成分的第一音频信号,无需终端对音频进行分离处理,降低了对终端的要求,也提高了终端处理音频的效率。

[0152] 并且,本申请实施例提供的音频处理方法,提供了多种对音频中成分的处理方式,能够对任一成分的音量、音效和音色进行处理,提高了处理的多样性和灵活性。

[0153] 上述图2至图3所示的实施例仅是对音频中任一个或者多个成分的处理过程进行示例性说明,下面通过图7至图8所示的实施例对从音频中分离出多个成分的过程进行示例性说明。

[0154] 图7是本申请实施例提供的一种音频处理方法的流程图。参见图7,该方法应用于计算机设备中,该方法包括:

[0155] 701、获取目标音频,目标音频由多个成分组成,该成分为人声成分或者任一乐器声成分。

[0156] 其中,目标音频为任一音频,例如该目标音频为任一歌曲的音频、任一交响乐的音频等,本申请实施例对目标音频不做限定。

[0157] 702、获取时域分离模型和频域分离模型,时域分离模型和频域分离模型用于从音频中获取相同类型的成分。

[0158] 其中,时域分离模型是用于基于音频的时域信息,对音频进行分离的模型,例如,该时域分离模型为Wave-U-Net (Wave-U-Network,信号波U型神经网络) 或TasNet (Time-domain audio separation Network,单通道语音分离神经网络) 等模型。频域分离模型是用于基于音频的频域信息,对音频进行分离的模型,例如,该频域分离模型为U-Net (U-Network,U型神经网络) 或open-unmix(一种频域分离模型) 等模型。

[0159] 703、调用时域分离模型和频域分离模型,从目标音频的第四音频信号中分离出至少一个成分中每个成分的第一音频信号。

[0160] 由于时域分离模型和频域分离模型是基于音频中不同的信息进行分离的,因此,该时域分离模型和频域分离模型具有互补性,若调用时域分离模型和频域分离模型,对音频进行分离,则能够更加准确地分离出各种成分。

[0161] 本申请实施例提供的音频处理方法,由于时域分离模型和频域分离模型能够基于音频中不同的信息将音频进行分离,因此,时域分离模型和频域分离模型具有互补性,从而通过时域分离模型和频域分离模型一起对音频进行分离,能够更加准确地分离出各种成分,提高了音频的分离效果。

[0162] 图8是本申请实施例提供的一种音频处理方法的流程图。在本申请实施例中,以计算机设备分离音频为例进行说明,参见图8,该方法包括:

[0163] 801、计算机设备获取目标音频,目标音频由多个成分组成,该成分为人声成分或者任一乐器声成分。

[0164] 802、计算机设备获取时域分离模型和频域分离模型,该时域分离模型和该频域分离模型用于从音频中获取相同类型的成分。

[0165] 其中,在步骤802中获取的时域分离模型和频域分离模型是经过训练的模型,该时域分离模型和频域分离模型具有一定的分离准确性。时域分离模型和频域分离模型用于从音频中获取相同类型的成分是指:从音频中分离出相同成分的音频信号。例如,时域分离模型用于从音频中分离出人声成分的音频信号、鼓声成分的音频信号和其他伴奏的音频信号,频域分离模型也用于从音频中分离出人声成分的音频信号、鼓声成分的音频信号和其他伴奏的音频信号。

[0166] 可选地,时域分离模型用于从音频中分离出目标成分的第一音频信号,而频域分离模型也用于从音频中分离出目标成分的第一音频信号,其中,目标成分为人声成分、或者伴奏成分、或者任一种乐器声成分。

[0167] 可选地,时域分离模型用于从音频中分离出多个成分的音频信号,而频域分离模型也用于从音频中分离出多个成分的音频信号。例如,时域分离模型用于从音频中分离出人声成分和贝斯声成分,频域分离模型也用于从音频中分离出人声成分和贝斯声成分。

[0168] 803、计算机设备调用频域分离模型,基于目标音频的频域信息,从目标音频的第四音频信号中分离出每个成分的第六音频信号。

[0169] 音频的音频信号表示音频的波形随着时间变化的规律,因此,音频信号为音频的时域信息。频谱是音频的频率分布曲线,表示音频的频域信息。通过对音频信号进行时频转换,即可将音频的时域信息转换为频域信息。

[0170] 无论是音频的时域信息还是频域信息,都包含该音频中各种成分的信息,因此,可以基于音频的时域信息或者频域信息,从音频中分离出各个成分的音频信号。在步骤803中,以基于音频的频域信息,从音频中分离出成分的音频信号为例进行说明,在步骤804中,以基于音频的时域信息,从音频中分离出成分的音频信号为例进行说明。

[0171] 在一种可能实现方式中,调用频域分离模型,基于目标音频的频域信息,从目标音频的第四音频信号中分离出每个成分的第六音频信号,包括:基于目标音频对应的第二频谱中的振幅信息,调用频域分离模型,从第二频谱中分离出每个成分对应的振幅信息,基于

每个成分的振幅信息,生成每个成分的第六音频信号。

[0172] 该第二频谱是目标音频的第四音频信号的振幅按照频率排布的曲线,因此,在调用频域分离模型之前,需要先生成第二频谱。可选地,生成第二频谱,包括:将目标音频的第四音频信号进行傅里叶转换,得到复数信号;获取复数信号的实部信息和虚部信息的平方和,将该平方和进行开平方运算,得到该第四音频信号的振幅信息,获取音频信号的振幅信息随着频率变化的曲线,得到第二频谱。

[0173] 由于频域分离模型仅能分离出振幅信息,因此,需要基于目标音频中第四音频信号的相位信息和分离出的振幅信息,来生成成分的第六音频信号。

[0174] 若通过目标音频中第四音频信号的相位信息,来生成成分的第六音频信号,会导致第六音频信号中引入相位噪声,因此,本申请实施例还提供了另一种更加准确地分离方法,在另一种可能实现方式中,调用频域分离模型,基于目标音频的频域信息,从目标音频的第四音频信号中分离出每个成分的第六音频信号,包括:确定目标音频对应的第一频谱的第一实部信号和第一虚部信号;调用频域分离模型,从第一频谱的第一实部信号和第一虚部信号中分离出每个成分的第二实部信号和第二虚部信号;基于每个成分的第二实部信号和第二虚部信号,确定每个成分的第六音频信号。

[0175] 可选地,确定目标音频对应的第一频谱的第一实部信号和第一虚部信号,包括:将目标音频的第四音频信号进行傅里叶转换,得到该第四音频信号对应的第一实部信号和第一虚部信号,获取第一实部信号和第一虚部信号随着频率变化的曲线,得到第一频谱。由于第一频谱即是第一实部信号和第一虚部信号随着频率变化的曲线,因此,得到第一频谱,即是确定了第一频谱中的第一实部信号和第一虚部信号。

