(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104080024 B
(45)授权公告日 2019.02.19

(21)申请号 201310100422.1

(22)申请日 2013.03.26

(65)同一申请的已公布的文献号
申请公布号 CN 104080024 A

(43)申请公布日 2014.10.01

(73)专利权人 杜比实验室特许公司
地址 美国加利福尼亚州

(72)发明人 王珺 芦烈 阿兰·西费尔特

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 李春晖 李德山

(51)Int.Cl.
H04R 1/20(2006.01)
H04S 7/00(2006.01)

(56)对比文件
US 2011218798 A1,2011.09.08,
CN 101211557 A,2008.07.02,
WO 2006079813 A1,2006.08.03,
EP 2367286 A1,2011.09.21,

审查员 王鑫

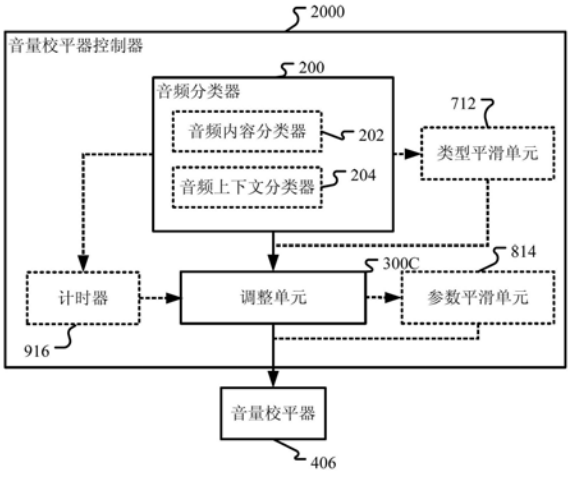
权利要求书6页 说明书59页 附图28页

(54)发明名称

音量校平器控制器和控制方法以及音频分类器

(57)摘要

公开了音量校平器控制器和控制方法、音频分类器和分类方法以及音频处理设备。在一个实施方式中,音量校平器控制方法包括:实时地识别音频信号的内容类型;以及通过随着音频信号的信息性内容类型的置信度值的增大或减小而分别增大或减小音量校平器的动态增益,并且随着音频信号的干扰性内容类型的置信度值的减小或增大而分别增大或减小音量校平器的动态增益,来基于所识别的内容类型以连续的方式调整音量校平器;其中,将音频信号分类到具有相应置信度值的多个内容类型中,并且调整的操作被配置成通过基于多个内容类型的重要性对多个内容类型的置信度值进行加权来考虑多个内容类型中的至少一些内容类型。



CN 104080024 B

1.一种音量校平器控制方法,包括:

实时地识别音频信号的内容类型;以及

通过随着所述音频信号的信息性内容类型的置信度值的增大或减小而分别增大或减小所述音量校平器的动态增益,并且随着所述音频信号的干扰性内容类型的置信度值的减小或增大而分别增大或减小所述音量校平器的动态增益,来基于所识别的内容类型以连续的方式调整音量校平器;

其中,将所述音频信号分类到具有相应置信度值的多个内容类型中,并且所述调整的操作被配置成通过基于所述多个内容类型的重要性对所述多个内容类型的置信度值进行加权来考虑所述多个内容类型中的至少一些内容类型。

2.根据权利要求1所述的音量校平器控制方法,其中,所述音频信号的所述内容类型包括语音、短期音乐、噪声和背景声音中的一个。

3.根据权利要求1所述的音量校平器控制方法,其中,噪声被视为干扰性内容类型。

4.根据权利要求1所述的音量校平器控制方法,其中,所述调整的操作被配置成基于所述内容类型的置信度值来调整所述音量校平器的动态增益。

5.根据权利要求4所述的音量校平器控制方法,其中,所述调整的操作被配置成通过所述内容类型的置信度值的传递函数来调整所述动态增益。

6.根据权利要求1所述的音量校平器控制方法,其中,所述调整的操作被配置成基于所述置信度值来考虑至少一个主导的内容类型。

7.根据权利要求1所述的音量校平器控制方法,其中,将所述音频信号分类到具有相应置信度值的多个干扰性内容类型中和/或具有相应置信度值的多个信息性内容类型中,并且所述调整的操作被配置成基于所述置信度值来考虑至少一个主导的干扰性内容类型和/或至少一个主导的信息性内容类型。

8.根据权利要求1至7中任一项所述的音量校平器控制方法,还包括,针对每个内容类型,基于所述音频信号的过去的置信度值来对所述音频信号的当次置信度值进行平滑。

9.根据权利要求8所述的音量校平器控制方法,其中,类型平滑操作被配置成通过计算当前的实际置信度值与上一次的经平滑的置信度值的加权和来确定所述音频信号当次的经平滑的置信度值。

10.根据权利要求2至7中任一项所述的音量校平器控制方法,还包括识别所述音频信号的上下文类型,其中,所述调整的操作被配置成基于所述上下文类型的置信度值来调整所述动态增益的范围。

11.根据权利要求2至7中任一项所述的音量校平器控制方法,还包括识别所述音频信号的上下文类型,其中,所述调整的操作被配置成基于所述音频信号的所述上下文类型来将所述音频信号的所述内容类型视为信息性的或者是干扰性的。

12.根据权利要求11所述的音量校平器控制方法,其中,所述音频信号的所述上下文类型包括VoIP、电影类媒体、长期音乐和游戏中的一个。

13.根据权利要求11所述的音量校平器控制方法,其中,在VoIP上下文类型的音频信号中,背景声音被视为干扰性内容类型;而在非VoIP上下文类型的音频信号中,所述背景声音和/或语音和/或音乐被视为信息性内容类型。

14.根据权利要求11所述的音量校平器控制方法,其中,取决于音频信号的上下文类

型，给不同上下文类型的音频信号中的所述内容类型分配不同的权重。

15.根据权利要求11所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个上下文类型中，并且所述调整的操作被配置成通过基于所述置信度值对所述多个上下文类型的影响进行加权来考虑所述多个上下文类型中的至少一些上下文类型。

16.根据权利要求11所述的音量校平器控制方法，其中，

识别所述内容类型的操作被配置成按所述音频信号的短期片段来识别所述内容类型；并且

识别所述上下文类型的操作被配置成至少部分地基于所识别的所述内容类型来按所述音频信号的短期片段识别所述上下文类型。

17.根据权利要求16所述的音量校平器控制方法，其中，识别所述内容类型的操作包括将短期片段分类到VoIP语音内容类型或者非VoIP语音内容类型中；并且

所述识别所述上下文类型的操作被配置成基于VoIP语音和非VoIP语音的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

18.根据权利要求17所述的音量校平器控制方法，其中，识别所述内容类型的操作还包括：

将短期片段分类到VoIP噪声内容类型和非VoIP噪声内容类型中；并且

识别所述上下文类型的操作被配置成基于VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

19.根据权利要求17所述的音量校平器控制方法，其中，识别所述上下文类型的操作被配置成：

如果VoIP语音的置信度值大于第一阈值，则将所述短期片段分类成所述VoIP上下文类型；

如果VoIP语音的置信度值不大于第二阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值；否则

将所述短期片段分类成上一个短期片段的上下文类型。

20.根据权利要求18所述的音量校平器控制方法，其中，识别所述上下文类型的操作被配置成：

如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值，则将所述短期片段分类成所述VoIP上下文类型；

如果VoIP语音的置信度值不大于第二阈值，或者如果VoIP噪声的置信度值不大于第四阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值，所述第四阈值不大于所述第三阈值；否则

将所述短期片段分类成上一个短期片段的上下文类型。

21.根据权利要求16至20中任一项所述的音量校平器控制方法，还包括基于所述内容类型的过去的置信度值来对所述内容类型的当次置信度值进行平滑。

22.根据权利要求21所述的音量校平器控制方法，其中，所述平滑的操作被配置成通过计算当前短期片段的置信度值与上一个短期片段的经平滑的置信度值的加权和来对当前短期片段的置信度值进行平滑。

23.根据权利要求22所述的音量校平器控制方法，还包括识别所述短期片段的语音内容类型，其中，在所述语音内容类型的置信度值低于第五阈值的情况下，将平滑之前的当前短期片段的VoIP语音的置信度值设定为预定的置信度值或者上一个短期片段的经平滑的置信度值。

24.根据权利要求17或18所述的音量校平器控制方法，其中，通过使用所述短期片段的所述内容类型的置信度值以及从所述短期片段提取的其他特征作为特征，基于机器学习模型对所述短期片段进行分类。

25.根据权利要求12至19中任一项所述的音量校平器控制方法，还包括测量所述识别所述上下文类型的操作连续地输出同一上下文类型的持续时间，其中，所述调整的操作被配置成继续使用当前的上下文类型直到新的上下文类型的持续时间的长度达到第六阈值为止。

26.根据权利要求25所述的音量校平器控制方法，其中，针对不同的从一个上下文类型到另一个上下文类型的转换对来设定不同的所述第六阈值。

27.根据权利要求25所述的音量校平器控制方法，其中，所述第六阈值与所述新的上下文类型的置信度值负相关。

28.根据权利要求19或20所述的音量校平器控制方法，其中，所述第一阈值和/或所述第二阈值取决于上一个短期片段的上下文类型而不同。

29.一种音量校平器控制方法，包括：

实时地识别音频信号的内容类型；以及

通过随着所述音频信号的信息性内容类型的置信度值的增大或减小而分别增大或减小所述音量校平器的动态增益，并且随着所述音频信号的干扰性内容类型的置信度值的减小或增大而分别增大或减小所述音量校平器的动态增益，来基于所识别的内容类型以连续的方式调整音量校平器；

其中，将所述音频信号分类到具有相应置信度值的多个音频内容中，并且所述调整的操作被配置成使用至少一个其他内容类型的置信度值来修改一个内容类型的权重。

30.根据权利要求29所述的音量校平器控制方法，其中，所述音频信号的所述内容类型包括语音、短期音乐、噪声和背景声音中的一个。

31.根据权利要求29所述的音量校平器控制方法，其中，噪声被视为干扰性内容类型。

32.根据权利要求29所述的音量校平器控制方法，其中，所述调整的操作被配置成基于所述内容类型的置信度值来调整所述音量校平器的动态增益。

33.根据权利要求32所述的音量校平器控制方法，其中，所述调整的操作被配置成通过所述内容类型的置信度值的传递函数来调整所述动态增益。

34.根据权利要求29所述的音量校平器控制方法，其中，所述调整的操作被配置成基于所述置信度值来考虑至少一个主导的内容类型。

35.根据权利要求29所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个干扰性内容类型中和/或具有相应置信度值的多个信息性内容类型中，并且所述调整的操作被配置成基于所述置信度值来考虑至少一个主导的干扰性内容类型和/或至少一个主导的信息性内容类型。

36.根据权利要求29至35中任一项所述的音量校平器控制方法，还包括，针对每个内容

类型，基于所述音频信号的过去的置信度值来对所述音频信号的当次置信度值进行平滑。

37.根据权利要求36所述的音量校平器控制方法，其中，类型平滑操作被配置成通过计算当前的实际置信度值与上一次的经平滑的置信度值的加权和来确定所述音频信号当次的经平滑的置信度值。

38.根据权利要求30至35中任一项所述的音量校平器控制方法，还包括识别所述音频信号的上下文类型，其中，所述调整的操作被配置成基于所述上下文类型的置信度值来调整所述动态增益的范围。

39.根据权利要求30至35中任一项所述的音量校平器控制方法，还包括识别所述音频信号的上下文类型，其中，所述调整的操作被配置成基于所述音频信号的所述上下文类型来将所述音频信号的所述内容类型视为信息性的或者是干扰性的。

40.根据权利要求39所述的音量校平器控制方法，其中，所述音频信号的所述上下文类型包括VoIP、电影类媒体、长期音乐和游戏中的一个。

41.根据权利要求39所述的音量校平器控制方法，其中，在VoIP上下文类型的音频信号中，背景声音被视为干扰性内容类型；而在非VoIP上下文类型的音频信号中，所述背景声音和/或语音和/或音乐被视为信息性内容类型。

42.根据权利要求29所述的音量校平器控制方法，其中，所述音频信号的上下文类型包括高质量音频或低质量音频。

43.根据权利要求39所述的音量校平器控制方法，其中，取决于音频信号的上下文类型，给不同上下文类型的音频信号中的所述内容类型分配不同的权重。

44.根据权利要求29所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个上下文类型中，并且所述调整的操作被配置成通过基于所述多个上下文类型的重要性对所述多个上下文类型的置信度值进行加权来考虑所述多个上下文类型中的至少一些上下文类型。

45.根据权利要求39所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个上下文类型中，并且所述调整的操作被配置成通过基于所述置信度值对所述多个上下文类型的影响进行加权来考虑所述多个上下文类型中的至少一些上下文类型。

46.根据权利要求39所述的音量校平器控制方法，其中，

识别所述内容类型的操作被配置成按所述音频信号的短期片段来识别所述内容类型；并且

识别所述上下文类型的操作被配置成至少部分地基于所识别的所述内容类型来按所述音频信号的短期片段识别所述上下文类型。

47.根据权利要求46所述的音量校平器控制方法，其中，识别所述内容类型的操作包括将短期片段分类到VoIP语音内容类型或者非VoIP语音内容类型中；并且

所述识别所述上下文类型的操作被配置成基于VoIP语音和非VoIP语音的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

48.根据权利要求47所述的音量校平器控制方法，其中，识别所述内容类型的操作还包括：

将短期片段分类到VoIP噪声内容类型和非VoIP噪声内容类型中；并且

识别所述上下文类型的操作被配置成基于VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

49.根据权利要求47所述的音量校平器控制方法，其中，识别所述上下文类型的操作被配置成：

如果VoIP语音的置信度值大于第一阈值，则将所述短期片段分类成所述VoIP上下文类型；

如果VoIP语音的置信度值不大于第二阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值；否则

将所述短期片段分类成上一个短期片段的上下文类型。

50.根据权利要求48所述的音量校平器控制方法，其中，识别所述上下文类型的操作被配置成：

如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值，则将所述短期片段分类成所述VoIP上下文类型；

如果VoIP语音的置信度值不大于第二阈值，或者如果VoIP噪声的置信度值不大于第四阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值，所述第四阈值不大于所述第三阈值；否则

将所述短期片段分类成上一个短期片段的上下文类型。

51.根据权利要求46至50中任一项所述的音量校平器控制方法，还包括基于所述内容类型的过去的置信度值来对所述内容类型的当次置信度值进行平滑。

52.根据权利要求51所述的音量校平器控制方法，其中，所述平滑的操作被配置成通过计算当前短期片段的置信度值与上一个短期片段的经平滑的置信度值的加权和来对当前短期片段的置信度值进行平滑。

53.根据权利要求52所述的音量校平器控制方法，还包括识别所述短期片段的语音内容类型，其中，在所述语音内容类型的置信度值低于第五阈值的情况下，将平滑之前的当前短期片段的VoIP语音的置信度值设定为预定的置信度值或者上一个短期片段的经平滑的置信度值。

54.根据权利要求47或48所述的音量校平器控制方法，其中，通过使用所述短期片段的所述内容类型的置信度值以及从所述短期片段提取的其他特征作为特征，基于机器学习模型对所述短期片段进行分类。

55.根据权利要求40至49中任一项所述的音量校平器控制方法，还包括测量所述识别所述上下文类型的操作连续地输出同一上下文类型的持续时间，其中，所述调整的操作被配置成继续使用当前的上下文类型直到新的上下文类型的持续时间的长度达到第六阈值为止。

56.根据权利要求55所述的音量校平器控制方法，其中，针对不同的从一个上下文类型到另一个上下文类型的转换对来设定不同的所述第六阈值。

57.根据权利要求55所述的音量校平器控制方法，其中，所述第六阈值与所述新的上下文类型的置信度值负相关。

58.根据权利要求49或50所述的音量校平器控制方法，其中，所述第一阈值和/或所述第二阈值取决于上一个短期片段的上下文类型而不同。

59.一种音量校平器控制器,包括:

音频内容分类器,用于实时地识别音频信号的内容类型;以及

调整单元,用于基于所识别的内容类型来以连续的方式调整音量校平器;

其中,所述音量校平器控制器被配置为执行权利要求1至58中任一项所述的方法。

60.一种音频处理设备,其包括根据权利要求59所述的音量校平器控制器。

音量校平器控制器和控制方法以及音频分类器

技术领域

[0001] 本申请总体上涉及音频信号处理。具体地，本申请的实施方式涉及用于音频分类和音频处理的设备和方法，尤其涉及对对话增强器、环绕声虚拟器、音量校平器和均衡器的控制。

背景技术

[0002] 为了提升音频的整体质量并且相应地提升用户体验，一些音频改善装置用于在时域中或者谱域中修改音频信号。已经针对各种目的开发出了各种音频改善装置。音频改善装置的一些常见示例包括：

[0003] 对话增强器：在电影和广播或者电视节目中，对于理解故事来说，对话是最重要的成分。为了提高其清晰度和其可理解性，尤其是对于听力下降的年长者，开发出了增强对话的方法。

[0004] 环绕声虚拟器：环绕声虚拟器使得能够在PC(个人电脑)的内置扬声器中或者耳机中渲染出环绕(多声道)声音信号。也就是说，通过立体声装置(例如扬声器和耳机)，环绕声虚拟器为用户生成虚拟的环绕声效果，提供电影的体验。

[0005] 音量校平器：音量校平器旨在对回放的音频内容的音量进行调节，并且基于目标响度值来使音量在时间轴上几乎保持一致。

[0006] 均衡器：均衡器提供被称为“音调”或者“音色”的谱平衡的一致性，并且使用户能够为了放大某些声音或者去除不期望的声音而在每个单独的频带上配置频率响应(增益)的整体模式(曲线或者形状)。在传统的均衡器中，可以针对不同的声音例如不同的音乐风格而提供不同的均衡器预置。一旦选择了预置，或者设置了均衡模式，则在信号上施加相同的均衡增益，直到该均衡模式被手动修改为止。相比之下，动态均衡器通过连续监测音频的谱平衡，将其与期望的音调相比较并且动态地调整均衡滤波器以将音频的原始音调转变为期望音调，来实现谱平衡一致性。

[0007] 通常，音频改善装置具有其自身的应用情景/上下文。也就是说，音频改善装置可能只适用于特定的内容集合而不适用于所有可能的音频信号，因为不同的内容可能需要以不同的方式来处理。例如，对话增强方法通常被应用于电影内容。如果将对话增强方法应用于其中没有对话的音乐，则对话增强方法可能错误地增强一些频率子带并且引入大量的音色变化和感知上的不一致性。类似地，如果将噪声抑制方法施加到音乐信号上，则能够听到强烈的畸变。