[0176] 另外,根据音频信号的振幅信息和相位信息的获取方式可知,第一实部信号和第一虚部信号中包含音频信号的振幅信息和相位信息,因此,能够直接根据每个成分的第二实部信号和第二虚部信号,确定每个成分的第六音频信号,避免了引入相位噪声,得到的第六音频信号更加准确。

[0177] 其中,基于每个成分的第二实部信号和第二虚部信号,确定每个成分的第六音频信号,包括:将每个成分的第二实部信号和第二虚部信号进行逆时频转换,得到每个成分的第六音频信号。

[0178] 804、计算机设备调用时域分离模型,对于每个成分,基于成分的时域信息,从成分的第六音频信号中分离出该成分的第一音频信号,得到每个成分的第一音频信号。

[0179] 在频域分离模型从音频中分离出每个成分的第六音频信号之后,为了保证分离效果,还可以通过时域分离模型,对频域分离模型的分离结果再次进行分离。例如,频域分离模型用于从音频中分离出人声成分,但是分离出的人声成分可能还夹杂一些鼓声成分,因此,将频域分离模型分离出的人声成分输入时域分离模型中,由时域分离模型对该人声成分继续进行分离。

[0180] 对于时域分离模型来说,可以直接将每个成分的第六音频信号输入至该时域分离模型,时域分离模型对于每个成分,从该成分的第六音频信号中分离出该成分的第一音频信号。

[0181] 需要说明的是,本申请实施例仅是以先调用频域分离模型,再调用时域分离模型,对调用时域分离模型和频域分离模型,从目标音频的第四音频信号中分离出至少一个成分

中每个成分的第一音频信号的过程进行示例性说明。在另一实施例中,可以先调用时域分离模型,再调用频域分离模型。其中,调用时域分离模型和频域分离模型,从目标音频的第四音频信号中分离出至少一个成分中每个成分的第一音频信号,包括:调用时域分离模型,基于目标音频的时域信息,从目标音频的第四音频信号中分离出每个成分的第五音频信号;调用频域分离模型,对于每个成分,基于成分的频域信息,从成分的第五音频信号中分离出成分的第一音频信号,得到每个成分的第一音频信号。

[0182] 也即是先调用时域分离模型,再调用频域分离模型。其中,调用频域分离模型,对于每个成分,基于成分的频域信息,从成分的第五音频信号中分离出成分的第一音频信号,得到每个成分的第一音频信号,包括:确定每个成分的第一音频信号对应的第三频谱的第三实部信号和第三虚部信号;调用频域分离模型,对于每个成分,从该成分的第三实部信号和第三虚部信号中分离出该成分的第四实部信号和第四虚部信号;基于每个成分的第三实部信号和第四虚部信号,确定每个成分的第一音频信号。

[0183] 在另一实施例中,还可以并行调用时域分离模型和频域分离模型,调用时域分离模型和频域分离模型,从目标音频的第四音频信号中分离出至少一个成分中每个成分的第一音频信号,包括:调用频域分离模型,基于目标音频的频域信息,从目标音频的第四音频信号中分离出每个成分的第六音频信号;调用时域分离模型,基于目标音频的时域信息,从目标音频的第四音频信号中分离出每个成分的第五音频信号;对于每个成分,将成分的第五音频信号与成分的第六音频信号进行融合处理,得到成分的第一音频信号。

[0184] 也就是说,时域分离模型和频域分离模型采用并联的方式对音频进行分离处理。其中,对于每个成分,将成分的第五音频信号与成分的第六音频信号进行融合处理,得到成分的第一音频信号是指:根据第五音频信号的权值和第六音频信号的权值,对该第五音频信号和第六音频信号进行加权处理,得到第一音频信号。

[0185] 需要说明的是,本申请实施例还提供了一种训练时域分离模型和频域分离模型的方式。在一种可能实现方式中,调用频域分离模型,基于目标音频的频域信息,从目标音频的第四音频信号中分离出每个成分的第六音频信号之前,方法还包括:获取样本数据,样本数据包括样本音频以及样本音频的至少一个成分中每个成分的样本音频信号;调用频域分离模型,基于样本音频的频域信息,从样本音频的样本音频信号中分离出至少一个成分中每个成分的第一预测音频信号;调用时域分离模型,基于样本音频的时域信息,从样本音频的样本音频信号中分离出每个成分的第二预测音频信号;对于每个成分,将成分的第一预测音频信号与成分的第二预测音频信号进行融合处理,得到成分的第三预测音频信号;根据每个成分的第三预测音频信号与样本数据中对应的样本音频信号之间的差异,对频域分离模型和时域分离模型进行训练,以使该第三预测音频信号与样本数据中对应的样本音频信号之间的差异收敛。

[0186] 需要说明的是,本申请实施例提供的时域分离模型和频域分离模型,可以从音频中分离出至少一个成分,本申请实施例还提供了一种当时域分离模型和频域分离模型仅能分离出一个成分时,如果对音频进行分离的方法。

[0187] 先以通过频域分离模型进行音频分离为例进行说明,可选地,频域分离模型为从音频中分离出一个成分的音频信号的模型,调用该频域分离模型从音频中分离出第一成分的第一音频信号,包括:确定目标音频对应的第一频谱的第一实部信号和第一虚部信号,基

于该第一实部信号和第一虚部信号,调用频域分离模型,从第一频谱的第一实部信号和第一虚部信号中分离出第一成分的第二实部信号和第二虚部信号,基于第一成分的第二实部信号和第二虚部信号,确定第一成分的第一音频信号。后续可以基于目标音频的第四音频信号和第一成分的第一音频信号,确定目标音频中剩余成分的第一音频信号,第一成分和剩余成分组成多个成分,从而实现了将音频分离为多个成分。

[0188] 再以通过时域分离模型和频域分离模型共同进行音频分离为例进行说明,可选地,频域分离模型和时域为从音频中分离出一个成分的音频信号的模型,调用该时域分离模型从音频中分离出第一成分的第一音频信号,包括:调用时域分离模型和频域分离模型,从目标音频的第四音频信号中分离出第一成分的第一音频信号。后续基于目标音频的第四音频信号和第一成分的第一音频信号,确定目标音频中剩余成分的第一音频信号,第一成分和剩余成分组成多个成分,从而实现了将音频分离为多个成分。

[0189] 本申请实施例提供的音频处理方法,由于时域分离模型和频域分离模型能够基于音频中不同的信息将音频进行分离,因此,时域分离模型和频域分离模型具有互补性,从而通过时域分离模型和频域分离模型一起对音频进行分离,能够更加准确地分离出各种成分,提高了音频的分离效果。

[0190] 图9是本申请提供的一种音频处理装置的结构示意图。参见图9,该装置包括:

[0191] 显示模块901,用于通过播放参数设置界面,显示目标音频中已分离出的多个成分的播放参数设置选项,该成分为人声成分或者任一乐器声成分;

[0192] 确定模块902,用于响应于对至少一个目标成分的播放参数设置选项的触发操作,确定为该至少一个目标成分设置的播放参数,该目标成分为多个成分中的任一成分;

[0193] 处理模块903,用于对于每个目标成分,根据为该目标成分设置的播放参数,对该目标成分的第一音频信号进行处理,得到该目标成分的第二音频信号;

[0194] 融合模块904,用于将该每个目标成分的第二音频信号与该目标音频中除该至少一个目标成分之外的其他成分的第三音频信号进行融合,得到处理后的目标音频。

[0195] 如图10所示,在一种可能实现方式中,该播放参数包括音量参数,该处理模块903,用于对于该每个目标成分,根据为该目标成分设置的音量参数,调整该目标成分的第一音频信号的振幅,得到该目标成分的第二音频信号;或者,

[0196] 该播放参数包括音效参数,该处理模块903,用于对于该每个目标成分,根据为该目标成分设置的音效参数,对该目标成分的第一音频信号进行音效处理,得到该目标成分的第二音频信号;或者,

[0197] 该播放参数包括音色参数,该音色参数指示音频的音色,该处理模块903,用于对于该每个目标成分,获取该目标成分对应的曲谱信息,该曲谱信息用于表示该目标成分的音高;根据该曲谱信息和该音色参数,生成该目标成分的第二音频信号。

[0198] 在一种可能实现方式中,该装置还包括:

[0199] 获取模块905,用于从服务器中获取从该目标音频中已分离出的多个成分的第一音频信号。

[0200] 在一种可能实现方式中,该装置还包括:

[0201] 分离模块906,用于调用时域分离模型和频域分离模型,从该目标音频的第四音频信号中分离出该多个成分中每个成分的第一音频信号;或者,

[0202] 该分离模块906,用于确定该目标音频对应的第一频谱的第一实部信号和第一虚部信号;基于该第一实部信号和该第一虚部信号,调用该频域分离模型,从该第一频谱的该第一实部信号和该第一虚部信号中分离出该多个成分中每个成分的第二实部信号和第二虚部信号;基于该每个成分的第二实部信号和第二虚部信号,确定该每个成分的第一音频信号。

[0203] 在一种可能实现方式中,该分离模块906,包括:

[0204] 时域分离单元9061,用于调用该时域分离模型,基于该目标音频的时域信息,从该目标音频的第四音频信号中分离出该每个成分的第五音频信号;

[0205] 频域分离单元9062,用于调用该频域分离模型,对于该每个成分,基于该成分的频域信息,从该成分的第五音频信号中分离出该成分的第一音频信号,得到该每个成分的第一音频信号;

[0206] 其中,该时域分离模型和该频域分离模型用于从音频中获取相同类型的成分。

[0207] 在一种可能实现方式中,该分离模块906,包括:

[0208] 频域分离单元9062,用于调用该频域分离模型,基于该目标音频的频域信息,从该目标音频的第四音频信号中分离出该每个成分的第六音频信号;

[0209] 时域分离单元9061,用于调用该时域分离模型,对于该每个成分,基于该成分的时域信息,从该成分的第六音频信号中分离出该成分的第一音频信号,得到该每个成分的第一音频信号;

[0210] 其中,该时域分离模型和该频域分离模型用于从音频中获取相同类型的成分。

[0211] 在一种可能实现方式中,该分离模块906,包括:

[0212] 频域分离单元9062,用于调用该频域分离模型,基于该目标音频的频域信息,从该目标音频的第四音频信号中分离出该每个成分的第六音频信号;

[0213] 时域分离单元9061,用于调用该时域分离模型,基于该目标音频的时域信息,从该目标音频的第四音频信号中分离出该每个成分的第五音频信号;

[0214] 融合单元9063,用于对于该每个成分,将该成分的第五音频信号与该成分的第六音频信号进行融合处理,得到该成分的第一音频信号。

[0215] 在一种可能实现方式中,该装置还包括:

[0216] 获取模块905,用于获取样本数据,该样本数据包括样本音频以及该样本音频的多个成分中每个成分的样本音频信号;

[0217] 该分离模块906,用于调用该频域分离模型,基于该样本音频的频域信息,从该样本音频的样本音频信号中分离出该多个成分中每个成分的第一预测音频信号;

[0218] 该分离模块906,还用于调用该时域分离模型,基于该样本音频的时域信息,从该样本音频的样本音频信号中分离出该每个成分的第二预测音频信号;

[0219] 该分离模块906,还用于对于该每个成分,将该成分的第一预测音频信号与该成分的第二预测音频信号进行融合处理,得到该成分的第三预测音频信号;

[0220] 训练模块907,用于根据该每个成分的第三预测音频信号与该样本数据中对应的样本音频信号之间的差异,对该频域分离模型和该时域分离模型进行训练。

[0221] 图11是本申请实施例提供的一种音频处理装置的结构示意图,参见图11,该装置包括:

[0222] 音频获取模块1101,用于获取目标音频,该目标音频由多个成分组成,该成分为人声成分或者任一乐器声成分;

[0223] 模型获取模块1102,用于获取时域分离模型和频域分离模型,该时域分离模型和该频域分离模型用于从音频中获取相同类型的成分;

[0224] 分离模块1103,用于调用该时域分离模型和该频域分离模型,从该目标音频的第四音频信号中分离出至少一个成分中每个成分的第一音频信号。

[0225] 如图12所示,在一种可能实现方式中,该分离模块1103,包括:

[0226] 时域分离单元1113,用于调用该时域分离模型,基于该目标音频的时域信息,从该目标音频的第四音频信号中分离出该每个成分的第五音频信号;

[0227] 频域分离单元1123,用于调用该频域分离模型,对于该每个成分,基于该成分的频域信息,从该成分的第五音频信号中分离出该成分的第一音频信号,得到该每个成分的第一音频信号。

[0228] 在一种可能实现方式中,该分离模块1103,包括:

[0229] 频域分离单元1123,用于调用该频域分离模型,基于该目标音频的频域信息,从该目标音频的第四音频信号中分离出该每个成分的第六音频信号;

[0230] 时域分离单元1113,用于调用该时域分离模型,对于该每个成分,基于该成分的时域信息,从该成分的第六音频信号中分离出该成分的第一音频信号,得到该每个成分的第一音频信号。

[0231] 在一种可能实现方式中,该频域分离单元1123,用于确定该目标音频对应的第一频谱的第一实部信号和第一虚部信号;调用该频域分离模型,从该第一频谱的第一实部信号和第一虚部信号中分离出该每个成分的第二实部信号和第二虚部信号;基于该每个成分的第二实部信号和该第二虚部信号,确定该每个成分的第六音频信号。