[0008] 但是，对于通常包括一组音频改善装置的音频处理系统来说，其输入不可避免地可能是所有可能类型的音频信号。例如，集成在PC中的音频处理系统将接收来自各种源的音频内容，包括电影、音乐、VoIP和游戏。因此，为了对相应内容应用较好的算法或者应用每个算法的较好的参数，重要的是识别或者区分这些被处理的内容。

[0009] 为了区分音频内容并且相应地应用较好的参数或者较好的音频改善算法，传统的系统通常预先设计一组预置，并且要求用户针对要播放的内容来选择预置。预置通常将一

组音频改善算法和/或其要应用的最佳参数进行编码，例如针对电影或者音乐回放而特别设计的“电影”预置和“音乐”预置。

[0010] 但是，对于用户来说，手动选择并不方便。用户通常不会在各种预定义的预置间进行频繁的切换，而是对所有内容保持使用一个预置。此外，即使在一些自动解决方案中，在预置中的参数或者算法设置通常是离散的 (例如，对针对特定内容的特定算法进行开启或者关闭)，其不能以基于内容的连续的方式来调整参数。

发明内容

[0011] 本申请的第一方面是基于回放的音频内容以连续的方式来自动地配置音频改善装置。通过该“自动”模式，用户可以不用疲于选择不同的预置，而只是享受他们的内容。另一方面，为了避免在转换点处的可听到的畸变，连续的调节更加重要。

[0012] 根据第一方面的实施方式，一种音频处理设备包括：音频分类器，用于将音频信号实时地分类到至少一个音频类型中；音频改善装置，用于改善听众体验；以及调整单元，用于基于该至少一个音频类型的置信度值来以连续的方式调整音频改善装置的至少一个参数。

[0013] 音频改善装置可以是对话增强器、环绕声虚拟器、音量校平器和均衡器中的任何装置。

[0014] 相应地，一种音频处理方法包括：将音频信号实时地分类到至少一个音频类型中；以及基于该至少一个音频类型的置信度值来以连续的方式调整至少一个用于音频改善的参数。

[0015] 根据第一方面的另一个实施方式，一种音量校平器控制器包括：音频内容分类器，用于实时地识别音频信号的内容类型；以及调整单元，用于基于所识别的内容类型来以连续的方式调整音量校平器。调整单元可以配置为使音量校平器的动态增益与音频信号的信息性内容类型正相关，且使音量校平器的动态增益与音频信号的干扰性内容类型负相关。

[0016] 还公开了一种包括上述音量校平器控制器的音频处理设备。

[0017] 相应地，一种音量校平器控制方法包括：实时地识别音频信号的内容类型；通过使音量校平器的动态增益与音频信号的信息性内容类型正相关，并且使音量校平器的动态增益与音频信号的干扰性内容类型负相关，而基于所识别的内容类型来以连续的方式调整音量校平器。

[0018] 根据第一方面的又一个实施方式，一种均衡器控制器包括：音频分类器，用于实时地识别音频信号的音频类型；以及调整单元，用于基于所识别的音频类型来以连续的方式调整均衡器。

[0019] 还公开了一种包括上述均衡器控制器的音频处理设备。

[0020] 相应地，一种均衡器控制方法包括：实时地识别音频信号的音频类型；以及基于所识别的音频类型来以连续的方式调整均衡器。

[0021] 本申请还提供了在其上记录有计算机程序指令的计算机可读介质，当由处理器来执行该指令时，该指令使处理器能够执行上述的音频处理方法、或者音量校平器控制方法、或者均衡器控制方法。

[0022] 根据第一方面的各个实施方式，可以根据音频信号的类型和/或该类型的置信度

值来连续地调整音频改善装置，该音频改善装置可以是对话增强器、环绕声虚拟器、音量校平器和均衡器中之一。

[0023] 本申请的第二方面是开发内容识别组件来识别多个音频类型，并且可以使用检测结果通过以连续的方式找到较好的参数来操纵/指导各种音频改善装置的工作方式。

[0024] 根据第二方面的实施方式，音频分类器包括：短期特征提取器，用于从各自包括音频帧序列的短期音频片段中提取短期特征；短期分类器，用于使用相应的短期特征来将长期音频片段中的短期音频片段序列分类到短期音频类型中；统计数据提取器，用于计算短期分类器针对该长期音频片段中的短期音频片段序列的结果的统计数据，作为长期特征；以及长期分类器，用于使用长期特征来将长期音频片段分类到长期音频类型中。

[0025] 还公开了一种包括上述音频分类器的音频处理设备。

[0026] 相应地，一种音频分类方法包括：从各自包括音频帧序列的短期音频片段中提取短期特征；使用相应的短期特征来将长期音频片段中的短期音频片段序列分类到短期音频类型中；计算短期分类器针对该长期音频片段中的短期音频片段序列的结果的统计数据，作为长期特征；以及使用长期特征来将长期音频片段分类到长期音频类型中。

[0027] 根据第二方面的另一个实施方式，一种音频分类器包括：音频内容分类器，用于识别音频信号的短期片段的内容类型；以及音频上下文分类器，用于至少部分地基于由音频内容分类器所识别的内容类型来识别该短期片段的上下文类型。

[0028] 还公开了包括上述音频分类器的音频处理设备。

[0029] 相应地，一种音频分类方法包括：识别音频信号的短期片段的内容类型；以及至少部分地基于所识别的内容类型来识别该短期片段的上下文类型。

[0030] 本公开内容还提供了其上记录有计算机程序指令的计算机可读介质，当由处理器来执行该指令时，该指令使处理器能够执行上述的音频分类方法。

[0031] 根据第二方面的各个实施方式，音频信号可以被分类到不同的长期类型或者上下文类型中，该长期类型或者上下文类型与短期类型或者内容类型不同。音频信号的类型和/或类型的置信度值还可以用于调整音频改善装置，例如对话增强器、环绕声虚拟器、音量校平器或者均衡器。

[0032] 根据一个实施例，一种音量校平器控制方法包括：实时地识别音频信号的内容类型；以及通过随着音频信号的信息性内容类型的置信度值的增大或减小而分别增大或减小音量校平器的动态增益，并且随着音频信号的干扰性内容类型的置信度值的减小或增大而分别增大或减小音量校平器的动态增益，来基于所识别的内容类型以连续的方式调整音量校平器；其中，将音频信号分类到具有相应置信度值的多个内容类型中，并且调整的操作被配置成通过基于多个内容类型的重要性对多个内容类型的置信度值进行加权来考虑多个内容类型中的至少一些内容类型。

[0033] 根据另一个实施例，一种音量校平器控制方法包括：实时地识别音频信号的内容类型；以及通过随着音频信号的信息性内容类型的置信度值的增大或减小而分别增大或减小音量校平器的动态增益，并且随着音频信号的干扰性内容类型的置信度值的减小或增大而分别增大或减小音量校平器的动态增益，来基于所识别的内容类型以连续的方式调整音量校平器；其中，将音频信号分类到具有相应置信度值的多个音频内容中，并且调整的操作被配置成使用至少一个其他内容类型的置信度值来修改一个内容类型的权重。

[0034] 根据一个实施例，提供一种音量校平器控制器，其包括：音频内容分类器，用于实时地识别音频信号的内容类型；以及调整单元，用于基于所识别的内容类型来以连续的方式调整音量校平器；其中，音量校平器控制器被配置为执行上述音量校平器控制方法。

[0035] 根据另一个实施例，提供一种音频处理设备，其包括上述音量校平器控制器。

[0036] 根据一个实施例，一种音频分类方法包括：识别音频信号的短期片段的内容类型；以及至少部分地基于所识别的内容类型来识别短期片段的上下文类型。

[0037] 根据另一个实施例，提供一种音频分类器，其包括：音频内容分类器，用于识别音频信号的短期片段的内容类型；以及音频上下文分类器，用于至少部分地基于由音频内容分类器识别的内容类型来识别短期片段的上下文类型；其中，音频分类器被配置为执行上述音频分类方法。

[0038] 根据又一个实施例，提供一种音频处理设备，其包括上述音频分类器。

附图说明

[0039] 在附图中，以示例的方式而非限制的方式图解了本申请，在附图中，相同的附图标记表示相似的元素，在附图中：

[0040] 图1的示意图图解了根据本申请的实施方式的音频处理设备；

[0041] 图2和图3的示意图图解了如图1所示的实施方式的变型；

[0042] 图4至图6的示意图图解了用于识别多个音频类型和计算置信度值的分类器的可能架构；

[0043] 图7至图9的示意图图解了本申请的音频处理设备的更多实施方式；

[0044] 图10的示意图图解了不同音频类型之间的转换延迟；

[0045] 图11至图14是根据本申请的实施方式的音频处理方法的流程图；

[0046] 图15的示意图图解了根据本申请的实施方式的对话增强控制器；

[0047] 图16和图17是在对对话增强器的控制中使用根据本申请的音频处理方法的流程图；

[0048] 图18的示意图图解了根据本申请的实施方式的环绕声虚拟器控制器；

[0049] 图19是在对环绕声虚拟器的控制中使用根据本申请的音频处理方法的流程图；

[0050] 图20的示意图图解了根据本申请的实施方式的音量校平器控制器；

[0051] 图21的示意图图解了根据本申请的音量校平器控制器的效果；

[0052] 图22的示意图图解了根据本申请的实施方式的均衡器控制器；

[0053] 图23示出了期望的谱平衡预置的若干示例；

[0054] 图24的示意图图解了根据本申请的实施方式的音频分类器；

[0055] 图25和图26的示意图图解了由根据本申请的音频分类器所使用的一些特征；

[0056] 图27至图29的示意图图解了根据本申请的音频分类器的更多实施方式；

[0057] 图30至图33是根据本申请的实施方式的音频分类方法的流程图；

[0058] 图34的示意图图解了根据本申请的另一个实施方式的音频分类器；

[0059] 图35的示意图图解了根据本申请的又一个实施方式的音频分类器；

[0060] 图36的示意图图解了本申请的音频分类器中使用的启发式规则；

[0061] 图37和图38的示意图图解了根据本申请的音频分类器的更多实施方式；

[0062] 图39和图40是根据本申请的实施方式的音频分类方法的流程图;以及

[0063] 图41是用于实现根据本申请的实施方式的示例性系统的框图。

具体实施方式

[0064] 以下参照附图描述本申请的实施方式。要注意的是,为了清楚起见,在附图和描述中省略了对本领域的技术人员所公知的且对于理解本申请并非必需的那些组件和处理的表示和描述。

[0065] 本领域的技术人员要理解的是,本申请的各个方面可以被实施为系统、装置(例如,蜂窝式电话、便携式媒体播放器、个人计算机、服务器、电视机顶盒或者数字录像机,或者任何其他媒体播放器)、方法或者计算机程序产品。因此,本申请的各个方面可以采取硬件实施方式的形式、软件实施方式(包括固件、驻留软件、微码等)的形式或者将软件与硬件方面组合起来的实施方式的形式,这里通常可以将它们称为“电路”、“模块”、“系统”。而且,本申请的各个方面可以采取其上包括了计算机可读程序编码的一个或者更多个计算机可读介质中所包括的计算机程序产品的形式。

[0066] 可以使用一个或者更多个计算机可读介质的任何组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是,但不限于,电子的、磁性的、光学的、电磁的、红外的或者半导体的系统、设备或者装置,或者是上述的任何适当的组合。计算机可读存储介质的更具体的示例(非穷举性的列举)可以包括:具有一个或者更多个导线的电连接、便携式计算机软盘、硬盘、随机存取存储器 (RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或者闪存)、光纤、光盘只读存储器(CD-ROM)、光存储器装置、磁性存储装置、或者上述的任何适当的组合。在本文档的语境中,计算机可读存储介质可以是能够包括或者存储用于由指令执行系统、设备或者装置所使用或者或者与之结合使用的程序的任何有形介质。

[0067] 计算机可读信号介质可以包括其中包含有计算机可读程序编码的传播数据信号,例如在基带中或者作为载波的一部分。这样的传播信号可以采取各种形式,包括但不限于,电磁信号或者光学信号,或者其任何合适的组合。

[0068] 计算机可读信号介质可以是除计算机可读存储介质之外的任何计算机可读介质,其能够通信、传播或者传输由指令执行系统、设备或者装置使用或者或者与之结合使用的程序。

[0069] 计算机可读介质中所包括的程序编码可以使用任何适当的介质被传送,适当的介质包括但不限于:无线线路、有线线路、光缆、RF(射频) 等,或者上述的任何合适的组合。

[0070] 用于针对本申请的各个方面而执行操作的计算机程序编码可以以一个或者更多个编程语言的任何组合来编写,编程语言包括面向对象的编程语言例如Java、Smalltalk、C++等,以及常规程序编程语言,例如“C”编程语言或者类似的编程语言。程序编码可以作为独立软件包来完全地在用户的计算机上执行,或者部分在用户的计算机上执行、部分在远程计算机上执行,或者完全在远程计算机或者服务器上执行。在后者的场景中,远程计算机可以通过任意类型的网络连接至用户的计算机,任意类型的网络包括局域网(LAN)或者广域网(WAN),或者可以连接至外部计算机 (例如,使用互联网服务运营商通过互联网连接)。

[0071] 以下,通过根据本申请的实施方式的方法、设备(系统)和计算机程序产品的流程

图图解和/或框图来描述本申请的各个方面。要理解的是，流程图图解和/或框图的每个框，以及流程图图解和/或框图的框的组合，可以由计算机程序指令来实现。可以将这些计算机程序指令提供给通用计算机、专用计算机或者其他可编程数据处理设备的处理器，以形成机器，使得通过计算机或者其他可编程数据处理设备的处理器执行的指令形成用于实现流程图和/或框图的一个块或者多个块中所指定的功能/动作的装置。

[0072] 这些计算机程序指令还可以存储在计算机可读介质中，其能够指导计算机、其他可编程数据处理设备、或者其他装置来以特定的方式工作，以使得在计算机可读介质中所存储的指令生产出一种制造品，该制造品包括实现流程图和/或框图的一个块或者多个块中所指定的功能/动作的指令。

[0073] 计算机编程指令还可以加载到计算机、其他可编程数据处理设备或者其他装置上，以引起一系列要在计算机、其他可编程数据处理设备或者其他装置上进行的运算操作，从而产生计算机实施的处理，以使得在计算机或其他可编程数据处理设备上执行的指令提供用于实现在流程图和/或框图的一个块或者多个块中所指定的功能/动作的处理。

[0074] 以下将详细描述本申请的实施方式，为了清楚起见，按照以下架构来组织描述：

[0075] 第1部分：音频处理设备和方法

[0076] 小节1.1音频类型

[0077] 小节1.2音频类型的置信度值和分类器的架构

[0078] 小节1.3对音频类型的置信度值进行平滑

[0079] 小节1.4参数调整

[0080] 小节1.5参数平滑

[0081] 小节1.6音频类型的转换

[0082] 小节1.7实施方式和应用场景的组合

[0083] 小节1.8音频处理方法

[0084] 第2部分：对话增强器控制器和控制方法

[0085] 小节2.1对话增强的级别

[0086] 小节2.2用于确定要增强的频带的阈值

[0087] 小节2.3对背景声级的调整

[0088] 小节2.4实施方式和应用场景的组合

[0089] 小节2.5对话增强器控制方法

[0090] 第3部分：环绕声虚拟器控制器和控制方法

[0091] 小节3.1环绕声增强量

[0092] 小节3.2起始频率

[0093] 小节3.3实施方式和应用场景的组合

[0094] 小节3.4环绕声虚拟器控制方法

[0095] 第4部分：音量校平器控制器和控制方法

[0096] 小节4.1信息性内容类型和干扰性内容类型

[0097] 小节4.2不同上下文中的内容类型

[0098] 小节4.3上下文类型

[0099] 小节4.4实施方式和应用场景的组合

[0100] 小节4.5音量校平器控制方法

[0101] 第5部分:均衡控制器和控制方法

[0102] 小节5.1基于内容类型的控制

[0103] 小节5.2音乐中存在主导源的可能性

[0104] 小节5.3均衡器的预置

[0105] 小节5.4基于上下文类型的控制

[0106] 小节5.5实施方式和应用场景的组合

[0107] 小节5.6均衡器控制方法

[0108] 第6部分:音频分类器和分类方法

[0109] 小节6.1基于内容类型分类的上下文分类器

[0110] 小节6.2长期特征的提取

[0111] 小节6.3短期特征的提取

[0112] 小节6.4实施方式和应用场景的组合

[0113] 小节6.5音频分类方法

[0114] 第7部分:VoIP分类器和分类方法

[0115] 小节7.1基于短期片段的上下文分类

[0116] 小节7.2使用VoIP语音和VoIP噪声的分类

[0117] 小节7.3使波动平滑

[0118] 小节7.4实施方式和应用场景的组合

[0119] 小节7.5VoIP分类方法

[0120] 第1部分:音频处理设备和方法

[0121] 图1示出了适应于内容的音频处理设备100的总体框架,该适应于内容的音频处理设备100支持基于回放的音频内容来以改善的参数自动地配置至少一个音频改善装置400。该总体框架包括三个主要部分:音频分类器200、调整单元300和音频改善装置400。

[0122] 音频分类器200用于将音频信号实时地分类到至少一个音频类型中。音频分类器200自动地识别回放内容的音频类型。任何音频分类技术,比如通过信号处理、机器学习和模式识别实现的音频分类技术,可以应用于识别音频内容。通常可以同时估算置信度值,置信度值代表音频内容针对一组预定义的目标音频类型的概率。

[0123] 音频改善装置400用于通过对音频信号进行处理来提升听众体验,稍后将会详细描述音频改善装置400。

[0124] 调整单元300用于基于至少一个音频类型的置信度值来以连续的方式调整音频改善装置的至少一个参数。调整单元300被设计用于操纵音频改善装置400的工作方式。调整单元300基于从音频分类器200获得的结果来估算相应音频改善装置的最适当的参数。

[0125] 在此设备中可以应用各种音频改善装置。图2示出了包括四个音频改善装置的示例性系统,该系统中包括对话增强器(Dialog Enhancer,DE) 402、环绕声虚拟器(Surround Virtualizer,SV)404、音量校平器(Volume Leveler,VL)406和均衡器(Equalizer,EQ)408。基于在音频分类器200 中获得的结果(音频类型和/或置信度值),能够以连续的方式自动地调整每个音频改善装置。