[0232] 在一种可能实现方式中,该分离模块1103,包括:

[0233] 频域分离单元1123,用于调用该频域分离模型,基于该目标音频的频域信息,从该目标音频的第四音频信号中分离出该每个成分的第六音频信号;

[0234] 时域分离单元1113,用于调用该时域分离模型,基于该目标音频的时域信息,从该目标音频的第四音频信号中分离出该每个成分的第五音频信号;

[0235] 融合单元1133,用于对于该每个成分,将该成分的第五音频信号与该成分的第六音频信号进行融合处理,得到该成分的第一音频信号。

[0236] 在一种可能实现方式中,该装置还包括:

[0237] 样本获取模块1104,用于获取样本数据,该样本数据包括样本音频以及该样本音频的至少一个成分中每个成分的样本音频信号;

[0238] 该分离模块1103,用于调用该频域分离模型,基于该样本音频的频域信息,从该样本音频的样本音频信号中分离出该至少一个成分中每个成分的第一预测音频信号;

[0239] 该分离模块1103,用于调用该时域分离模型,基于该样本音频的时域信息,从该样本音频的样本音频信号中分离出该每个成分的第二预测音频信号;

[0240] 融合模块1105,用于对于该每个成分,将该成分的第一预测音频信号与该成分的第二预测音频信号进行融合处理,得到该成分的第三预测音频信号;

[0241] 训练模块1106,用于根据该每个成分的第三预测音频信号与该样本数据中对应的

样本音频信号之间的差异,对该频域分离模型和该时域分离模型进行训练。

[0242] 本申请实施例还提供了一种计算机设备,该计算机设备包括处理器和存储器,存储器中存储有至少一条程序代码,该至少一条程序代码由处理器加载并执行,以实现如上述实施例的音频处理方法中所执行的操作。

[0243] 可选地,计算机设备被提供为终端。图13是本申请实施例提供的一种终端的结构示意图。该终端1300可以是便携式移动终端,比如:智能手机、平板电脑、MP3播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、笔记本电脑或台式电脑。终端1300还可能被称为用户设备、便携式终端、膝上型终端、台式终端等其他名称。

[0244] 终端1300包括有:处理器1301和存储器1302。

[0245] 处理器1301可以包括一个或多个处理核心,比如4核心处理器、8核心处理器等。处理器1301可以采用DSP(Digital Signal Processing,数字信号处理)、FPGA(Field-Programmable Gate Array,现场可编程门阵列)、PLA(Programmable Logic Array,可编程逻辑阵列)中的至少一种硬件形式来实现。处理器1301也可以包括主处理器和协处理器,主处理器是用于对在唤醒状态下的数据进行处理的处理器,也称CPU(Central Processing Unit,中央处理器);协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中,处理器1301可以集成有GPU(Graphics Processing Unit,图像处理器),GPU用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中,处理器1301还可以包括AI(Artificial Intelligence,人工智能)处理器,该AI处理器用于处理有关机器学习的计算操作。

[0246] 存储器1302可以包括一个或多个计算机可读存储介质,该计算机可读存储介质可以是非暂态的。存储器1302还可包括高速随机存取存储器,以及非易失性存储器,比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中,存储器1302中的非暂态的计算机可读存储介质用于存储至少一个程序代码,该至少一个程序代码用于被处理器1301所执行以实现本申请中方法实施例提供的音频处理方法。

[0247] 在一些实施例中,终端1300还可选包括有:外围设备接口1303和至少一个外围设备。处理器1301、存储器1302和外围设备接口1303之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口1303相连。具体地,外围设备包括:射频电路1304、显示屏1305、摄像头组件1306、音频电路1307、定位组件1308和电源1309中的至少一种。

[0248] 本领域技术人员可以理解,图13中示出的结构并不构成对终端1300的限定,可以包括比图示更多或更少的组件,或者组合某些组件,或者采用不同的组件布置。

[0249] 可选地,计算机设备被提供为服务器。图14是根据一示例性实施例示出的一种服务器的结构示意图,该服务器1400可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上处理器(Central Processing Units,CPU)1401和一个或一个以上的存储器1402,其中,存储器1402中存储有至少一条程序代码,至少一条程序代码由处理器1401加载并执行以实现上述各个方法实施例提供的方法。当然,该服务器还可以具有有线或无线网络接口、键盘以及输入输出接口等部件,以便进行输入输出,该服务器还可以包括其他用于

实现设备功能的部件,在此不做赘述。

[0250] 本申请实施例还提供了一种计算机可读存储介质,该计算机可读存储介质中存储有至少一条程序代码,该至少一条程序代码由处理器加载并执行,以实现上述实施例的音频处理方法中所执行的操作。

[0251] 本申请实施例还提供了一种计算机程序,该计算机程序中存储有至少一条程序代码,该至少一条程序代码由处理器加载并执行,以实现上述实施例的音频处理方法中所执行的操作。

[0252] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,该程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0253] 以上仅为本申请实施例的可选实施例,并不用以限制本申请实施例,凡在本申请实施例的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