[0126] 当然,音频处理设备可以不必包括所有类别的音频改善装置,而可以只包括其中

的一个或者更多个音频改善装置。另一方面，音频改善装置不限于本公开内容中给出的那些装置，而可以包括更多类型的音频改善装置，其也在本申请的范围内。此外，本公开内容中讨论的那些音频改善装置的名称，包括对话增强器(DE)402、环绕声虚拟器(SV)404、音量校平器(VL)406和均衡器(EQ)408，不应构成限制，它们中的每个应被理解为覆盖实现相同或相似功能的任何其他装置。

[0127] 1.1音频类型

[0128] 为了适当地控制各种类型的音频改善装置，本发明还提供了音频类型的新的架构，然而现有技术中的那些音频类型也可以应用于此。

[0129] 具体地，对不同语意级别的音频类型进行了建模，包括代表音频信号中的基本组分的低级别音频元素和代表实际生活中用户的娱乐应用中最普遍的音频内容的高级别音频类型。前者也可以被命名为“内容类型”，基本的音频内容类型可以包括语音(speech)、音乐(music，包括歌曲)、背景声音(background sound，或者音效)和噪声(noise)。

[0130] 语音和音乐的含义不言而喻。在本申请中的噪声意指物理噪声，而不是指语意的噪声。在本申请中，物理噪声可以包括来自例如空调的噪声，以及发自技术原因的噪声例如由于信号传输路径所导致的粉红噪声。相比之下，本申请中的“背景声音”是那些可以是发生在听者注意力的核心目标周围的听觉事件的音效。例如，在电话通话中的音频信号中，除了通话者的声音，还可以有一些其他的非有意的声音，例如与该电话通话无关的一些其他人的声音、键盘的声音、脚步的声音等。这些不需要的声音被称为“背景声音”，而不是噪声。换言之，可以将“背景声音”定义为并非目标(或者听者注意力的核心目标)的或者甚至是不希望的，但是仍有一些语意含义的声音；而“噪声”可以定义为除了目标声音和背景声音之外的那些不需要的声音。

[0131] 有时背景声音真的不是“不需要的”而是有意生成的并且承载一些有用的信息，例如电影、电视节目或者无线电广播节目中的背景声音。所以，有时背景声音也可以被称为“音效”。在本公开内容的下文中，为了简洁性而只使用“背景声音”，并且也可简称为“背景”。

[0132] 进一步，音乐还可以被分为没有主导源的音乐和有主导源的音乐。如果在音乐片段中有一个源(嗓音或乐器)远比其他源更强，则该音乐被称为“有主导源的音乐”，否则就被称为“无主导源的音乐”。例如，在伴有歌唱声和各种乐器的复调音乐中，如果其是和声平衡的，或者若干最主要的源的能量是彼此相当的，则其被视为没有主导源的音乐；相比之下，如果一个源(例如，嗓音)响度高得多而其他源安静得多，则其被视为包括了主导源。作为另一个示例，单个的或者是突出的乐器音调是“具有主导源的音乐”。

[0133] 音乐还可以基于不同的标准被分为不同的类型。其可以基于音乐的风格来分类，例如摇滚、爵士、说唱和民谣，但不限于此。其还可以基于乐器被分类，例如声乐和器乐。器乐可以包括以不同乐器演奏的各种音乐，例如钢琴音乐和吉他音乐。其他示例性的标准包括音乐的节奏、速度、音色和/或任何其他音乐特征，以使得音乐可以基于这些特征的相似性而被归类。例如，根据音色，声乐可以被分为男高音、男中音、男低音、女高音、女中音和女低音。

[0134] 音频信号的内容类型可以针对例如包括多个帧的短期音频片段来分类。通常，音频帧的长度是多个毫秒，例如20ms，而要被音频分类器分类的短期音频片段的长度可以具

有从数百个毫秒到数秒的长度，例如1秒。

[0135] 为了以适应于内容的方式来控制音频改善装置，音频信号可以被实时地分类。针对以上所陈述的内容类型，当前的短期音频片段的内容类型代表当前的音频信号的内容类型。因为短期音频片段的长度不是很长，所以音频信号可以被相继划分为非重叠的短期音频片段。但是，短期音频片段也可以沿着音频信号的时间轴被连续地/半连续地取样。也就是说，短期音频片段可以用以一个或者更多个帧的步长沿着音频信号的时间轴移动的预定长度(所要的短期音频片段长度)的窗来取样。

[0136] 高级别音频类型也可以被命名为“上下文类型”，因为其指示音频信号的长期类型，并且可以被当作是可以分类到上述内容类型的瞬时声音事件的环境或者上下文。根据本申请，上下文类型可以包括最普遍的音频应用，例如电影类媒体(movie-like media)、音乐(music，包括歌曲)、游戏 (game)和VoIP(互联网协议语音)。

[0137] 音乐、游戏和VoIP的含义不言而喻。电影类媒体可以包括电影、电视节目、无线电广播节目或者与前面提到的类似的任何其他音频媒体。电影类媒体的主要特征是混合了可能的语音、音乐和各种类型的背景声音 (音效)。

[0138] 需要注意的是，内容类型和上下文类型都包括音乐(包括歌曲)。在本申请的下文中，使用词汇“短期音乐(short-term music)”和“长期音乐 (long-term music)”来分别区分这两者。

[0139] 针对本申请的一些实施方式，还提出了一些其他的上下文类型架构。

[0140] 例如，音频信号可以被分类为高质量的音频(例如电影类媒体和音乐 CD)或者低质量的音频(例如VoIP、低比特率的在线流音频和用户生成的内容)，其可以被统称为“音频质量类型”。

[0141] 作为另一个示例，音频信号可以被分类为VoIP或者非VoIP，其可以被视为上述的4上下文类型架构(VoIP、电影类媒体、(长期)音乐和游戏)的变形。与VoIP或者非VoIP的上下文相关地，音频信号可以被分为与VoIP相关的音频内容类型，例如VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声。VoIP音频内容类型的架构对于区分VoIP和非VoIP上下文尤其有用，因为VoIP上下文通常是音量校平器(一种音频改善装置) 的最具挑战性的应用场景。

[0142] 通常，音频信号的上下文类型可以针对比短期音频片段更长的长期音频片段来分类。长期音频片段包括的多个帧的数量比短期音频片段中的帧的数量更多。长期音频片段也可以包括多个短期音频片段。通常，长期音频片段可以具有秒数量级的长度，例如数秒至数十秒，如10秒。

[0143] 类似地，为了以自适应的方式来控制音频改善装置，音频信号可以被实时地分类到上下文类型中。类似地，当前的长期音频片段的上下文类型代表当前的音频信号的上下文类型。因为长期音频片段的长度相对地长，所以音频信号可以沿着音频信号的时间轴被连续地/半连续地取样，以避免其上下文类型的急剧变化以及因此导致的音频改善装置的工作参数的急剧变化。也就是说，长期音频片段可以使用预定长度(想要的长期音频片段长度)的窗以一个或者更多个帧的步长，或者以一个或者更多个短期片段的步长沿着音频信号的时间轴移动来取样。

[0144] 以上已经描述了内容类型和上下文类型两者。在本申请的实施方式中，调整单元300可以基于各种内容类型中的至少一个内容类型和/或各种上下文类型中的至少一个上

下文类型来调整音频改善装置的至少一个参数。因此，如图3所示，在图1所示的实施方式的变形中，音频分类器200 可以包括音频内容分类器202或者音频上下文分类器204，或者两者。

[0145] 以上已经提到了基于不同标准(例如针对上下文类型)的不同音频类型，也提到了基于不同层次级别(例如针对内容类型)的不同音频类型。但是，所述标准和所述层次级别都是为了这里描述的方便而显然并非限定。换言之，在本申请中，上述的任何两个或者更多个音频类型可以由音频分类器200同时识别，并且由调整单元300同时考虑，如后文所要描述的。换言之，不同层次级别中的所有音频类型可以是并列的，或者在同一级别中。

[0146] 1.2音频类型的置信度值和分类器的架构

[0147] 音频分类器200可以输出硬判决结果，或者调整单元300可以将音频分类器200的结果当作是硬判决结果。即使是对于硬判决，也可以将多个音频类型分配到音频片段。例如，音频片段可以被标记为“语音”和“短期音乐”两者，因为其可以是语音和短期音乐的混合信号。所获得的标签可以被直接用于操纵音频改善装置400。简单的示例是当出现语音时启用对话增强器402而当不存在语音时关闭对话增强器402。但是，如果没有仔细的平滑方案(将在稍后论述)，该硬判决方法可能在从一个音频类型到另一个音频类型的转换点处引入一些不自然的声音。

[0148] 为了具有更大的灵活性以及能以连续的方式来调节音频改善装置的参数，可以估算每个目标音频类型的置信度值(软判决)。置信度值代表待识别音频内容和目标音频类型之间的匹配水平，其值从0到1。

[0149] 如前所述，许多分类技术可以直接输出置信度值。也可以根据各种方法来计算置信度值，这些方法可以被视为分类器的一部分。例如，如果通过一些概率建模技术例如高斯混合模型(Gaussian Mixture Models, GMM)来训练音频模型，则后验概率可以被用于表示置信度值，如：

[0150] $$p(c_i \mid x) = \frac{p(x \mid c_i)}{\sum_{i=1}^{N} p(x \mid c_i)} \qquad (1)$$

[0151] 其中，x是一个音频片段，c_i是目标音频类型，N是目标音频类型的数量，$p(x|c_i)$是音频片段x属于音频类型c_i的可能性，而$p(c_i|x)$是相应的后验概率。

[0152] 另一方面，如果通过一些有识别力的方法如支持向量机(Support Vector Machine，SVM)和adaBoost来训练音频模式，则根据模型的对照只能获得得分(实际值)。在这些情况下，通常使用S形函数(sigmoid function)来将所获得的得分(理论上从-∞到∞)映射到期望的置信度(从 0到1)：

[0153] $$conf = \frac{1}{1 + e^{Ay+B}} \qquad (2)$$

[0154] 其中，y是来自SVM或者adaBoost的输出得分，A和B是需要通过使用一些众所周知的技术从训练数据集中估算出来的两个参数。

[0155] 对于本申请的一些实施方式，调整单元300可以使用多于两个的内容类型和/或多于两个的上下文类型。那么，音频内容分类器202就需要识别多于两个的内容类型，并且/或

者音频上下文分类器204需要识别多于两个的上下文类型。在这种情形中，音频内容分类器202或者音频上下文分类器204可以是以某种架构组织的一组分类器。

[0156] 例如，如果调整单元300需要所有的四种上下文类型：电影类媒体、长期音乐、游戏和VoIP，则音频上下文分类器204可以具有以下不同的架构：

[0157] 首先，音频上下文分类器204可以包括：如图4所示的那样组织的6 个一对一的二元分类器(每个分类器将一个目标音频类型与另一个目标音频类型进行鉴别)；如图5所示的那样组织的3个“一对其他” (one-to-others)的二元分类器(每个分类器将一个目标音频类型与其他目标音频类型进行鉴别)；以及如图6所示的那样组织的4个“一对其他”分类器。还有其他的架构例如决策有向无环图(Decision Directed Acyclic Graph，DDAG)架构。注意，在图4至图6以及以下的相应描述中，为了简洁起见使用“电影(movie)”而不是“电影类媒体”。

[0158] 每个二元分类器将给出置信度得分H(x)作为其输出(x代表音频片段)。在获得每个二元分类器的输出之后，需要将其映射到所识别的上下文类型的最终的置信度值。

[0159] 通常，假设音频信号要被分类到M个上下文类型中(M是正整数)。传统的一对一的架构构造出各自由来自两个类别的数据训练的M(M-1)/2 个分类器，然后每个一对一分类器投出其所倾向的类别的一票，并且最终结果是在M(M-1)/2个分类器的分类中得票最多的类别。与传统的一对一架构相比较，图4中的层次架构也需要构造M(M-1)/2个分类器。但是，测试迭代可以缩短到M-1次，因为片段x将在每个层级级别中被判定是/ 不是在相应类别中，并且整体的级别数是M-1。可以根据二元分类置信度 $H_k(x)$ 来计算针对各种上下文类型的最终的置信度值，例如(k＝1,2,…6，代表不同的上下文类型)：

[0160] $$C_{MOVIE} = (1 - H_1(x)) \cdot (1 - H_3(x)) \cdot (1 - H_6(x))$$

[0161] $$C_{VOIP} = H_1(x) \cdot H_2(x) \cdot H_4(x)$$

[0162]

$$C_{MUSIC} = H_1(x) \cdot (1 - H_2(x)) \cdot (1 - H_5(x)) + H_3(x) \cdot (1 - H_1(x)) \cdot (1 - H_5(x)) + H_6(x) \cdot (1 - H_1(x)) \cdot (1 - H_3(x))$$

[0163] $C_{GAME}＝H_1(x) \cdot H_2(x) \cdot (1-H_4(x))+H_1(x) \cdot H_3(x) \cdot (1-H_2(x))+H_3(x) \cdot H_3(x) \cdot (1-H_1(x))$

[0164] 在如图5所示的架构中，从二元分类结果$H_k(x)$到最终的置信度值的映射函数可以被定义为如下示例：

[0165] $$C_{MOVIE} = H_1(x)$$

[0166] $$C_{MUSIC} = H_2(x) \cdot (1 - H_1(x)) \quad C_{VOIP} = H_3(x) \cdot (1 - H_2(x)) \cdot (1 - H_1(x))$$

[0167] $C_{GAME}＝(1-H_3(x)) \cdot (1-H_2(x)) \cdot (1-H_1(x))$

[0168] 在如图6所示的架构中，最终的置信度值可以等于相应的二元分类结果$H_k(x)$，或者如果要求针对所有类别的置信度值的和是1，则最终的置信度值可以基于所估算的$H_k(x)$来简单地进行归一化：

[0169] $$C_{MOVIE} = H_1(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

[0170] $C_{MUSIC} = H_2(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$

[0171] $C_{VOIP} = H_3(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$

[0172] $C_{GAME} = H_4(x) / (H_1(x) + H_2(x) + H_3(x) + H_4(x))$

[0173] 具有最大置信度值的一个或者更多个类别可以被确定为最终的所识别的类。

[0174] 应该注意的是，在如图4至图6所示的架构中，不同二元分类器的顺序不一定如图所示，而是可以为其他顺序，该顺序可以根据各种应用的不同需求通过手工分配或者自动学习来选择。

[0175] 以上描述针对音频上下文分类器204。对于音频内容分类器202来说，情况类似。

[0176] 可替代地，不论是音频内容分类器202还是音频上下文分类器204都可以被实现为同时识别所有内容类型/上下文类型并且同时给出相应置信度值的单个分类器。有许多用于做到这一点的现有技术。

[0177] 使用置信度值，音频分类器200的输出可以用矢量来表示，每个维度代表每个目标音频类型的置信度值。例如，如果目标音频类型依次是语音、短期音乐、噪声、背景，则示例输出结果可以是(0.9,05,0.0,0.0)，表示其 90％地确定该音频内容是语音，50％地确定该音频是音乐。注意，输出的向量的所有维度的和不一定是1(例如，图6的结果不一定是归一化的)，表示该音频信号可能是语音和短期音乐的混合信号。

[0178] 在后续的第6部分和第7部分中，将详细论述音频上下文分类和音频内容分类的新颖的实现方式。

[0179] 1.3音频类型的置信度值的平滑

[0180] 可选地，在每个音频片段已经被分类到预定义的音频类型中之后，附加的步骤是使分类结果沿着时间轴平滑，以避免从一个类型到另一个类型的急剧跃变，并且使音频改善装置中的参数的估算更平滑。例如，一个长的选段除了只有一个片段被分类为VoIP以外均被分类为电影类媒体，因此可以通过平滑处理将突兀的VoIP决策修改为电影类媒体。

[0181] 因此，在如图7所示的实施方式的变型中，还设置了类型平滑单元712，用于针对每个音频类型，对当次音频信号的置信度值进行平滑。

[0182] 常用的平滑方法基于加权平均，例如计算当前的实际置信度值与上一次的经平滑的置信度值的加权和，如下：

[0183] $\mathrm{smoothConf}(t) = \beta \cdot \mathrm{smoothConf}(t-1) + (1-\beta) \cdot \mathrm{conf}(t)$ (3)

[0184] 其中，t代表当次(当前的音频片段)，t-1代表上一次(上一个音频片段)，β是权重，conf和smoothConf分别是平滑之前的置信度值和平滑之后的置信度值。

[0185] 从置信度值的角度来看，来自分类器的硬判决的结果也可以用置信度值来表示，其值为0或1。也就是说，如果某目标音频类型被选择分配给某音频片段，则相应的置信度是1；否则，置信度是0。因此，即使音频分类器200不给出置信度值而只给出关于音频类型的硬判决，也可以通过类型平滑单元712的平滑操作来进行调整单元300的连续调整。

[0186] 通过针对不同情况使用不同的平滑权重，平滑算法可以是“不对称的”。例如，用于计算加权和的权重可以基于音频信号的音频类型的置信度值来自适应地改变。当前片段的置信度值越大，则其权重越大。

[0187] 从另一个角度来看，用于计算加权和的权重可以基于不同的从一个音频类型到另

一个音频类型的转换对来自适应地改变，尤其是当基于由音频分类器200所识别的多个内容类型，而不是基于单个内容类型存在或不存在来调整一个或多个音频改善装置时。例如，对于从在某个上下文中较频繁出现的音频类型到在该上下文中不那么频繁出现的另一个音频类型的转换，可以对后者的置信度值进行平滑，以使得其不会太快地增加，因为其可能只是个偶然的干扰。

[0188] 另一个因素是变化(增加或者减少)趋势，包括变化速率。假设更关心音频类型出现时(也就是说，当其置信度值增加时)的延迟，可以以如下方式来设计平滑算法：

[0189]

$$smoothConf(t)=\begin{cases} conf(t) & conf(t)\geq smoothConf(t-1) \\ \beta\cdot smoothConf(t-1)+(1-\beta)\cdot conf(t) & 否则 \end{cases} \quad (4)$$

[0190] 以上公式使得当置信度值增加时经平滑的置信度值能够快速响应当前状态，而当置信度值减小时经平滑的置信度值能够缓慢地消失。可以以相似的方式容易地设计平滑函数的变型。例如，公式(4)可以被修改以使得当conf(t)>＝smoothConf(t-1)时conf(t)的权重变得更大。实际上，在公式(4)中，可以认为β＝0并且conf(t)的权重变为最大，即1。