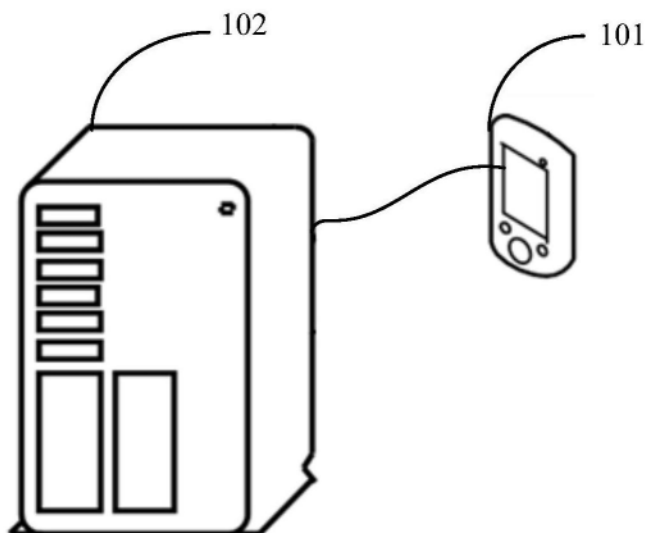


图1

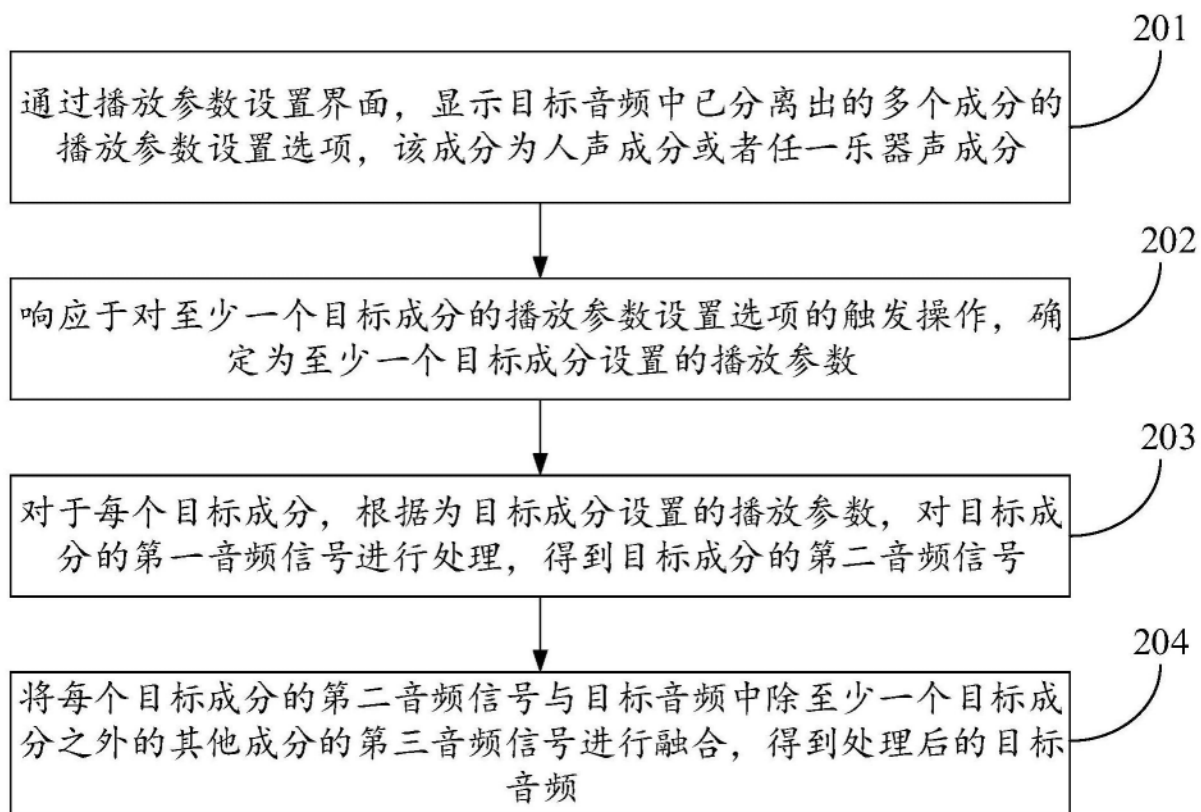


图2

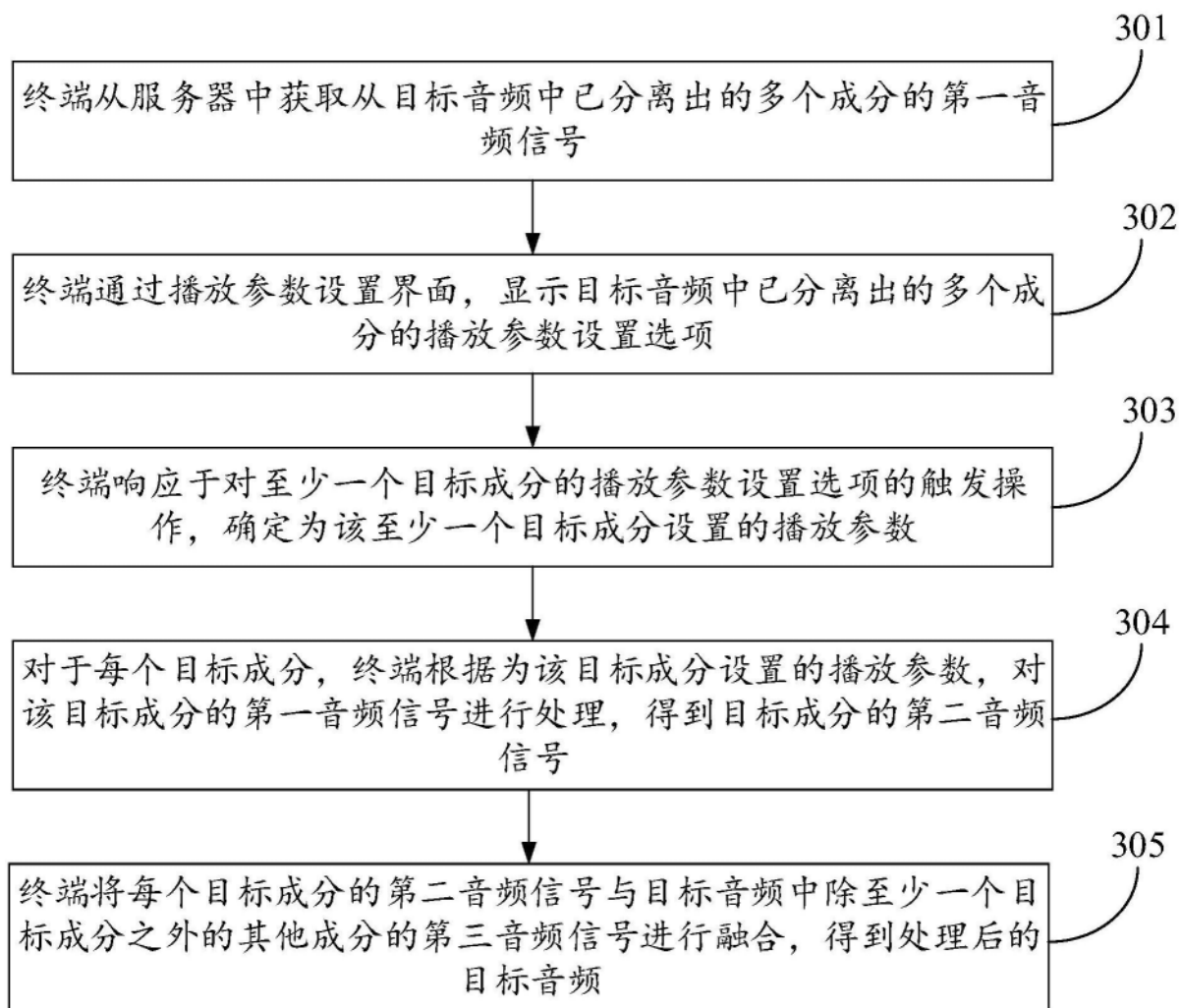


图3

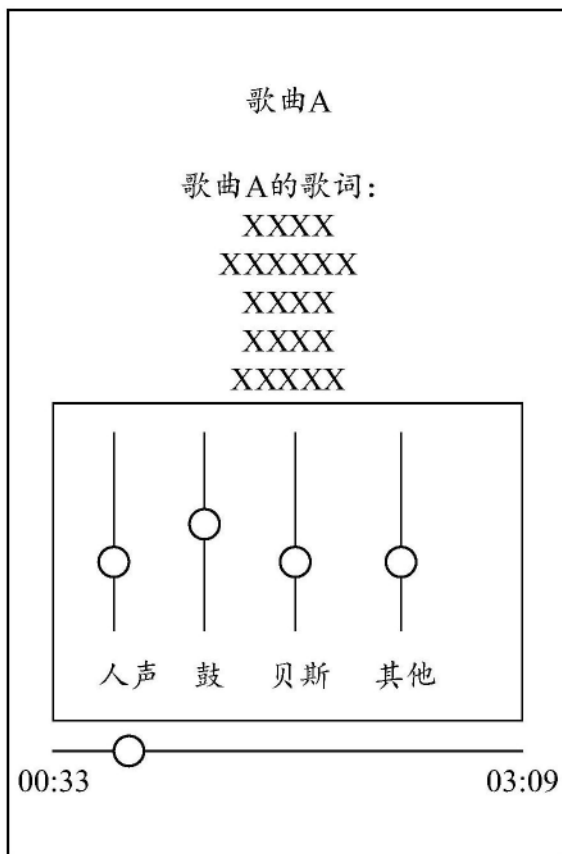