[0191] 换一个视角来看，考虑某个音频类型的改变趋势只是考虑音频类型的不同转换对的具体示例。例如，类型A的置信度值的增加可以被看做是从非A转换到A，而类型A的置信度值的减小可以被看做是从A转换到非 A。

[0192] 1.4参数调整

[0193] 调整单元300被设计用于基于从音频分类器200获得的结果来估算或者调整音频改善装置400的适当参数。通过使用内容类型或者上下文类型，或者使用内容类型和上下文类型两者用于联合决策，可以针对不同的音频改善装置来设计不同的调整算法。例如，使用上下文类型信息例如电影类媒体和长期音乐，可以自动地选择如前所述的预置并且将其施加到相应内容上。使用可获得的内容类型信息，可以以更精准的方式调节每个音频改善装置的参数，如后续部分将要介绍的。在调整单元300中还可以联合地使用内容类型信息和上下文信息，以平衡长期信息和短期信息。用于特定音频改善装置的特定调整算法可以被当作独立的调整单元，或者不同的调整算法可以被共同当作是联合的调整单元。

[0194] 也就是说，调整单元300可以被配置成基于至少一个内容类型的置信度值和/或至少一个上下文类型的置信度值来调整音频改善装置的至少一个参数。对于特定的音频改善装置，一些音频类型是信息性的而一些音频类型是干扰性的。因此，特定的音频改善装置的参数可以与信息性的音频类型的置信度值正相关或者与干扰性的音频类型的置信度值负相关。这里，“正相关”意指参数以线性的方式或者以非线性的方式随着音频类型的置信度值的增大或减小而增大或减小。“负相关”意指参数以线性的方式或者以非线性的方式随着音频类型的置信度值的减小或增大而分别增大或减小。

[0195] 这里，通过正相关或者负相关将置信度值的减小和增加直接“传递”到待调整的参数。在数学上，这种相关性或者“传递”可以具体表现为线性比例或者反比例，加运算或者减运算(加法或者减法)，乘运算或者除运算或者非线性函数。所有这些形式的关联可以被称为“传递函数”。为了确定置信度值的增加或减少，还可以将当前置信度值或者其数学变形与上一置信度值或者多个历史置信度值或其数学变形进行比较。在本申请的语境中，术语“比较”是指通过减法运算的比较或者通过除法运算的比较。通过确定差是否大于0或者比

率是否大于1可以确定是增加还是减少。

[0196] 在具体的实现中，可以通过适当的算法(例如传递函数)将参数与置信度的值或者其比率或差直接进行关联，因此“外部观察者”不一定需要明确地知道具体的置信度值和/或具体的参数是增加了还是减少了。将在接下来的关于具体音频改善装置的第2部分至第5部分中给出一些具体的示例。

[0197] 如前面的小节所述，对于同一音频片段，分类器200可以识别出具有相应置信度值的多个音频类型，其置信度值可以不一定合计为1，因为该音频片段可以同时包括多个成分，例如音乐和语音和背景声音。在这样的情形中，应该在不同音频类型之间平衡音频改善装置的参数。例如，调整单元300可以被配置为通过基于至少一个音频类型的重要性对至少一个音频类型的置信度值进行加权来考虑多个音频类型中的至少一些音频类型。特定音频类型越重要，则参数被其影响的程度越大。

[0198] 权重还可以反映出音频类型的信息性影响和干扰性影响。例如，对于干扰性音频类型，可以给它负的权重。将在接下来的关于具体音频改善装置的第2部分至第5部分中给出一些具体的示例。

[0199] 请注意在本申请的语境中“权重”具有比多项式中的系数更广泛的含义。除了多项式中的系数的形式，其还可以是指数或者幂的形式。当是多项式中的系数时，权重系数可以是归一化的或者可以不是归一化的。简言之，权重只代表被加权的对象对于待调整的参数具有多少影响。

[0200] 在一些其他实施方式中，针对在同一音频片段中所包含的多个音频类型，其置信度值可以通过归一化被转换为权重，然后，可以通过计算针对每个音频类型预定义的并且被基于置信度值的权重加权的参数预置值的和来确定最终的参数。也就是说，调整单元300可以被配置成通过基于置信度值将多个音频类型的作用进行加权来考虑多个音频类型。

[0201] 作为加权的具体示例，调整单元被配置成基于置信度值来考虑至少一个主导的音频类型。对于具有过低置信度值(小于阈值)的音频类型，其可以不被考虑。其等同于将置信度值小于阈值的其他音频类型的权重设置为零。将在接下来的关于具体音频改善装置的第2部分至第5部分中给出一些具体的示例。

[0202] 可以一起考虑内容类型和上下文类型。在一个实施方式中，内容类型和上下文类型可以被当作是在同一级别并且其置信度值可以具有相应的权重。在另一个实施方式中，正如其命名所示出的，“上下文类型”是“内容类型”所处的上下文或者环境，因此可以配置调整单元200以使得取决于音频信号的上下文类型而给不同上下文类型的音频信号中的内容类型分配不同的权重。一般来说，任何音频类型可以构成另一个音频类型的上下文，因此调整单元200可以被配置成根据另一个音频类型的置信度值来更改一个音频类型的权重。将在接下来的关于具体音频改善装置的第2部分至第5部分中给出一些具体的示例。

[0203] 在本申请的语境中，“参数”具有比其字面含义更广泛的含义。除了具有单一值的参数，其还可以指如前所述的预置，包括不同的参数的集合、由不同参数构成的向量、或者模式(profile)。具体地，在接下来的第2部分至第5部分中将论述以下参数，但是本申请不限于此：对话增强的级别、用于确定要对话增强的频带的阈值、背景声级、环绕声增强量、用于环绕声虚拟器的起始频率、音量校平器的动态增益或者动态增益的范围、表示音频信号是新的可察觉音频事件的程度的参数、均衡级别、均衡模式和谱平衡预置。

[0204] 1.5参数平滑

[0205] 在小节1.3中，已经论述了对音频类型的置信度值进行平滑以避免其剧烈变化，并且因此避免音频改善装置的参数的剧烈变化。其他的方式也是可以的。一种方式是将基于音频类型调整的参数进行平滑，将在本小节中论述这种方式；另一种方式是配置音频分类器和/或调整单元以延迟音频分类器的结果的变化，将在小节1.6中论述这种方式。

[0206] 在一个实施方式中，参数还可以被进一步平滑以避免可能在转换点处引入可听见的畸变的快速变化，如：

[0207] $$\widetilde{L}(t)=\tau\widetilde{L}(t-1)+(1-\tau)L(t) \qquad (3')$$

[0208] 其中，$\widetilde{L}(t)$是经平滑的参数，L(t)是未经平滑的参数，τ是代表时间常数的系数，t是当次，而t-1是上一次。

[0209] 也就是说，如图8所示，音频处理设备可以包括参数平滑单元814，用于：针对由调整单元300所调整的音频改善装置(例如对话增强器402、环绕声虚拟器404、音量校平器406和均衡器408中的至少一个)的参数，通过计算当次由调整单元所确定的参数值与上一次的经平滑的参数值的加权和来对当次由调整单元300确定的参数进行平滑。

[0210] 时间常数τ可以是基于应用的具体要求和/或基于音频改善装置400的实现方式的固定值。其也可以基于音频类型，尤其是基于不同的从一个音频类型到另一个音频类型的转换类型——例如从音乐到语音以及从语音到音乐——而自适应地改变。

[0211] 以均衡器作为示例(进一步的细节可以参考第5部分)。均衡器适合于应用到音乐内容而不适合应用到语音内容。因此，为了对均衡的级别进行平滑，当音频信号从音乐转换到语音时，时间常数可以相对较小，以使得更快地对语音内容应用更小的均衡级别。另一方面，为了避免在转换点处产生可听到的畸变，针对从语音到音乐的转换的时间常数可以相对较大。

[0212] 为了估计转换类型(例如，从语音到音乐，或者从音乐到语音)，可以直接使用内容分类结果。也就是说，将音频内容分类到或音乐或语音使其直截了当地得到转换类型。为了以更连续的方式估计所述转换，也可以依赖于所估计的未经平滑的均衡级别，而不是直接比较音频类型的硬判决。总体思想是：如果未经平滑的均衡级别是增加的，其表示从语音到音乐(或者更像音乐)的转换；否则，其更可能是从音乐到语音(或者更像语音)的转换。通过区分不同的转换类型，可以相应地设置时间常数，一个示例是：

[0213] $$\tau(t)=\begin{cases}\tau_1 & L(t)\geq L(t-1)\\ \tau_2 & L(t)<L(t-1)\end{cases} \qquad (4')$$

[0214] 其中τ(t)是取决于内容的随时间变化的时间常数，τ1和τ2是两个预置的时间常数值，通常满足τ1>τ2。直观地，以上函数表示当均衡级别增加时进行相对慢的转换，以及当均衡级别减小时进行相对快的转换，但是本申请不限于此。另外，参数不限于均衡级别，而可以是其他参数。也就是说，可以配置参数平滑单元814以使得用于计算加权和的权重基于由调整单元300所确定的参数的增加趋势或者减小趋势而自适应地改变。

[0215] 1.6音频类型的转换

[0216] 参照图9和图10，将描述为了避免音频类型的急剧变化，并且因此避免音频改善装

置的参数的急剧变化的另一个方案。

[0217] 如图9所示，音频处理设备100还可以包括计时器916，用于测量音频分类器200连续输出同一新音频类型的持续时间，其中，调整单元300 可以被配置成继续使用当前的音频类型，直到新的音频类型的持续时间的长度达到阈值为止。

[0218] 换言之，引入了如图10所示的观察期(或者维持期)。使用观察期(与持续时间的长度的阈值对应)，进一步在一段连续的时间中监测音频类型的变化，以确认音频类型是否确实已经改变，然后才能在调整单元300实际使用新的音频类型。

[0219] 如图10所示，箭头(1)示出了当前状态是类型A并且音频分类器 200的结果没有变化的情形。

[0220] 如果当前状态是类型A并且音频分类器200的结果变成类型B，则计时器916开始计时，或者如图10所示，处理进行到观察期(箭头(2))，并且设置逗留计数cnt的初始值，其表示观察期长度(等于阈值)。

[0221] 然后，如果音频分类器200连续地输出类型B，则cnt连续地减小(箭头(3))直到cnt等于0(也就是说，新类型B的持续时间的长度达到阈值)，则调整单元300可以使用新的音频类型B(箭头(4))，或者换言之，直到现在才可以认为音频类型真的变化为类型B。

[0222] 否则，如果在cnt变为0之前(持续时间的长度达到阈值之前)，音频分类器的输出200变回到原来的类型A，则该观察期结束，并且调整单元 300仍然使用原来的类型A(箭头(5))。

[0223] 从类型B到类型A的变化可以与上述处理类似。

[0224] 在以上处理中，可以基于应用需求来设置阈值(或者逗留计数)。该阈值可以被预定义为固定值。该阈值也可以被自适应地设置。在一个变型中，针对不同的从一个音频类型到另一个音频类型的转换对，该阈值不同。例如，当从类型A变化到类型B时，该阈值可以是第一值；而当从类型B 变化到类型A时，该阈值可以是第二值。

[0225] 在另一个变型中，逗留计数(阈值)可以与新的音频类型的置信度值负相关。总体思想是：如果置信度表现出两个类型之间的混淆(例如，当置信度值只在0.5左右)，则观察期需要长；否则，观察期可以相对较短。根据此指导方针，可以通过以下公式来设置示例性的逗留计数：

[0226] $HangCnt = C \cdot |0.5 - Conf| + D$

[0227] 其中，HangCnt是逗留期或者阈值，C和D是能够基于应用需求而设置的两个参数，通常C是负的值而D是正的值。

[0228] 顺便提及，上面将计时器916(以及因此上述的转换处理)描述成作为音频处理设备的一部分但是在音频分类器200的外部。在一些其他的实施方式中，正如小节7.3中所描述的，计时器916可以被视为音频分类器 200的一部分。

[0229] 1.7实施方式和应用场景的组合

[0230] 以上所述的所有实施方式及其变型可以以其任意的组合来实现，并且在不同的部分/实施方式中所提到但是具有相同或者类似功能的任何组件可以作为同一组件或者单独的组件来实现。

[0231] 特别地，当在上文中描述实施方式及其变型时，省略了具有与前面的实施方式或者变型中已经描述的组件的附图标记类似的附图标记的组件，而只是描述了不同的组件。

实际上，这些不同的组件可以与其他实施方式或者变型的组件进行合并，也可以独自构成单独的解决方案。例如，参照图1至图10所描述的任何两个或者更多个解决方案可以互相合并。作为最完整的解决方案，音频处理设备可以包括音频内容分类器202和音频上下文分类器204两者，以及类型平滑单元712、参数平滑单元814和计时器916。

[0232] 如前面所提到的，音频改善装置400可以包括对话增强器402、环绕声虚拟器404、音量校平器406和均衡器408。音频处理设备100可以包括它们中的任何一个或更多个，以及适于它们的调整单元300。当涉及多个音频改善装置400时，调整单元300可以被视为包括专用于相应的音频改善装置400的多个子单元300A至300D(图15、图18、图20和图22)，或者仍被视为一个联合的调整单元。当专用于音频改善装置时，调整单元 300连同音频分类器200，以及其他可能的组件可以被视为特定音频改善装置的控制器，其将在接下来的第2部分至第5部分中详细论述。

[0233] 此外，音频改善装置400不限于已经提到的示例，而是可以包括任何其他的音频改善装置。

[0234] 另外，任何已经论述的解决方案或者其任何组合也可以与本公开内容的其他部分中所描述或者所暗示的实施方式组合。特别地，在第6部分和第7部分中将要论述的音频分类器的实施方式可以用在音频处理设备中。

[0235] 1.8音频处理方法

[0236] 在描述以上实施方式中的音频处理设备的过程中，显然也公开了一些过程和方法。在下文中，在不重复已经论述的细节的情况下给出这些方法的概要，但是应该注意的是，尽管在描述音频处理设备的过程中公开了方法，但是这些方法不一定采用所描述的组件或者不一定由这些组件来执行。例如，音频处理设备的实施方式可以部分地或者完全地以硬件和/或固件来实现，而下述的音频处理方法可以完全地由计算机可执行程序来实现，尽管这些方法也可以采用音频处理设备的硬件和/或固件。

[0237] 以下将参照图11至图14来描述这些方法。请注意，对应于音频信号的流属性，当实时地实现所述方法时重复地进行各种操作，并且不同的操作不一定针对同一音频片段。

[0238] 如图11所示的实施方式中，提供了音频处理方法。首先，将待处理的音频信号实时地分类到至少一个音频类型中(操作1102)。基于至少一个音频类型的置信度值，可以连续地调整至少一个用于音频改善的参数 (操作1104)。音频改善可以是对话增强(操作1106)、环绕声虚拟(操作 1108)、音量校平(1110)和/或均衡(操作1112)。对应地，该至少一个参数可以包括用于对话增强处理、环绕声虚拟处理、音量校平处理和均衡处理中的至少一个处理的至少一个参数。

[0239] 这里，“实时地”和“连续地”意指音频类型(从而所述参数)将根据音频信号的具体内容而实时地变化，并且“连续地”还意指调整是基于置信度值的连续调整，而不是突变的或者离散的调整。

[0240] 音频类型可以包括内容类型和/或上下文类型。相应地，调整操作1104 可以被配置成基于至少一个内容类型的置信度值和至少一个上下文类型的置信度值来调整至少一个参数。内容类型还可以包括短期音乐、语音、背景声音和噪声中的至少一个内容类型。上下文类型还可以包括长期音乐、电影类媒体、游戏和VoIP中的至少一个上下文类型。

[0241] 也可以提出其他的上下文类型方案，例如包括VoIP和非-VoIP的与 VoIP相关的上

下文类型，以及包括高质量音频或者低质量音频的音频质量类型。

[0242] 短期音乐还可以根据不同的标准而被进一步分为各种子类型。取决于主导源的存在，短期音乐可以包括没有主导源的音乐和有主导源的音乐。另外，短期音乐可以包括至少一个基于风格的类型，或者至少一个基于乐器的类型，或者至少一个基于音乐的节奏、速度、音色和/或任何其他音乐特征而分类的音乐类型。

[0243] 当既识别内容类型又识别上下文类型时，可以通过内容类型所处的上下文类型来确定内容类型的重要度。也就是说，取决于音频内容的上下文类型，给不同上下文类型的音频信号中的内容类型分配不同的权重。更一般地，一个音频类型可以影响另一个音频类型，或者一个音频类型可以是另一个音频类型的前提。因此，调整操作1104可以被配置为根据另一个音频类型的置信度值来更改一个音频类型的权重。

[0244] 当音频信号同时(也就是针对同一音频片段)被分类到多个音频类型中时，为了调整参数以改善该音频片段，调整操作1104可以考虑所识别的音频类型中的一些或者全部。例如，调整操作1104可以被配置为基于至少一个音频类型的重要性来对至少一个音频类型的置信度值进行加权。或者，调整操作1104可以被配置成通过基于音频类型的置信度值对其进行加权来考虑音频类型中的至少一些音频类型。在特殊的情况中，调整操作1104可以被配置为基于置信度值来考虑至少一个主导音频类型。

[0245] 为了避免结果的急剧变化，可以引入平滑方案。

[0246] 可以对经调整的参数值进行平滑(图12中的操作1214)。例如，当次由调整操作1104确定的参数值可以被替换为当次由调整操作所确定的参数值与上一次经平滑的参数值的加权和。因此，通过迭代的平滑操作，在时间轴上平滑了参数值。

[0247] 用于计算加权和的权重可以基于音频信号的音频类型，或者基于不同的从一个音频类型到另一个音频类型的转换对，而自适应地变化。或者，用于计算加权和的权重基于由调整操作确定的参数值的增加趋势或者减小趋势来自适应地变化。

[0248] 图13中示出了另一个平滑方案。也就是说，该方法还可以包括：针对每个音频类型，通过计算当前的实际置信度值与上一次的经平滑的置信度值的加权和，来对当次音频信号的置信度值进行平滑(操作1303)。与参数平滑操作1214类似地，用于计算加权和的权重可以基于音频信号的音频类型的置信度值，或者基于不同的从一个音频类型到另一个音频类型的转换对，而自适应地变化。