图4

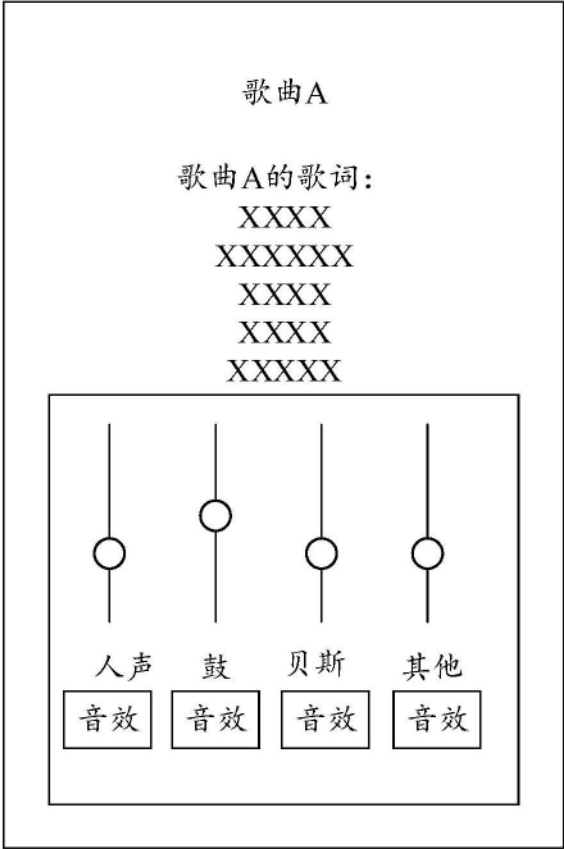


图5

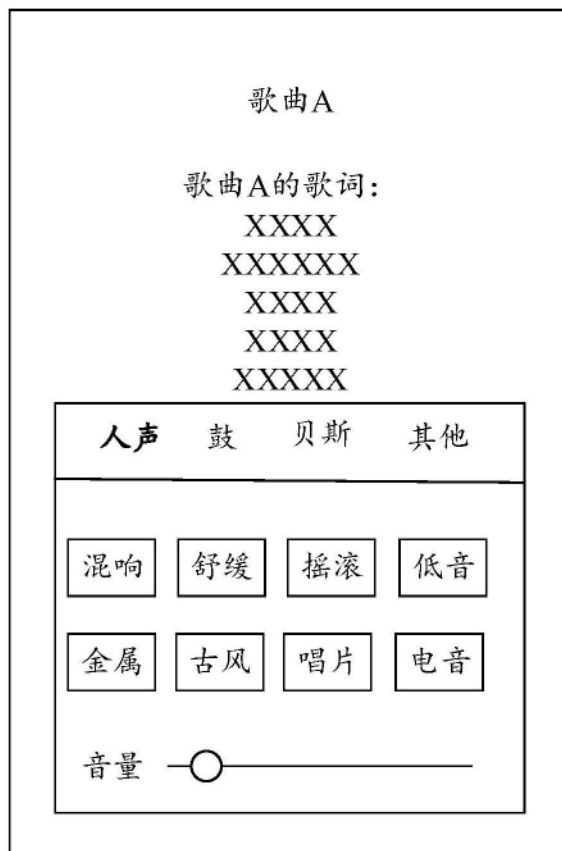


图6

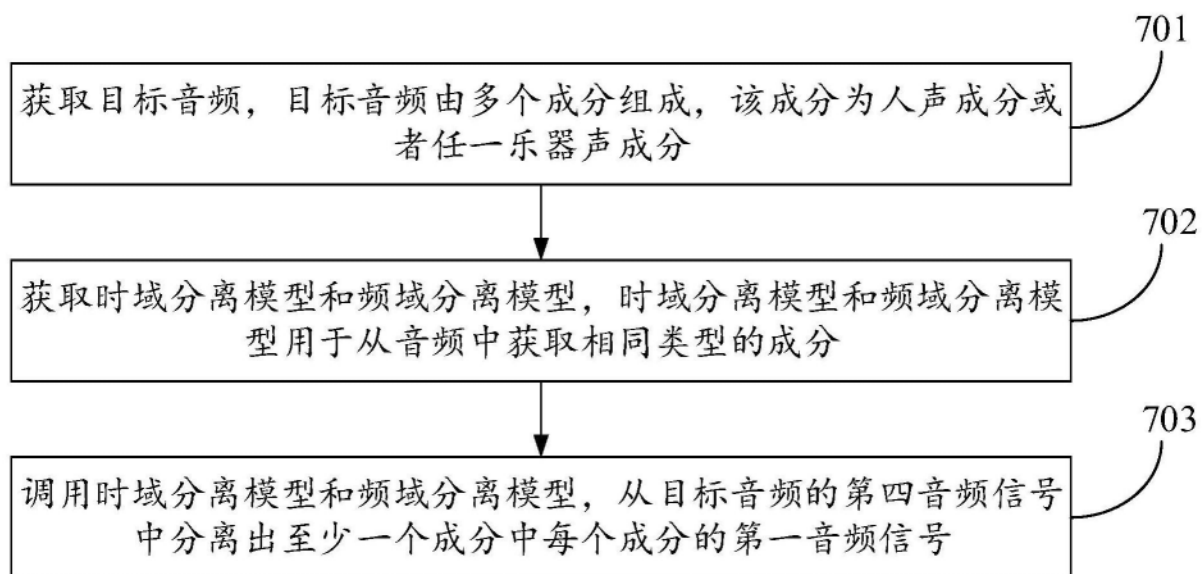


图7