[0249] 另一个平滑方案是用于即使音频分类操作1102的输出变化了但是延迟从一个音频类型到另一个音频类型的转换的缓冲机制。也就是说，调整操作1104不立即使用新的音频类型，而是等待音频分类操作1102的输出的稳定。

[0250] 具体地，该方法可以包括：对分类操作连续地输出同一新音频类型的持续时间进行测量(图14中的操作1403)，其中，调整操作1104被配置为继续使用当前的音频类型(操作14035中的“N”和操作11041)直到新的音频类型的持续时间的长度达到阈值(操作14035中的“Y”和操作 11042)。具体地，当来自音频分类操作1102的音频类型输出相对于音频参数调整操作1104中所使用的当前音频类型改变时(操作14031中的“Y”)，则计时开始(操作14032)。如果音频分类操作1102继续输出该新音频类型，也就是说，如果在操作14031中的判断继续为“Y”，则计时继续(操作14032)。最终当该新的音频类型的持续时间达到阈值(操作14035中的“Y”)时，调整操作1104使用该新音频类型(操作11042)，并且计时复位(操作

14034)，用于为下一次音频类型的转换做准备。在达到阈值之前 (操作14035中的“N”)，调整操作1104继续使用当前的音频类型(操作 11041)。

[0251] 这里，计时可以通过计时器的机制来实现(向上计数或者向下计数)。如果在计时开始之后但是在达到阈值之前，音频分类操作1104的输出变回到当前的调整操作1104中所使用的当前音频类型，则应该视为没有相对于调整操作1104中所使用的当前音频类型的变化(操作14031中的“N”)。但是当前的分类结果(对应于音频信号中的待分类的当前音频片段)相对于音频分类操作1102的前一输出(对应于音频信号中的待分类的前一个音频片段)变化了(操作14033中的“Y”)，因此，计时复位(操作14034)，直到下一次改变时(操作14031中的“Y”)开始计时。当然，如果音频分类操作1102的分类结果既没有相对于音频参数调整操作1104 中所使用的当前音频类型变化(操作14031中的“N”)，也没有相对于前一分类变化(操作14033中的“N”)，则表示音频分类处于稳定的状态并且继续使用当前的音频类型。

[0252] 这里所使用的阈值也可以针对不同的从一个音频类型到另一个音频类型的转换对而不同，因为当状态不是很稳定时，通常可能更希望音频改善装置处于其默认状态而不是处于其他状态。另一方面，如果该新的音频类型的置信度值相对较高，则转换到新的音频类型更安全。因此，该阈值可以与新的音频类型的置信度值负相关。置信度越高，则阈值越低，意味着音频类型可以更快地转换到新的音频类型。

[0253] 与音频处理设备的实施方式类似地，一方面，音频处理方法的实施方式与实施方式的变型的任何组合都是可行的；另一方面，音频处理方法的实施方式和实施方式的变型的每个方面也都可以是单独的解决方案。特别地，在所有的音频处理方法中，可以使用如第6部分和第7部分中所论述的音频分类方法。

[0254] 第2部分：对话增强器控制器和控制方法

[0255] 音频改善装置的一个示例是对话增强器(DE)，其旨在连续地监测回放的音频，检测对话的存在，以及增强该对话以提高其清晰度和可理解性 (使该对话更容易被听到和被理解)，尤其是对于听力衰退的年长者。除了检测是否存在对话之外，如果对话存在的话还检测对于可理解性来说最重要的频率，然后相应地增强该频率(使用动态谱再平衡)。在H.Muesch 的公开号为WO 2008/106036A2的“Speech Enhancement in Entertainment Audio”中给出了对话增强方法的一个示例，该文献的全部内容通过引用合并到本文中。

[0256] 通常对于电影类媒体的内容启用对对话增强器的普通手动配置，而对于音乐内容则禁用，因为对话增强对音乐信号可能过多地错误触发。

[0257] 在可获得音频类型信息的情况下，可以基于所识别的音频类型的置信度值来调节对话增强的级别和其他参数。作为之前论述的音频处理设备和方法的具体示例，对话增强器可以使用第1部分中所论述的所有实施方式以及这些实施方式的任意组合。具体地，在控制对话增强器的情况下，如图1至图10所示的音频处理设备100的音频分类器200和调整单元300 可以组成如图15所示的对话增强器控制器1500。在这个实施方式中，因为调整单元是专用于对话增强器的，所以其可以被称为300A。并且，如前一部分所论述的，音频分类器200可以包括音频内容分类器202和音频上下文分类器204中的至少一个，并且对话增强器控制器1500还可以包括类型平滑单元712、参数平滑单元814和计时器916中的至少一个。

[0258] 因此，在这个部分中，将不重复在前一部分已经描述的这些内容，而只给出其一些具体的示例。

[0259] 对于对话增强器，可调整的参数包括但不限于：对话增强的级别、背景声级和用于确定待增强的频带的阈值。参见H.Muesch的公开号为WO 2008/106036 A2的“Speech Enhancement in Entertainment Audio”，其全部内容通过引用合并到本文中。

[0260] 2.1对话增强的级别

[0261] 当涉及对话增强的级别时，调整单元300A可以被配置为使对话增强器的对话增强的级别与语音的置信度值正相关。附加地或者可替代地，该级别可以与其他内容类型的置信度值负相关。因此，对话增强的级别可以被设置为与语音的置信度成比例(线性的或者非线性的)，以使得在非语音信号例如音乐和背景声音(音效)中，对话增强不那么有效。

[0262] 对于上下文类型，调整单元300A可以被配置成使对话增强器的对话增强的级别与电影类媒体和/或VoIP的置信度值正相关，并且/或者使对话增强器的对话增强的级别与长期音乐和/或游戏的置信度值负相关。例如，对话增强的级别可以被设定为与电影类媒体的置信度值成比例(线性的或者非线性的)。当电影类媒体的置信度值是0时(例如，在音乐内容中)，对话增强的级别也是0，其相当于停用对话增强。

[0263] 如在前一部分所描述的，可以联合考虑内容类型和上下文类型。

[0264] 2.2用于确定待增强的频带的阈值

[0265] 在对话增强器的工作期间，针对每个频带存在用于确定该频带是否要被增强的阈值(通常是能量阈值或者响度阈值)，也就是说，将对相应的能量/响度阈值以上的这些频带进行增强。为了调整阈值，调整单元300A 可以被配置为使该阈值与短期音乐和/或噪声和/或背景声音的置信度值正相关，并且/或者使阈值与语音的置信度值负相关。例如，如果语音置信度高(意味着更可靠的语音检测)，则可以降低阈值，以使得更多的频带能够被增强；另一方面，当音乐的置信度值高时，可以升高阈值以使得更少的频带被增强(因此有更少的畸变)。

[0266] 2.3对背景声级的调整

[0267] 如图15所示，对话增强器中的另一个组件是最小量追踪单元4022，其用于估计音频信号中的背景声级(背景声级用于SNR(信噪比)的估计，和小节2.2中所提到的频带阈值估计)。其还可以基于音频内容类型的置信度值来调节。例如，如果语音的置信度高，则最小量追踪单元可以更确信地将背景声级设定到当前的最小量。如果音乐的置信度高，则背景声级被设定到比当前的最小量高一点，或者以另一种方式，背景声级被设定成当前最小量与当前帧的能量的加权平均值，其中当前最小量被施以大权重。如果噪声和背景的置信度高，则背景声级可以被设定得比当前最小量的值高很多，或者以另一种方式，背景声级被设定成当前最小量与当前帧的能量的加权平均值，其中当前最小量被施以小权重。

[0268] 因此，调整单元300A可以被配置成对由最小量追踪单元估计的背景声级施加一个调整量，其中，调整单元还被配置为使该调整量与短期音乐和/或噪声和/或背景声音的置信度值正相关，并且/或者使该调整量与语音的置信度值负相关。在变型中，调整单元300A可以被配置为使该调整量与噪声和/或背景声音的置信度值比与短期音乐更加正相关。

[0269] 2.4实施方式和应用场景的组合

[0270] 与第1部分类似，以上所述的所有实施方式和实施方式的变型可以以其任意的组合来实现，并且在不同的部分/实施方式中所提到的但是具有相同或者类似功能的任何组件可以作为同一组件或者单独的组件来实现。

[0271] 例如，在小节2.1至小节2.3中描述的任何两个或更多个解决方案可以彼此组合。并且这些组合还可以与在第1部分中描述的和暗示的以及在稍后其他部分中将要描述的任何实施方式进行组合。特别地，很多公式实际上可应用于每种音频改善装置或方法，但是不一定在本公开内容的每个部分中都引用或者论述了这些公式。在这种情形中，本公开的各个部分可以相互参考，以将一个部分中论述的特定公式应用到另一个部分中，只是需要根据具体应用的具体要求，适当地调整相关参数、系数、幂(指数) 和权重。

[0272] 2.5对话增强器控制方法

[0273] 与第1部分类似，在描述上文实施方式中的对话增强控制器的过程中，显然也公开了一些过程和方法。在下文中，在不重复已经论述的细节的情况下给出这些方法的概要。

[0274] 首先，在第1部分中所论述的音频处理方法的实施方式可以用于对话增强器，对话增强器的参数是要由音频处理方法调整的目标之一。根据这一点，音频处理方法也是对话增强器控制方法。

[0275] 在本小节中，将只论述特定于对话增强器的控制的那些方面。关于控制方法的一般方面，可以参考第1部分。

[0276] 根据一个实施方式，音频处理方法还可以包括对话增强处理，并且调整操作1104包括使对话增强的级别与电影类媒体和/或VoIP的置信度值正相关，并且/或者使对话增强的级别与长期音乐和/或游戏的置信度值负相关。也就是说，对话增强主要针对上下文类型为电影类媒体或者VoIP的音频信号。

[0277] 更具体地，调整操作1104可以包括使对话增强器的对话增强的级别与语音的置信度值正相关。

[0278] 本申请还可以在对话增强处理中调整待增强的频带。如图16所示，根据本申请，可以基于所识别的音频类型(操作1602)的置信度值来调整阈值(通常是能量或者响度)，该阈值用于确定相应的频带是否要被增强。然后，在对话增强器中，基于所调整的阈值，选择(操作1604)并增强(操作1606)相应阈值以上的频带。

[0279] 特别地，调整操作1104可以包括使阈值与短期音乐和/或噪声和/或背景声音的置信度值正相关，并且/或者使阈值与语音的置信度值负相关。

[0280] 音频处理方法(尤其是对话增强处理)通常还包括估计音频信号中的背景声级，通常由最小量追踪单元4022来实现该处理，最小量追踪单元4022在对话增强器402中实现，并且用于SNR估计或者频带阈值估计。本申请还可以用于调整背景声级。在这样的情形中，在估计背景声级之后 (操作1702)，首先基于音频类型的置信度值来调整背景声级(操作1704)，然后将背景声级用于SNR估计和/或频带阈值估计(操作1706)。特别地，调整操作1104可以被配置成对所估计的背景声级施加一个调整量，其中调整操作1104还可以被配置成使该调整量与短期音乐和/或噪声和/或背景声音正相关，并且/或者使该调整量与语音的置信度值负相关。

[0281] 更具体地，调整操作1104可以被配置成使该调整量与噪声和/或背景的置信度值比与短期音乐更加正相关。

[0282] 与音频处理设备的实施方式类似地，一方面，音频处理方法的实施方式与实施方式的变型的任何组合都是可行的的；另一方面，音频处理方法的实施方式和实施方式的变型的每个方面都可以是单独的解决方案。另外，本小节中所描述的任何两个或更多个解决

方案可以彼此组合，并且这些组合还可以与在第1部分中以及在稍后将要描述的其他部分中所描述的和所暗示的任何实施方式进行组合。

[0283] 第3部分：环绕声虚拟器控制器和控制方法

[0284] 环绕声虚拟器使得能够在PC的内置扬声器或者耳机中渲染出环绕声信号(例如多声道5.1和多声道7.1)。也就是说，通过立体声装置例如内置便携式电脑扬声器或者耳机，环绕声虚拟器为用户生成虚拟的环绕声效果并且提供电影的体验。在环绕声虚拟器中通常利用头部相关传递函数 (Head Related Transfer Function，HRTF)来模拟来自与多通道音频信号相关联的各种扬声器位置的声音在耳朵处的波至。

[0285] 虽然现有的环绕声虚拟器在耳机上工作良好，但是对于内置扬声器上，环绕声虚拟器对于不同的内容不一样地工作。通常，电影类媒体内容针对扬声器启用环绕声虚拟器，而音乐不这样做，因为音乐可能听起来太单薄。

[0286] 因为环绕声虚拟器的相同参数不能针对电影类媒体内容和音乐内容两者同时生成好的声像，所以需要基于内容更精确地调节参数。使用可获得的音频类型信息，尤其是音乐置信度值和语音置信度值，以及一些其他的内容类型信息和上下文信息，可以使用本申请来完成该工作。

[0287] 与第2部分类似地，作为第1部分中所论述的音频处理设备和方法的具体示例，环绕声虚拟器404可以使用第1部分中所论述的所有实施方式以及在第1部分中所公开的这些实施方式的任何组合。特别地，在控制环绕声虚拟器404的情况下，如图1至图10所示的音频处理设备100的音频分类器200和调整单元300可以组成如图18所示的环绕声虚拟器控制器1800。在这个实施方式中，因为调整单元是专用于环绕声虚拟器404的，所以其可以被称为300B。并且，与第2部分类似，音频分类器200可以包括音频内容分类器202和音频上下文分类器204中的至少一个，并且环绕声虚拟器控制器1800还可以包括类型平滑单元712、参数平滑单元814和计时器916中的至少一个。

[0288] 因此，在这个部分中，将不重复第1部分已经描述的这些内容，而只给出其一些具体示例。

[0289] 对于环绕声虚拟器，可调整的参数包括但不限于：用于环绕声虚拟器 404的起始频率和环绕声增强量。

[0290] 3.1环绕声增强量

[0291] 当涉及环绕声增强量时，调整单元300B可以被配置成使环绕声虚拟器404的环绕声增强量与噪声和/或背景和/或语音的置信度值正相关，并且/或者使环绕声增强量与短期音乐的置信度值负相关。

[0292] 具体地，为了修改环绕声虚拟器404以使音乐(内容类型)听起来是可接受的，调整单元300B的示例实现可以基于短期音乐置信度值来调节环绕声增强量，例如：

[0293] $SB \propto (1-Conf_{music})$ (5)

[0294] 其中，SB表示环绕声增强量，$Conf_{music}$是短期音乐的置信度值。

[0295] 其有助于针对音乐减弱环绕声增强，防止其听起来模糊。

[0296] 类似地，也可以利用语音置信度值，例如：

[0297] $SB \propto (1-Conf_{music}) * Conf_{speech}^{\alpha}$ (6)

[0298] 其中，$Conf_{speech}$是语音的置信度值，α是指数形式的权重系数，其范围可以是1至2。

该公式表示环绕声增强量只对纯语音(高的语音置信度且低的音乐置信度)是高的。

[0299] 或者可以只考虑语音的置信度值：

[0300] $SB \propto Conf_{speech}$ (7)

[0301] 可以以类似的方式设计各种变型。尤其是,对于噪声或者背景声音,可以构造与公式(5)至公式(7)类似的公式。此外,可以以任何组合联合考虑该四个内容类型的效果。在这样的情形下,噪声和背景声音是环境声音,所以可以更安全地具有大的增强量;假设说话人通常位于屏幕的前面,所以语音可以具有中等的增强量;而音乐使用较少的增强量。因此,调整单元300B可以被配置成使环绕声增强量与噪声和/或背景的置信度值比与语音的内容类型更加正相关。

[0302] 假设针对每个内容类型预定义了期望的增强量(即,相当于权重),也可以应用另一个可替代的公式：

[0303]

$$\hat{a} = \frac{a_{speech} \cdot Conf_{speech} + a_{music} \cdot Conf_{music} + a_{noise} \cdot Conf_{noise} + a_{bkg} \cdot Conf_{bkg}}{Conf_{speech} + Conf_{music} + Conf_{noise} + Conf_{bkg}} \quad (8)$$

[0304] 其中,$\hat{a}$是估计的增强量,带有内容类型下标的α是内容类型的期望/预定义的增强量(权重),带有内容类型下标的Conf是内容类型的置信度值 (bkg代表background sound,即背景声音)。视情况而定,a_{music}可以(但不一定)被设定为0,表示对于纯音乐(内容类型)将禁用环绕声虚拟器 404。

[0305] 从另一个角度来看,公式(8)中的带有内容类型下标的α是内容类型的期望/预定义的增强量,并且相应内容类型的置信度值被所有所识别的内容类型的置信度值的和所除的商可以被视为相应内容类型的预定义~/期望的增强量的归一化的权重。也就是说,调整单元300B可以被配置成通过基于置信度值对多个内容类型的预定义的增强量进行加权,来考虑多个内容类型中的至少一些内容类型。

[0306] 对于上下文类型来说,调整单元300B可以被配置成使环绕声虚拟器 404的环绕声增强量与电影类媒体和/或游戏的置信度值正相关,并且/或者使环绕声增强量与长期音乐和/或VoIP的置信度值负相关。然后,可以构造与公式(5)至公式(8)类似的公式。

[0307] 作为特殊的示例,可以对纯电影类媒体和/或游戏启用环绕声虚拟器 404,但是对音乐和/或VoIP禁用环绕声虚拟器404。同时,可以针对电影类媒体和游戏不同地设置环绕声虚拟器404的增强量。电影类媒体使用更高的增强量,而游戏使用更少的增强量。因此,调整单元300B可以被配置成使环绕声增强量与电影类媒体的置信度值比与游戏更加正相关。

[0308] 与内容类型类似,音频信号的增强量还可以被设定为上下文类型的置信度值的加权平均值：

[0309]

$$\hat{a} = \frac{a_{MOVIE} \cdot Conf_{MOVIE} + a_{MUSIC} \cdot Conf_{MUSIC} + a_{GAME} \cdot Conf_{GAME} + a_{VOIP} \cdot Conf_{VOIP}}{Conf_{MOVIE} + Conf_{MUSIC} + Conf_{GAME} + Conf_{VOIP}} \quad (9)$$

[0310] 其中,$\hat{a}$是估计的增强量,带有上下文类型下标的α是上下文类型的期望/ 预定义的增强量(权重),带有上下文类型下标的Conf是上下文类型的置信度值。视情况而定,a_{MUSIC}和a_{VOIP}可以(但不一定)被设定为0,表示对于纯音乐(内容类型)和/或纯VoIP禁用环绕声虚