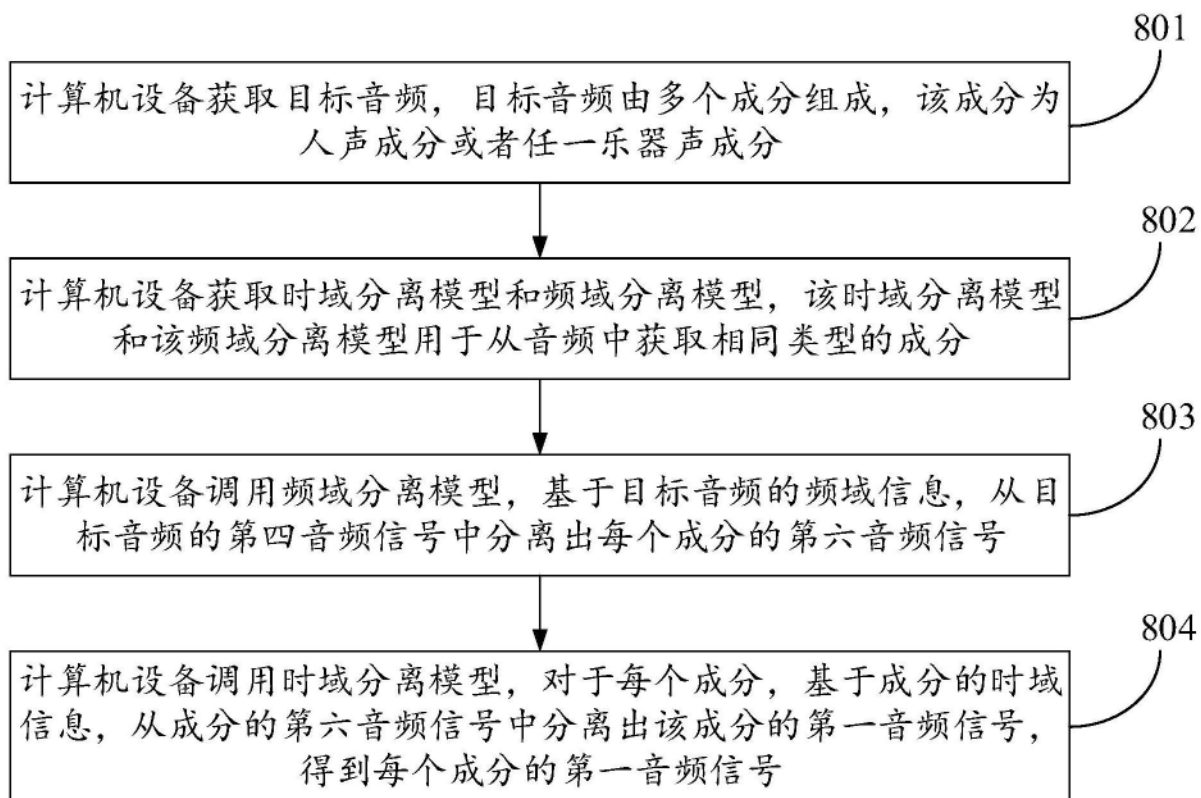


图8

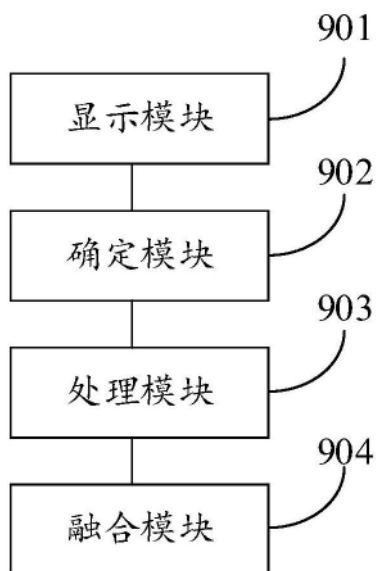


图9

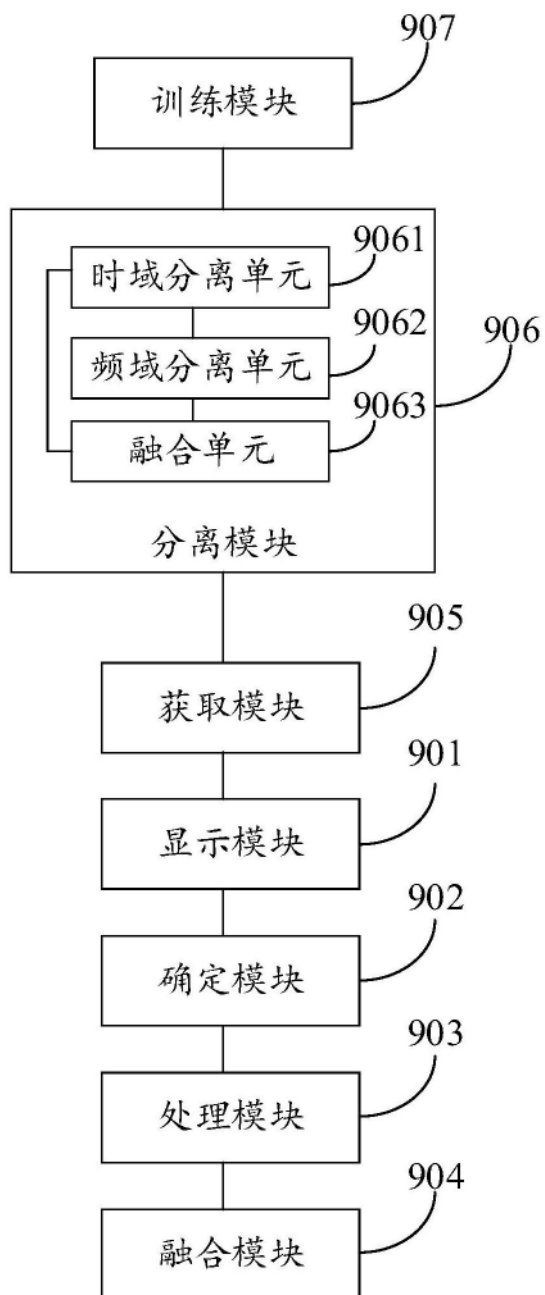


图10

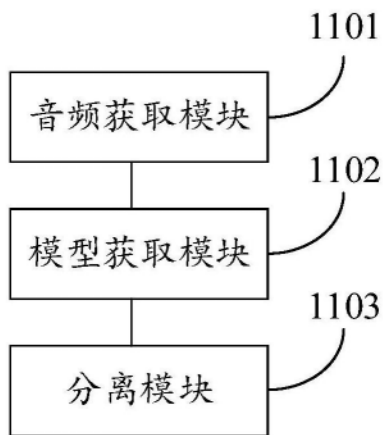