拟器404。

[0311] 同样，与内容类型类似，公式(9)中的带有上下文类型下标的α是上下文类型的期望/预定义的增强量，并且相应上下文类型的置信度值被所有所识别的上下文类型的置信度值的和所除的商可以被视为相应上下文类型的预定义/期望的增强量的归一化的权重。也就是说，调整单元300B可以被配置成通过基于置信度值对多个上下文类型的预定义的增强量进行加权，来考虑多个上下文类型中的至少一些上下文类型。

[0312] 3.2起始频率

[0313] 在环绕声虚拟器中还可以修改其他参数，例如起始频率。通常，音频信号中的高频分量更适合于被空间渲染。例如，在音乐中，如果对低音进行空间渲染而使之具有更多的环绕声效果，则其将听起来很奇怪。因此，对于特定的音频信号，环绕声虚拟器需要确定频率阈值，对该阈值以上的分量进行空间渲染而保持该阈值以下的分量。该频率阈值就是起始频率。

[0314] 根据本申请的实施方式，可以对音乐内容增大环绕声虚拟器的起始频率，使得对于音乐信号能够保持更多的低音。因此，调整单元300B可以被配置为使环绕声虚拟器的起始频率与短期音乐的置信度值正相关。

[0315] 3.3实施方式和应用场景的组合

[0316] 与第1部分类似，以上所述的所有实施方式和实施方式的变型可以以其任意组合来实现，并且在不同的部分/实施方式中所提到、但是具有相同或者类似功能的任何组件可以作为同一组件或者单独的组件来实现。

[0317] 例如，在小节3.1至小节3.2中所述的任何两个或更多个解决方案可以彼此组合。并且这些组合还可以与在第1部分、第2部分中所描述的和所暗示的以及在稍后其他部分中将要描述的任何实施方式进行组合。

[0318] 3.4环绕声虚拟器控制方法

[0319] 与第1部分类似，在描述上文实施方式中的环绕声虚拟器控制器的过程中，显然也公开了一些过程和方法。在下文中，在不重复已经论述的细节的情况下给出这些方法的概要。

[0320] 首先，在第1部分中所论述的音频处理方法的实施方式可以用于环绕声虚拟器，环绕声虚拟器的参数是要由音频处理方法调整的目标之一。根据这一点，音频处理方法也是环绕声虚拟器控制方法。

[0321] 在这个小节中，将只论述专用于控制环绕声虚拟器的那些方面。关于控制方法的一般方面，可以参考第1部分。

[0322] 根据一个实施方式，音频处理方法还可以包括环绕声虚拟处理，并且调整操作1104可以被配置成使环绕声虚拟处理的环绕声增强量与噪声和/ 或背景和/或语音的置信度值正相关，并且/或者使环绕声增强量与短期音乐的置信度值负相关。

[0323] 具体地，调整操作1104可以被配置成使环绕声增强量与噪声和/或背景的置信度值比与语音的内容类型更加正相关。

[0324] 可替代地或者附加地，还可以基于上下文的置信度值来调整环绕声增强量。具体地，调整操作1104可以被配置成使环绕声虚拟处理的环绕声增强量与电影类媒体和/或游戏的置信度值正相关，并且/或者使环绕声增强量与长期音乐和/或VoIP的置信度值负相

关。

[0325] 更具体地，调整操作1104可以被配置成使环绕声增强量与电影类媒体的置信度值比与游戏更加正相关。

[0326] 另一个要调整的参数是环绕声虚拟处理的起始频率。如图19所示，首先基于音频类型的置信度值来调整起始频率(操作1902)，然后环绕声虚拟器处理起始频率以上的那些音频分量(操作1904)。具体地，调整操作1104可以被配置成使环绕声虚拟处理的起始频率与短期音乐的置信度值正相关。

[0327] 与音频处理设备的实施方式类似，一方面，音频处理方法的实施方式与实施方式的变型的任何组合都是可行的；另一方面，音频处理方法的实施方式和实施方式的变型的每个方面也可以是单独的解决方案。另外，在本小节中所描述的任何两个或者更多个解决方案可以彼此组合，并且这些组合还可以与在本公开内容的其他部分中所描述的和所暗示的任何实施方式进行组合。

[0328] 第4部分：音量校平器控制器和控制方法

[0329] 不同音频源的音量或者同一音频源中的不同片段的音量有时变化很大。因为用户不得不频繁地调整音量，所以很麻烦。音量校平器(VL)旨在对回放的音频内容的音量进行调节，并且基于目标响度值来使音量在时间轴上保持一致。在A.J.Seefeldt等人的公开号为US2009/0097676A1的“Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal”、B.G.Grockett等人的公开号为WO2007/127023A1的“Audio Gain Control Using Specific-Loudness-Based Auditory Event Detection”以及A.Seefeldt等人的公开号为 WO2009/011827A1的“Audio Processing Using Auditory Scene Analysis and Spectral Skewness”中给出了示例的音量校平器。这三个文档的全部内容通过引用合并到本文中。

[0330] 音量校平器以某种方式连续地测量音频信号的响度，然后以增益量来修改该信号，该增益量是用来修改音频信号的响度的缩放因子，并且通常是所测量的响度、期望的目标响度和若干其他因素的函数。在既要达到目标响度又要保持动态范围的潜在条件下，需要考虑多个因素来估计合适的增益。音量校平器通常包括若干子元素，例如动态增益控制(AGC)、听觉事件检测、动态范围控制(DRC)。

[0331] 在音量校平器中通常应用控制信号来控制音频信号的“增益”。例如，控制信号可以是由纯信号分析得出的音频信号的幅度的变化的指示。控制信号也可以是通过心理声学分析例如听觉情景分析或者基于具体响度的听觉事件检测来表示是否出现新的听觉事件的听觉事件指示。在音量校平器中应用这样的控制信号来进行增益控制，例如，通过确保在听觉事件中增益几乎恒定，以及通过对事件边界附近的大部分增益变化进行限制，以减小由音频信号中的增益的快速变化导致的可能的可听到的畸变。

[0332] 但是，得出控制信号的常用方法不能对信息性的听觉事件和非信息性 (干扰性)的听觉事件进行区分。这里，信息性听觉事件代表包含有意义的信息并且可能被用户更加关注的音频事件，例如对话和音乐，而非信息性的信号不包含对用户有意义的信息，例如VoIP中的噪声。其结果是，非信息性的信号也可能被施加大的增益并且被提高到接近目标响度。在一些应用中这将是非常令人不悦的。例如，在VoIP电话中，在由音量校平器处理之后，出现在通话间歇中的噪声经常被提高到响亮的音量。这对用户来说是不希望有的。

[0333] 为了至少部分地解决该问题，本申请提出基于第1部分中所论述的实施方式来控制音量校平器。

[0334] 与第2部分和第3部分类似，作为第1部分中所论述的音频处理设备和方法的具体示例，音量校平器406可以使用第1部分中所论述的所有实施方式以及在第1部分中所公开的这些实施方式的任何组合。特别地，在控制音量校平器406的情况下，如图1至图10所示的音频处理设备100 的音频分类器200和调整单元300可以组成如图20所示的音量校平器406的控制器2000。在这个实施方式中，因为调整单元是专用于音量校平器 406的，所以其可以被称为300C。

[0335] 也就是说，基于第1部分的公开内容，音量校平器控制器2000可以包括：音频分类器200，用于连续地识别音频信号的音频类型(例如内容类型和/或上下文类型)；以及调整单元300C，用于基于所识别的音频类型的置信度值来以连续的方式调整音量校平器。类似地，音频分类器200可以包括音频内容分类器202和音频上下文分类器204中的至少一个，并且音量校平器控制器2000还可以包括类型平滑单元712、参数平滑单元814 和计时器916中的至少一个。

[0336] 因此，在这个部分中，将不重复在第1部分已经描述的这些内容，而只给出其一些具体示例。

[0337] 可以基于分类结果自适应地调节音量校平器406的不同参数。例如，通过减小非信息性信号的增益，可以调节与动态增益或者动态增益的范围直接有关的参数。也可以调节指示信号是新的可感知的音频事件的程度的参数，然后间接地控制动态增益(该增益将在音频事件中缓慢变化，但是可能在两个音频事件的边界处快速地变化)。在本申请中，给出了参数调节或者音量校平器控制机制的若干实施方式。

[0338] 4.1信息性内容类型和干扰性内容类型

[0339] 如以上所提到的，与音量校平器的控制有关地，音频内容类型可以被分类为信息性内容类型和干扰性内容类型。并且调整单元300C可以被配置成使音量校平器的动态增益与音频信号的信息性内容类型正相关，并且使音量校平器的动态增益与音频信号的干扰性内容类型负相关。

[0340] 作为示例，认为噪声是干扰性的(非信息性的)并且将噪声提高到响亮的音量是令人不快的，直接控制动态增益的参数或者指示新的音频事件的参数可以被设定为与噪声置信度值($Conf_{noise}$)的递减函数成比例，例如：

[0341] $GainControl \propto 1-Conf_{noise}$ (10)

[0342] 这里，为了简单起见，使用符号GainControl来表示与音量校平器中的增益控制有关的所有参数，因为音量校平器的不同实现方式可以使用具有不同潜在含义的不同参数名称。使用单个术语GainControl可以使表达简短而不失其普遍性。实质上，调整这些参数相当于对原始增益施加线性的或非线性的权重。作为一个示例，GainControl可以被直接用于缩放增益，以使得如果GainControl小则增益小。作为另一个具体示例，在B.G. Grockett等人的公开号为WO2007/127023A1的“Audio Gain Control Using Specific-Loudness-Based Auditory Event Detection”中描述了通过用 GainControl来缩放事件控制信号来间接地控制增益，该文献的全部内容通过引用合并到本文中。在此情况中，当GainControl小时，修改对音量校平器的增益的控制以防止增益随着时间显著变化。当GainControl高

时，修改控制以使得校平器的增益能够更自由的变化。

[0343] 使用在公式(10)中所述的增益控制(要么直接将原始增益进行缩放，要么缩放事件控制信号)，音频信号的动态增益与噪声置信度值相关(线性的或者非线性的)。如果信号是具有高置信度值的噪声，则由于因子 $(1-Conf_{noise})$ 而最终的增益将会小。以此方式，避免了将噪声信号提高到令人不悦的响亮的音量。

[0344] 作为公式(10)的示例性变型，如果在应用中(例如在VoIP中)对背景声音也不感兴趣，可以类似地处理背景声音并且对其也施加小的增益。控制函数可以既考虑噪声的置信度值 $(Conf_{noise})$ 又考虑背景置信度值 $(Conf_{bkg})$，例如：

[0345] $GainControl \propto (1-Conf_{noise}) \cdot (1-Conf_{bkg})$ (11)

[0346] 在上面的公式中，因为噪声和背景声音都是不期望的，所以GainControl 同等地受噪声的置信度值和背景的置信度值的影响，并且可以被认为噪声和背景声音具有同一权重。视情况而定，噪声和背景声音可以具有不同的权重。例如，可以对噪声的置信度值和背景声音的置信度值(或者它们与 1的差)给出不同的系数或者不同的指数(α和γ)。也就是说，公式(11) 可以重写成：

[0347] $GainControl \propto (1-Conf_{noise})^{\alpha} \cdot (1-Conf_{bkg})^{\gamma}$ (12)

[0348] 或者

[0349] $GainControl \propto (1-Conf_{noise}^{\alpha}) \cdot (1-Conf_{bkg}^{\gamma})$ (13)

[0350] 或者，调整单元300C可以被配置成基于置信度值来考虑至少一个主导内容类型。例如：

[0351] $GainControl \propto 1-\max(Conf_{noise}, Conf_{bkg})$ (14)

[0352] 公式(11)(及其变型)和公式(14)两者都表示给噪声信号和背景声音信号以小的增益，并且只有当噪声的置信度和背景的置信度两者都小(例如在语音信号和音乐信号中)时才保持音量校平器的原来的工作方式，以使得GainControl接近于1。

[0353] 以上示例是要考虑主导的干扰性内容类型。视情况而定，调整单元 300C也可以被配置成基于置信度值来考虑主导的信息性内容类型。为了更具普遍性，调整单元300C可以被配置成基于置信度值来考虑至少一个主导的内容类型，而不论所识别的音频类型是否是/包括信息性音频类型和 /或干扰性音频类型。

[0354] 作为公式(10)的另一个示例性变型，假设语音信号是最具有信息性的内容并且需要对音量校平器的默认工作方式做较少的修改，控制函数可以考虑噪声置信度值 $(Conf_{noise})$ 和语音置信度值 $(Conf_{speech})$ 两者，如：

[0355] $GainControl \propto 1-Conf_{noise} \cdot (1-Conf_{speech})$ (15)

[0356] 使用该函数，只对具有高噪声置信度且具有低语音置信度(例如，纯噪声) 的那些信号获得小的GainControl，并且如果语音置信度高，则 GainControl将会接近于1(从而因此保持了音量校平器的原来的工作方式)。更一般地，其可以被视为可以根据至少另一个内容类型的置信度值 (例如 $Conf_{speech}$) 来修改一个内容类型(例如 $Conf_{noise}$) 的权重。在以上的公式(15)中，其可以被视为语音的置信度改变了噪声的置信度的权重系数(与公式12和公式13中的权重相比是另一种权重)。换言之，在公式(10)中，$Conf_{noise}$ 的系数可以被视为1；而在公式(15)中，一些其他音频类型(例如语音，但不限于此)将影响噪声的置信度值的重要性，因此可以说 $Conf_{noise}$ 的权重被语音的置信度值修改了。在本公开内容的语境中，术语“权

重”应该被解释为包括这一点。也就是说，其指示了值的重要性，但是不一定是归一化的。可以参考小节1.4。

[0357] 从另一个角度看，与公式(12)和公式(13)类似，在以上函数中可以对置信度值施加指数形式的权重，以表示不同音频信号的优先级(或者重要性)，例如，公式(15)可以被改成：

[0358] $GainControl \propto 1-Conf_{noise}^{\alpha} \cdot (1-Conf_{speech})^{\gamma}$ (16)

[0359] 其中，α和γ是两个权重，如果期望针对校平器参数的修改有更快的响应，则这两个权重可以被设定得更小。

[0360] 可以自由地组合公式(10)至公式(16)以形成可以适合于不同应用的各种控制函数。也可以以类似的方式容易地将其他音频内容类型的置信度值，例如音乐置信度值，合并到控制函数中。

[0361] 在GainControl用于调节表示信号是新的可察觉的音频事件的程度的参数，然后间接地控制动态增益(在音频事件中该增益将缓慢变化，但是在两个音频事件的边界处可能快速变化)的情况下，可以认为在内容类型的置信度值与最终的动态增益之间有另一个传递函数。

[0362] 4.2不同上下文中的内容类型

[0363] 在以上公式(10)至公式(16)的控制函数中考虑了音频内容类型，例如噪声、背景声音、短期音乐和语音的置信度值，但是没有考虑声音所来源的音频上下文，例如电影类媒体和VoIP。有可能需要在不同的音频上下文中对同一音频内容类型例如背景声音进行不同的处理。背景声音包括各种声音，例如汽车引擎、爆炸和鼓掌。在VoIP中，背景信号可能是无意义的，但是在电影类媒体中，背景信号可能是重要的。这表示需要识别感兴趣的音频上下文并且需要针对不同的音频上下文设计不同的控制函数。

[0364] 因此，调整单元300C可以被配置成基于音频信号的上下文类型来将音频信号的内容类型视为信息性的或者干扰性的。例如，通过考虑噪声置信度值和背景置信度值，并且区分VoIP上下文和非VoIP上下文，依赖于音频上下文的控制函数可以是：

[0365] 如果音频上下文是VoIP

[0366] $GainControl \propto 1-\max(Conf_{noise}, Conf_{bkg})$

[0367] 否则 (17)

[0368] $GainControl \propto 1-Conf_{noise}$

[0369] 也就是说，在VoIP上下文中，噪声和背景声音都被视为干扰性的内容类型；而在非VoIP上下文中，背景声音被视为信息性的内容类型。

[0370] 作为另一个示例，考虑语音、噪声和背景的置信度值并且区分VoIP 和非VoIP上下文的依赖于音频上下文的控制函数可以是：

[0371] 如果音频上下文是VoIP

[0372] $GainControl \propto 1-\max(Conf_{noise}, Conf_{bkg})$

[0373] 否则 (18)

[0374] $GainControl \propto 1-Conf_{noise} \cdot (1-Conf_{speech})$

[0375] 这里，语音作为信息性的内容类型被强调。

[0376] 假设在非VoIP上下文中，音乐也是重要的信息性信息，可以将公式 (18)的第2部

分扩展为：

[0377] $GainControl \propto 1-Conf_{noise} \cdot (1-\max(Conf_{speech}, Conf_{music}))$ (19)

[0378] 实际上，控制函数(10)至控制函数(16)中的每个控制函数或者其变型可以应用于不同的/相应的音频上下文。因此，可以产生大量的组合来形成依赖于音频上下文的控制函数。

[0379] 除了在公式(17)和公式(18)中区分和利用的VoIP上下文和非VoIP 上下文之外，可以以相似的方式利用其他音频上下文，例如电影类媒体、长期音乐和游戏，或者低质量音频和高质量音频。

[0380] 4.3上下文类型

[0381] 上下文类型还可以直接用于控制音量校平器以避免那些令人不悦的声音(例如噪声)被提升地太多。例如，VoIP置信度值可以用于操纵音量校平器，使音量校平器在VoIP置信度高时较不灵敏。

[0382] 特别地，使用VoIP置信度值$Conf_{VOIP}$，音量校平器的级别可以被设定成与$(1-Conf_{VOIP})$成比例。也就是说，音量校平器在VoIP内容(当VoIP 置信度值高时)中几乎被停用，这与针对VoIP上下文禁用音量校平器的传统的手工设置(预置)一致。

[0383] 或者，可以针对音频信号的不同上下文设定动态的增益范围。通常，VL(音量校平器)量还调整施加于音频信号的增益的量，并且可以被看做是增益上的另一(非线性)权重。在一个实施方式中，设置可以是：

[0384] 表1

[0385]