图11

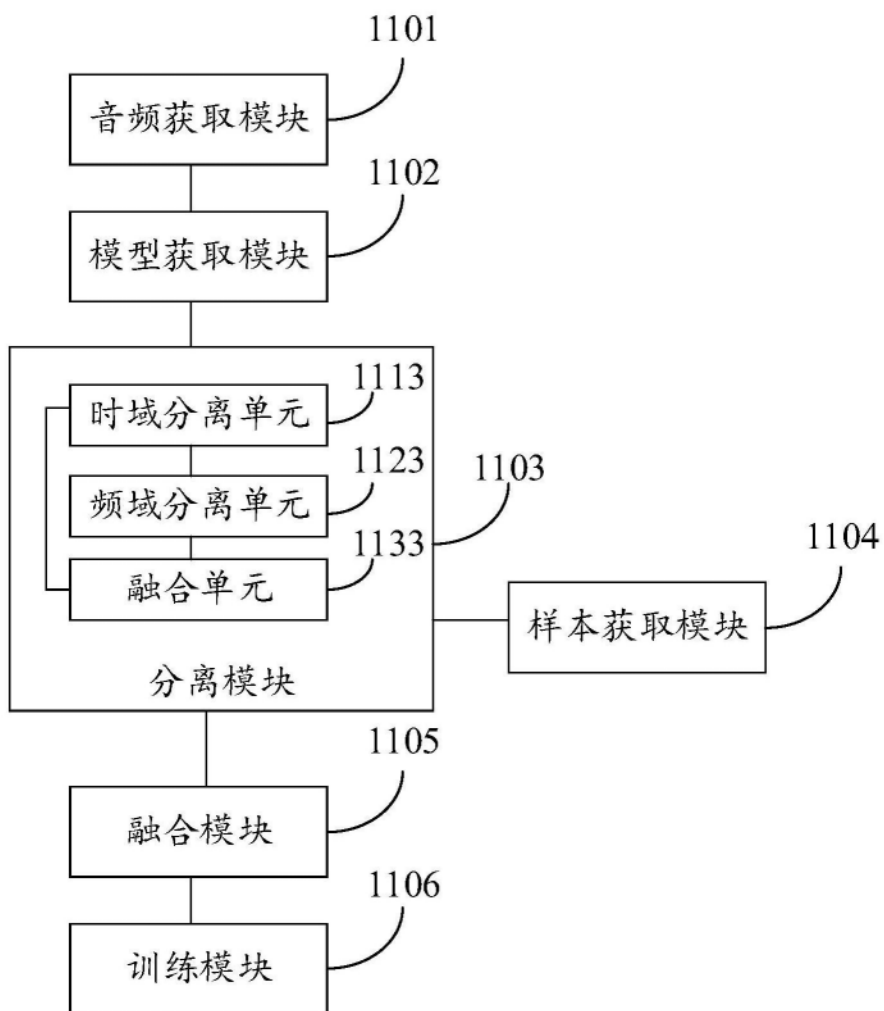


图12

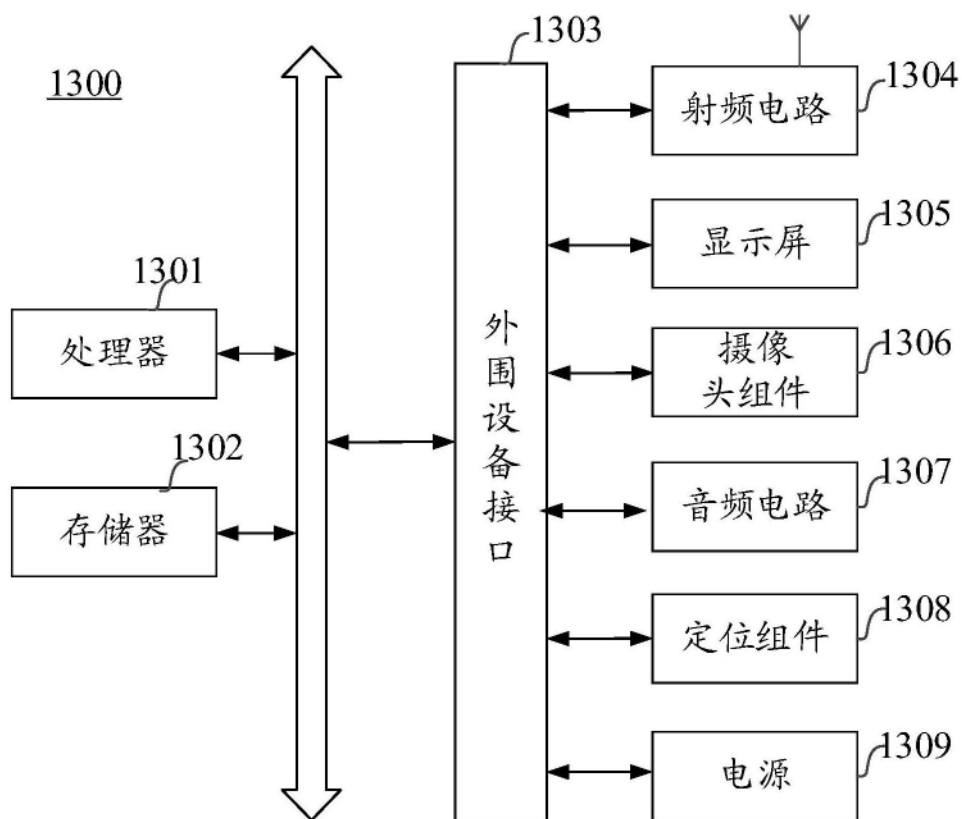


图13

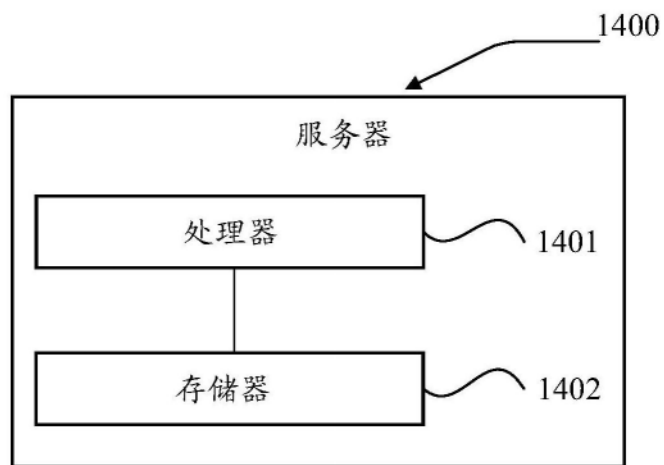


图14