	电影类媒体	长期音乐	VOIP	游戏
VL量	高	中等	关闭(或最低)	低

[0386] 此外，假设针对每个上下文类型预定义了期望的VL量。例如，对于电影类媒体VL量被设定为1，对于VoIP为0，对于音乐为0.6，而对于游戏为0.3，但是本申请不限于此。根据此示例，如果电影类媒体的动态增益的范围是100％，则VoIP的动态增益的范围为60％，以此类推。如果音频分类器200的分类是基于硬判决的，则动态增益的范围可以直接设置为上述示例。如果音频分类器200的分类是基于软判决的，则动态增益的范围可以基于上下文类型的置信度值来调整。

[0387] 类似地，音频分类器200可以从音频信号中识别多个上下文类型，并且调整单元300C可以被配置成通过基于该多个上下文类型的重要性将该多个上下文类型的置信度值进行加权，来调整动态增益的范围。

[0388] 通常，对于上下文类型，这里也可以使用与公式(10)至公式(16) 类似的函数——将公式中的内容类型替换为上下文类型——来自适应地设定合适的VL量。实际上，表1反映了不同上下文类型的重要性。

[0389] 从另一个角度看，置信度值可以用于得到在小节1.4中所论述的归一化的权重。假设在表1中针对每个上下文类型预定义了具体的量，则可以应用与公式(9)类似的公式。顺便提及，类似的解决方案也可以被应用于多个内容类型和任何其他音频类型。

[0390] 4.4实施方式和应用场景的组合

[0391] 与第1部分类似，以上所述的所有实施方式和实施方式的变型可以以其任意的组

合来实现，并且在不同的部分/实施方式中所提到的但是具有相同或者类似功能的任何组件可以作为同一组件或者单独的组件来实现。例如，在小节4.1至小节4.3中所述的任何两个或者更多个解决方案可以彼此组合。并且这些组合还可以与在第1部分至第3部分以及在稍后将要描述其他部分中所描述的和所暗示的任何实施方式进行组合。

[0392] 图21通过将原始的短期片段(图21(A))、由不修改参数的常规音量校平器处理的短期片段(图21(B))、和由本申请提出的音量校平器处理的短期片段(图21(C))进行比较，示出了本申请中所提出的音量校平器控制器的效果。可以看出，在如图21(B)所示的常规音量校平器中，噪声(音频信号的后半部分)的音量也被提高了，这是令人不悦的。相比之下，在如图21(C)所示的新的音量校平器中，音频信号的有效部分的音量被提高而没有明显提高噪声的音量，给听众带来良好体验。

[0393] 4.5音量校平器控制方法

[0394] 与第1部分类似，在描述上文实施方式中的音量校平器控制器的过程中，显然也公开了一些处理和方法。在下文中，在不重复已经论述的细节的情况下给出这些方法的概要。

[0395] 首先，在第1部分中所论述的音频处理方法的实施方式可以用于音量校平器，音量校平器的参数是要由音频处理方法调整的目标之一。根据这一点，音频处理方法也是音量校平器控制方法。

[0396] 在这个小节中，将只论述专用于控制音量校平器的那些方面。关于控制方法的一般方面，可以参考第1部分。

[0397] 根据本申请，提供了音量校平器控制方法，包括：实时地识别音频信号的内容类型；以及通过使音量校平器的动态增益与音频信号的信息性内容类型正相关，并且使音量校平器的动态增益与音频信号的干扰性内容类型负相关，来基于所识别的内容类型以连续的方式调整音量校平器。

[0398] 内容类型可以包括语音、短期音乐、噪声和背景声音。通常，噪声被视为干扰性的内容类型。

[0399] 当调整音量校平器的动态增益时，可以基于内容类型的置信度值来直接调整，或者可以通过内容类型的置信度值的传递函数来调整。

[0400] 如已经描述的，音频信号可能被同时分类到多个音频类型中。当涉及多个内容类型时，调整操作1104可以被配置成通过基于该多个内容类型的重要性将该多个内容类型的置信度值进行加权，或者通过基于置信度值将该多个内容类型的影响进行加权，来考虑该多个音频内容类型中的至少一些音频内容类型。特别地，调整操作1104可以被配置成基于置信度值来考虑至少一个主导的内容类型。当音频信号既包括干扰性内容类型又包括信息性内容类型时，调整操作可以被配置成基于置信度值来考虑至少一个主导的干扰性内容类型，并且/或者基于置信度值来考虑至少一个主导的信息性内容类型。

[0401] 不同的音频类型可能彼此影响。因此，调整操作1104可以被配置成使用至少一个其他内容类型的置信度值来修改一个内容类型的权重。

[0402] 如在第1部分中所描述的，可以对音频信号的音频类型的置信度值进行平滑。关于平滑操作的细节请参考第1部分。

[0403] 该方法还可以包括识别音频信号的上下文类型，其中，调整操作1104 可以被配置成基于上下文类型的置信度值来调整动态增益的范围。

[0404] 内容类型的角色受限于其所处的上下文类型。因此，当对于音频信号同时(即针对同一音频片段)既获得内容类型信息又获得上下文类型信息时，基于音频信号的上下文类型可以将音频信号的内容类型确定为信息性的或者干扰性的。另外，取决于音频信号的上下文类型，可以给不同上下文类型的音频信号中的内容类型分配不同的权重。从另一个角度看，可以使用不同的权重(较大的或者较小的，正值或者负值)来反映内容类型的信息性质或者干扰性质。

[0405] 音频信号的上下文类型可以包括VoIP、电影类媒体、长期音乐和游戏。并且在VoIP上下文类型的音频信号中，背景声音被视为干扰性内容类型；而在非VoIP上下文类型的音频信号中，背景和/或语音和/或音乐被视为信息性内容类型。其他上下文类型可以包括高质量音频或者低质量音频。

[0406] 与多个内容类型类似，当音频信号同时(针对同一音频片段)被分类到具有相应置信度值的多个上下文类型时，调整操作1104可以被配置成通过基于该多个上下文类型的重要性将该多个上下文类型的置信度值进行加权，或者通过基于置信度值将该多个上下文类型的影响进行加权，来考虑该多个上下文类型中的至少一些上下文类型。特别地，调整操作可以被配置成基于置信度值来考虑至少一个主导的上下文类型。

[0407] 最后，如本小节所述的方法的实施方式可以使用如将在第6部分和第 7部分中所论述的音频分类方法，这里省略其详细描述。

[0408] 与音频处理设备的实施方式类似地，一方面，音频处理方法的实施方式与实施方式的变型的任何组合都是可行的；另一方面，音频处理方法的实施方式和实施方式的变型的每个方面可以是单独的解决方案。另外，在本小节中所描述的任何两个或者更多个解决方案可以彼此组合，并且这些组合还可以与在本公开内容的其他部分中所描述的和所暗示的任何实施方式进行组合。

[0409] 第5部分：均衡控制器和控制方法

[0410] 均衡器通常应用于音乐信号以调整或者修改音乐信号的谱平衡，该谱平衡被称为“音调”或者“音色”。传统的均衡器允许用户为了强调某个声音或者去除不期望的声音而在每个单独的频带上配置频率响应(增益) 的整体模式(曲线或者形状)。流行的音乐播放器例如Windows(视窗) 媒体播放器通常提供图形均衡器来调整每个频带处的增益，而且也提供一组针对不同音乐风格例如摇滚、说唱、爵士和民谣的预置，以在倾听不同风格的音乐过程中得到最佳体验。一旦选择了预置或者设定了模式，则在信号上施加同一均衡增益直到手动地修改该模式为止。

[0411] 相比之下，为了保持与期望的音调或音色相关的谱平衡的整体一致性，动态均衡器提供了自动的调整每个频带处的均衡增益的方式。通过连续地监测音频的谱平衡，将其与期望的预置谱平衡相比较，以及动态地调整所施加的均衡增益以将音频的原始谱平衡转换为期望谱平衡，从而实现所述一致性。期望谱平衡是在处理前手工地选择或者预置的。

[0412] 两种类型的均衡器都有以下缺点：必须手动地选择最佳均衡模式、期望谱平衡或者相关的参数，并且不能基于回放的音频内容来动态地修改最佳均衡模式、期望谱平衡或者相关的参数。为了针对不同类别的音频信号来提供整体的高质量，区分音频内容类型是非常重要的。例如，不同的音乐作品需要不同的均衡模式，例如不同风格音乐的均衡模式。

[0413] 在有可能输入各种音频信号(不仅是音乐)的均衡器系统中，需要基于内容类型来

调整均衡器参数。例如，通常对音乐信号启用均衡器，但是对语音信号禁用均衡器，因为均衡器可能过多地改变语音的音色而相应地使信号听起来不自然。

[0414] 为了至少部分地解决这个问题，本申请提出了基于第1部分中所论述的实施方式来控制均衡器。

[0415] 与第2部分至第4部分类似，作为第1部分中所论述的音频处理设备和方法的具体示例，均衡器408可以使用第1部分中所论述的所有实施方式以及在第1部分中所公开的这些实施方式的任何组合。特别地，在控制均衡器408的情况下，如图1至图10所示的音频处理设备100的音频分类器200和调整单元300可以组成如图22所示的均衡器408的控制器2200。在这个实施方式中，因为调整单元是专用于均衡器408的，所以其可以被称为300D。

[0416] 也就是说，基于第1部分的公开内容，均衡器控制器2200可以包括：音频分类器200，用于连续地识别音频信号的音频类型；以及调整单元 300D，用于基于所识别的音频类型的置信度值来以连续的方式调整均衡器。类似地，音频分类器200可以包括音频内容分类器202和音频上下文分类器204中的至少一个，并且音量均衡器控制器2200还可以包括类型平滑单元712、参数平滑单元814和计时器916中的至少一个。

[0417] 因此，在这个部分中，将不重复在第1部分已经描述的这些内容，而只给出它们的一些具体的示例。

[0418] 5.1基于内容类型的控制

[0419] 一般而言，针对一般的音频内容类型例如音乐、语音、背景声音和噪声，应该对不同的内容类型不同地设置均衡器。与传统设定类似，对音乐信号能够自动地启用均衡器，但是对语音自动地禁用均衡器；或者以更加连续的方式，对音乐信号设置高的均衡级别但是对语音信号设置低的均衡级别。以此方式，可以针对音频内容来自动地设定均衡器的均衡级别。

[0420] 尤其是对于音乐，观察到均衡器对具有主导源的音乐片段效果不是很好，因为如果施加了不适当的均衡，可能显著地改变主导源的音色并且听起来不自然。考虑到这一点，较好的是在具有主导源的音乐片段上设定低的均衡级别，而对没有主导源的音乐片段保持高的均衡级别。使用这个信息，均衡器可以针对不同的音乐内容来自动地设定均衡级别。

[0421] 还可以基于不同的属性——例如风格、乐器和包括节奏、速度和音色的一般音乐特征——来对音乐分类。以相同的方式可以针对不同的音乐风格使用不同的均衡预置，这些音乐群/类型还可以具有其自身的最优均衡模式或者均衡曲线(在传统的均衡器中)或者最优的期望谱平衡(在动态均衡器中)。

[0422] 如以上所述，通常对音乐内容启用均衡器，但是对语音禁用均衡器，因为由于音色变化，对于对话，均衡器可能使之听上去不佳。自动实现这一点的一个方法是使均衡级别与内容相关，内容具体指从音频内容分类模块获得的音乐置信度值和/或语音置信度值。这里，均衡级别可以被解释为所施加的均衡增益的权重。级别越高，则所施加的均衡越强。针对该示例，如果均衡级别是1，则施加完整的均衡模式；如果均衡级别是0，相应地所有的增益是0dB而因此没有施加均衡。在均衡器算法的不同实现方式中可以用不同参数来表示均衡级别。该参数的一个示例性实施方式是如在 A.Seefeldt等人的公开号为US 2009/0097676A1的“Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal”中所实现的均衡器的权重，该文献的

全部内容通过引用合并到本文中。

[0423] 可以设计各种控制方案来调节均衡级别。例如，使用音频内容类型信息，可以使用语音置信度值或者音乐置信度值来设定均衡级别，如：

[0424] $L_{eq} \propto Conf_{music}$ (20)

[0425] 或者

[0426] $L_{eq} \propto 1-Conf_{speech}$ (21)

[0427] 其中，L_{eq}是均衡级别，并且$Conf_{music}$和$Conf_{speech}$代表音乐和语音的置信度值。

[0428] 也就是说，调整单元300D可以被配置成使均衡级别与短期音乐的置信度值正相关，或者使均衡级别与语音的置信度值负相关。

[0429] 还可以结合地使用语音置信度值和音乐置信度值来设定均衡级别。总体思想是：只有当音乐置信度值高且语音置信度低时，均衡级别才应该高，否则均衡级别低。例如：

[0430] $L_{eq} = Conf_{music}(1-Conf_{speech}^{\alpha})$ (22)

[0431] 其中，为了处理在音乐信号中可能经常出现的非0的语音置信度，给语音置信度上加指数α。使用以上公式，对没有任何语音成分的纯音乐信号完整地施加均衡(级别等于1)。如在第1部分中所陈述的，α可以被视为基于内容类型的重要性的权重系数，并且通常可以被设定到1至2。

[0432] 如果对语音的置信度值设置更大的权重，则调整单元300D可以被配置成当针对语音内容类型的置信度值大于阈值时禁用均衡器408。

[0433] 在以上描述中，将音乐内容类型和语音内容类型作为示例。可替代地或者附加地，也可以考虑背景声音和/或噪声的置信度值。具体地，调整单元300D可以被配置成使均衡级别与背景声音的置信度值正相关，并且/或者使均衡级别与噪声的置信度值负相关。

[0434] 作为另一个实施方式，置信度值可以用于得出如小节1.4中所述的归一化的权重。假设针对每个内容类型预定义了期望的均衡级别(例如，对音乐是1、对语音是0、对噪声和背景声音是0.5)，则完全可以应用与公式(8)类似的公式。

[0435] 还可以对均衡级别进行平滑，以避免在可能在转换点处引入可听见畸变的快速变化。可以使用如小节1.5所述的参数平滑单元814。

[0436] 5.2音乐中的主导源的可能性

[0437] 为了避免具有主导源的音乐被施加高的均衡级别，还可以使均衡级别与指示音乐片段是否包括主导源的置信度值$Conf_{dom}$相关，例如：

[0438] $L_{eq} = 1-Conf_{dom}$ (23)

[0439] 以此方式，均衡级别对具有主导源的音乐片段较低，而对没有主导源的音乐片段较高。

[0440] 这里，虽然描述的是具有主导源的音乐的置信度值，但是也可以使用没有主导源的音乐的置信度值。也就是说，调整单元300D可以被配置成使均衡级别与没有主导源的短期音乐的置信度值正相关，并且/或者使均衡级别与具有主导源的短期音乐的置信度值负相关。

[0441] 如小节1.1所述，虽然作为一方面的语音和音乐和作为另一方面的具有主导源的音乐或不具有主导源的音乐是不同层次级别上的内容类型，但是可以平行地考虑它们。通过如上所述联合考虑主导源的置信度值以及语音置信度值和音乐置信度值，可以通过将公

式(20)至公式(21)中的至少一个与公式(23)进行合并来设定均衡级别。一个示例是合并所有三个公式：

[0442] $L_{eq}=Conf_{music}(1-Conf_{speech})(1-Conf_{dom})$ (24)

[0443] 为了更具一般性，还可以对不同的置信度值施加基于内容类型的重要性的不同的权重，例如以公式(22)中的方式。

[0444] 作为另一个示例，假设只有当音频信号是音乐时才计算$Conf_{dom}$，可以设置分段函数，如：

[0445] $$L_{eq}=\begin{cases}(1-Conf_{dom}) & Conf_{music}>threshold\\ Conf_{music}(1-conf_{speech}^{\alpha}) & \text{否则}\end{cases} \quad (25)$$

[0446] 如果分类系统相当确定该音频是音乐(音乐置信度值大于阈值)，则该函数基于主导源的置信度值来设定均衡级别；否则，基于音乐置信度值和语音置信度值来设定均衡级别。也就是说，调整单元300D可以被配置成当短期音乐的置信度值大于阈值时考虑具有/不具有主导源的短期音乐。当然，可以以公式(20)至(24)的方式来修改公式(25)的前半部分或者后半部分。

[0447] 也可以应用如小节1.5中所述的同一平滑方案，并且还可以基于转换类型来设定常数α，转换类型例如位从具有主导源的音乐到不具有主导源的音乐的转换，或者从不具有主导源的音乐到具有主导源的音乐的转换。为此目的，也可以应用与公式(4')类似的公式。

[0448] 5.3均衡器的预置

[0449] 除了基于音频内容类型的置信度值来自适应地调节均衡级别之外，还可以取决于其风格、乐器或者其他特性，针对不同的音频内容来自动地选择适当的均衡模式预置或者期望谱平衡预置。具有同样风格的音乐、包括同一乐器的音乐或者具有同一音乐特性的音乐可以共享同一均衡模式预置或者期望谱平衡预置。

[0450] 为了具有一般性，使用术语“音乐类型”来表示具有同一风格、同一乐器或者相似音乐属性的音乐群，并且它们可以被视为如小节1.1中所述的音频内容类型的另一个层次级别。适当的均衡模式、均衡级别、和/或期望谱平衡预置可以与每个音乐类型关联。均衡模式是施加在音乐信号上的增益曲线，它可以是针对不同音乐类型(例如经典、摇滚、爵士和民谣) 而使用的均衡器预置中的任何一个均衡器预置。期望谱平衡预置表示针对每个类型的期望的音色。图23示出了在杜比家庭剧院(Dolby Home Theater)技术中实现的期望谱平衡预置的几个示例。每一个示例描述了在整个可听频率范围上的期望谱形状。连续地将该形状与输入的音频的谱形状进行比较，并且根据该比较计算出均衡增益，从而将输入的音频的谱形状变换到预置的谱形状。

[0451] 对于一个新的音乐片段，可以确定最接近的类型(硬判决)，或者可以计算与每个音乐类型相关的置信度值(软判决)。基于这个信息，可以针对给定的音乐作品来确定合适的均衡模式或者期望谱平衡预置。最简单的方式是给其分配最匹配的类型的相应模式，如：

[0452] $$P_{eq}=P_{c^*} \quad (26)$$

[0453] 其中，P_{eq}是估算的均衡模式或期望谱平衡预置，并且c^*是最匹配的音乐类型的索引(主导音频类型)，可以通过选择具有最高置信度值的类型来得到该索引。

[0454] 而且，可能有多于一个的具有大于0的置信度值的音乐类型，意味着该音乐作品具有与那些类型或多或少类似的属性。例如，音乐作品可以具有多个乐器，或者其可以具有多种风格的属性。这启发了通过考虑所有类型而不是只使用一个最接近的类型来估计合适的均衡模式的另一方式。例如，可以使用加权和：

[0455] $$P_{eq}=\sum_{c=1}^{N}w_cP_c \qquad (27)$$

[0456] 其中，N是预定义的类型的数量，w_c是与每个预定义的音乐类型(索引是 c)相关的设计模式P_c的权重，w_c应该基于其相应置信度值被归一化到1。以此方式，估计的模式将是各种音乐类型的模式的混合。例如，对于既具有爵士属性又具有摇滚属性的音乐作品来说，估计的模式将是介于两者之间的模式。

[0457] 在一些应用中，可能不希望如公式(27)所示的那样涉及所有的类型。只有类型的一个子集——与当前的音乐作品最相关的类型——需要被考虑，公式(27)可稍微改进成：

[0458] $$P_{eq}=\sum_{c'=1}^{N'}w_{c'}P_{c'} \qquad (28)$$

[0459] 其中，N’是要考虑的类型的数量，c’是基于类型的置信度值而降序排列的类型索引。通过使用子集，可以更加关注最接近的类型而排除那些不太相关的类型。换言之，调整单元300D可以被配置成基于置信度值来考虑至少一个主导音频类型。

[0460] 在以上的描述中，将音乐类型作为示例。实际上，该方案对于如小节 1.1所示的任何层次级别上的音频类型都是可应用的。因此，一般地，调整单元300D可以被配置成给每个音频类型分配均衡级别和/或均衡模式和 /或谱平衡预置。

[0461] 5.4基于上下文类型的控制

[0462] 在之前的小节中，讨论的焦点在于各种内容类型。在本小节要讨论的更多实施方式中，将替代地或者附加地考虑上下文类型。

[0463] 通常，针对音乐启用均衡器但是针对电影类媒体内容禁用均衡器，因为由于明显的音色变化，均衡器可能使电影类媒体中的对话听起来不佳。这表示均衡级别可以与长期音乐的置信度和/或电影类媒体的置信度关联：

[0464] $L_{eq}\propto Conf_{MUSIC}$ (29)

[0465] 或者

[0466] $L_{eq}\propto 1-Conf_{MOVIE}$ (30)

[0467] 其中，L_{eq}是均衡级别，$Conf_{MUSIC}$和$Conf_{MOVIE}$代表长期音乐和电影类媒体的置信度值。

[0468] 也就是说，调整单元300D可以被配置成使均衡级别与长期音乐的置信度值正相关，或者使均衡级别与电影类媒体的置信度值负相关。

[0469] 也就是说，针对电影类媒体信号，电影类媒体置信度值高(或者音乐置信度低)，因此均衡级别低；另一方面，针对音乐信号，电影类媒体置信度值低(或者音乐置信度高)，因此均衡级别高。

[0470] 可以以与公式(22)至公式(25)相同的方式来修改公式(29)和公式(30)中所示的

解决方案，并且/或者公式(29)和公式(30)可以与公式(22)至公式(25)所示的方案中的任意一个方案进行组合。

[0471] 附加地或者可选地，调整单元300D可以被配置成使均衡级别与游戏的置信度值负相关。

[0472] 作为另一个实施方式，置信度值可以用于得出如小节1.4所述的归一化的权重。假设针对每个上下文类型预定义了期望的均衡级别/模式(在下面的表2中示出了均衡模式)，也可以应用与公式(9)类似的公式。

[0473] 表2：

[0474]

	电影类媒体	长期音乐	VoIP	游戏
均衡模式	模式1	模式2	模式3	模式4

[0475] 这里，在一些模式中，作为针对某个上下文类型(例如电影类媒体和游戏)禁用均衡器的一种方式，所有的增益可以被设定为零。

[0476] 5.5实施方式和应用场景的组合

[0477] 与第1部分类似，以上所述的所有实施方式和实施方式的变型可以以其任意的组合来实现，并且在不同的部分/实施方式中所提到的但是具有相同或者类似功能的任何组件可以作为同一组件或者单独的组件来实现。

[0478] 例如，在小节5.1至小节5.4中所述的任何两个或者更多个解决方案可以彼此组合。并且这些组合还可以与在第1部分至第4部分以及在稍后将要描述的其他部分中所描述的和所暗示的任何实施方式进行组合。

[0479] 5.6均衡器控制方法

[0480] 与第1部分类似，在描述上文实施方式中的均衡器控制器的过程中，显然也公开了一些处理和方法。在下文中，在不重复已经论述的细节的情况下给出这些方法的概要。

[0481] 首先，在第1部分中所论述的音频处理方法的实施方式可以用于均衡器，均衡器的参数是要由音频处理方法调整的目标之一。根据这一点，音频处理方法也是均衡器控制方法。

[0482] 在这个小节中，将只论述专用于控制均衡器的那些方面。关于控制方法的一般方面，可以参考第1部分。

[0483] 根据一些实施方式，均衡器控制方法可以包括：实时地识别音频信号的音频类型；以及基于所识别的音频类型以连续的方式调整均衡器。

[0484] 与本申请的其他部分类似，当涉及具有相应置信度值的多个音频类型时，调整操作1104可以被配置成通过基于该多个音频类型的重要性对该多个音频类型的置信度值进行加权，或者通过基于置信度值将该多个音频类型的影响进行加权，来考虑该多个音频类型中的至少一些音频类型。特别地，调整操作1104可以被配置成基于置信度值来考虑至少一个主导音频类型。

[0485] 如第1部分所述，可以对经调整的参数值进行平滑。可以参考小节1.5 和小节1.8，并且在这里省略详细描述。

[0486] 音频类型既可以是内容类型也可以是上下文类型，或者是两者。当涉及内容类型时，调整操作1104可以被配置成使均衡级别与短期音乐的置信度值正相关，并且/或者使均

衡级别与语音的置信度值负相关。附加地或者可替代地，调整操作还可以被配置成使均衡级别与背景声音的置信度值正相关，并且/或者使均衡级别与噪声的置信度值负相关。

[0487] 当涉及上下文类型时，调整操作1104可以被配置成使均衡级别与长期音乐的置信度值正相关，并且/或者使均衡级别与电影类媒体和/或游戏的置信度值负相关。

[0488] 针对短期音乐的内容类型，调整操作1104可以被配置成使均衡级别与不具有主导源的短期音乐的置信度值正相关，并且/或者使均衡级别与具有主导源的短期音乐的置信度值负相关。可以只当短期音乐的置信度值大于阈值时才完成这一点。

[0489] 除了调整均衡级别，可以基于音频信号的音频类型的置信度值来调整均衡器的其他方面。例如，调整操作1104可以被配置成给每个音频类型分配均衡级别和/或均衡模式和/或谱平衡预置。

[0490] 关于音频类型的具体实例，可以参考第1部分。

[0491] 与音频处理设备的实施方式类似地，一方面，音频处理方法的实施方式与实施方式的变型的任何组合都是可行的；另一方面，音频处理方法的实施方式和实施方式的变型的每个方面也可以是单独的解决方案。另外，在本小节中所描述的任何两个或者更多个解决方案可以彼此组合，并且这些组合还可以与在本公开的其他部分中所描述的和所暗示的任何实施方式进行组合。

[0492] 第6部分：音频分类器和分类方法

[0493] 如小节1.1和小节1.2所述，在本申请中论述的包括各种层次级别的内容类型和上下文类型的音频类型可以通过现有分类方案包括基于机器学习的方法来分类或者识别。在本部分和接下来的部分中，如在之前的部分中所提到的，本申请提出了用于对上下文类型进行分类的分类器和方法的一些新颖的方面。

[0494] 6.1基于内容类型分类的上下文分类

[0495] 如在之前的部分所提到的，音频分类器200用于识别音频信号的内容类型和/或识别音频信号的上下文类型。因此，音频分类器200可以包括音频内容分类器202和/或音频上下文分类器204。当采用现有技术来实现音频内容分类器202和音频上下文分类器204时，这两个分类器可以彼此独立，即使它们可能共享一些特征而因此可能共享用于提取特征的一些方案。

[0496] 在本部分和接下来的第7部分中，根据在本申请中所提出的新颖的方面，音频上下文分类器204可以使用音频内容分类器202的结果，也就是说，音频分类器200可以包括：音频内容分类器，用于识别音频信号的内容类型；以及音频上下文分类器204，用于基于音频内容分类器202的结果来识别音频信号的上下文类型。这样，音频内容分类器202的分类结果可以既由音频上下文分类器204使用，又由如在前面的部分中所述的调整单元300(或者调整单元300A至调整单元300D)来使用。但是，虽然在附图中没有示出，音频分类器200还可以包括分别由调整单元300和音频上下文分类器204使用的两个音频内容分类器202。

[0497] 另外，如小节1.2所述，尤其是当分类多个音频类型时，音频内容分类器202或者音频上下文分类器204可以包括一组彼此协作的分类器，虽然它们也可以被实现成单个分类器。

[0498] 如小节1.1所述，内容类型是针对通常具有数帧至数十帧量级的长度 (例如1s)的短期音频片段的一种音频类型，而上下文类型是针对通常具有数秒至数十秒量级的长度

(例如10s)的长期音频片段的一种音频类型。因此,与“内容类型”和“上下文类型”相应地,必要时分别使用术语“短期”和“长期”。但是,如在接下来的第7部分所要描述的,虽然上下文类型是用于指示在相对长的时间尺度上的音频信号的性质,但是也可以基于从短期音频片段提取的特征来识别上下文类型。

[0499] 现在,参照图24说明音频内容分类器202和音频上下文分类器204 的结构。

[0500] 如图24所示,音频内容分类器202可以包括短期特征提取器2022,用于从各自包括音频帧序列的短期音频片段提取短期特征;以及短期分类器2024,用于使用相应的短期特征来将长期音频片段中的短期片段序列分类到短期音频类型中。短期特征提取器2022和短期分类器2024两者都可以以现有技术来实现,但是在接下来的小节6.3中也针对短期特征提取器 2022提出了一些改进。

[0501] 短期分类器2024可以被配置成将该短期片段序列中的每个短期片段分类到已经在小节1.1中解释过的以下短期音频类型(内容类型)中的至少一个短期音频类型:语音、短期音乐、背景声音和噪声。每个内容类型还可以被进一步分类到更低层次级别的内容类型中,例如在小节1.1中所论述的,但是不限于此。

[0502] 如在本领域已知的,还可以由短期分类器2024获得经分类的音频类型的置信度值。在本申请中,当提到任何分类器的操作时,应该理解的是无论是否进行了明确的说明,如果有必要则同时也获得了置信度值。可以在L.Lu,H.-J.Zhang和S.Li的″Content-based Audio Classification and Segmentation by Using Support Vector Machines″,ACM Multimedia Systems Journal 8(6),pp.482-492(2003.3)中找到音频类型分类的示例,该文献的全部内容通过引用合并到本文中。

[0503] 另一方面,如图24所示,音频上下文分类器204可以包括统计数据提取器2042,用于计算短期分类器针对该长期音频片段中的短期片段序列的结果的统计数据,作为长期特征;以及长期分类器2044,用于使用长期特征将长期音频片段分类到长期音频类型中。类似地,统计数据提取器 2042和长期分类器2044两者都可以以现有技术来实现,但是在接下来的小节6.2中也针对统计数据提取器2042提出了的一些改进。

[0504] 长期分类器2044可以被配置成将长期音频片段分类到以下已经在小节1.1中解释过的长期音频类型(上下文类型)中的至少一个长期音频类型:电影类媒体、长期音乐、游戏和VoIP。可替代地或者附加地,长期分类器2044可以被配置成将长期音频片段分类到已经在小节1.1中解释过的 VoIP或者非VoIP中。可替代地或者附加地,长期分类器2044可以被配置成将长期音频片段分类到已经在小节1.1中解释过的高质量音频或者低质量音频中。在实际中,可以基于应用/系统的需求来选择和训练各种目标音频类型。

[0505] 关于短期片段和长期片段(以及在小节6.3中要论述的帧)的含义和选择,可以参考小节1.1。

[0506] 6.2长期特征的提取

[0507] 如图24所示,在一个实施方式中,只有统计数据提取器2042被用于从短期分类器2024的结果提取长期特征。作为长期特征,可以由统计数据提取器2042计算出以下数据中的至少一个:待分类的长期片段中的短期片段的短期音频类型的置信度值的平均值和方差,由短期片段的重要程度加权的上述平均值和方差,在待分类的长期片段中每个短期音频类型的出现频率和在不同短期音频类型之间转换的频率。

[0508] 图25中示出了在每个短期片段(长度为1s)中的语音置信度值和短期音乐置信度值的平均值。为了对比,从三个不同的音频上下文提取该片段:电影类媒体(图25(A))、长期音乐(图25(B))、和VoIP(图25 (C))。可以观察出,对于电影类媒体上下文,不论是针对语音类型还是音乐类型都可以得到高的置信度值,而其在这两个音频类型中频繁地交替。相比之下,长期音乐的片段给出稳定且高的短期音乐置信度值和相对稳定且低的语音置信度值。而VoIP的片段给出平稳且低的短期音乐置信度值,但是给出了由于VoIP对话期间的停顿而涨落的语音置信度值。

[0509] 针对每个音频类型的置信度值的方差也是用于分类不同音频上下文的特征。图26给出了电影类媒体、长期音乐和VoIP音频上下文中的语音、短期音乐、背景和噪声的置信度值的方差的柱状图(横坐标是数据集中的置信度值的方差,纵坐标是数据集中的方差值的每个区间(bin)的发生数量,其可以被归一化以指示方差值的每个区间的发生概率)。对于电影类媒体,所有语音、短期音乐和背景的置信度的方差都相对较高并且分布较宽,说明这些音频类型的置信度值强烈地变化;对于长期音乐,所有语音、短期音乐和背景的置信度的方差都相对较低并且分布较窄,指示这些音频类型的置信度值保持稳定:语音置信度值保持恒定的低而音乐置信度值保持恒定的高;对于VoIP,短期音乐的置信度值的方差较低并且分布较窄,而语音的置信度值的方差相对地分布较宽,这是由于在VoIP通话期间的经常的停顿。

[0510] 关于用于计算加权平均值和方差的权重,其是基于每个短期片段的重要度来确定的。可以通过短期片段的能量和响度来测量短期片段的重要度,可以用很多现有技术来估计短期片段的能量和响度。

[0511] 待分类的长期片段中的每个短期音频类型的出现频率是:对长期片段中的短期片段被分类到的每个音频类型的计数,由长期片段的长度进行归一化。

[0512] 在待分类的长期片段中的在不同短期音频类型间转换的频率是:对待分类的长期片段中的两个相邻短期片段改变音频类型的计数,由长期片段的长度进行归一化。

[0513] 当参照图25论述置信度值的平均值和方差时,实际上也涉及了每个短期音频类型的出现频率和在那些不同短期音频类型之间的转换频率。这些特征也与音频上下文分类高度相关。例如,长期音乐主要包括短期音乐音频类型,所以长期音乐具有高的短期音乐出现频率,而VoIP主要包括语音和停顿,所以VoIP具有高的语音或噪声出现频率。作为另一个示例,电影类媒体比长期音乐或者VoIP更频繁地在不同短期音频类型之间转换,所以电影类媒体通常具有更高的在短期音乐、语音和背景间的转换频率; VoIP通常比其他类型更频繁地在语音和噪声之间转换,因此VoIP通常具有更高的在语音和噪声之间的转换频率。

[0514] 通常,假设在同一应用/系统中长期片段具有同一长度。如果是这种情况,则每个短期音频类型的发生计数以及在长期片段中不同短期音频类型间的转换计数可以被直接使用而不需要归一化。如果长期片段的长度是可变的,则应该使用如上所提到的出现频率和转换频率。本申请的权利要求应该被解释为涵盖这两种情况。

[0515] 附加地或者可替代地,音频分类器200(或者音频上下文分类器204) 还可以包括长期特征提取器2046(图27),用于基于长期音频片段中的短期片段序列的短期特征进一步从长期音频片段中提取长期特征。换言之,长期特征提取器2046不使用短期分类器2024的分类结果,而是直接使用由短期特征提取器2022提取的短期特征,来得到要由长期分类器

2044使用的一些长期特征。长期特征提取器2046和统计数据提取器2042可以独立地使用或者联合的使用。换言之，音频分类器200既可以包括长期特征分类器2046也可以包括统计数据提取器2042，或者可以包括两者。

[0516] 可以由长期特征提取器2046提取任何特征。在本申请中，提出了计算来自短期特征提取器2022的短期特征的下述统计数据的至少之一作为长期特征：平均值、方差、加权平均、加权方差、高平均(high average)、低平均(low average)、以及高平均与低平均之间的比率(对比度)。

[0517] 从待分类的长期片段中的短期片段提取的短期特征的平均值和方差；

[0518] 从待分类的长期片段中的短期片段提取的短期特征的加权平均值和加权方差。基于刚提到的使用短期片段的能量或者响度而测量的每个短期片段的重要度来对短期特征进行加权；

[0519] 高平均：从待分类的长期片段中的短期片段提取的被选中的短期特征的平均值。当短期特征满足以下条件的至少一个条件时被选中：大于阈值；或者在不低于所有其他短期特征的预定比例的短期特征中，例如，最高的 10％的短期特征；

[0520] 低平均：从待分类的长期片段中的短期片段提取的被选中的短期特征的平均值。当短期特征满足以下条件的至少一个条件时被选中：小于阈值；或者在不高于所有其他短期特征的预定比例的短期特征中，例如，最低的 10％的短期特征；以及

[0521] 对比度：高平均与低平均之间的比率，表示长期片段中的短期特征的动态。

[0522] 可以用现有技术来实现短期特征提取器2022，并且可以由此提取任何特征。尽管如此，在接下来的小节6.3中针对短期特征提取器2022提出了一些改进。

[0523] 6.3短期特征的提取

[0524] 如图24和图27所示，短期特征提取器2022可以被配置成直接从每个短期音频片段提取以下特征中的至少一个特征：节奏特性、中断/静音特性和短期音频质量特征。

[0525] 节奏特性可以包括节奏强度、节奏规律性、节奏清晰性(参见L.Lu,D. Liu,和H.-J.Zhang.“Automatic mood detection and tracking of music audio signals”.IEEE Transactions on Audio,Speech,and Language Processing,14(1):518,2006，其全部内容通过引用合并到本文中)和2D 子带调制(M.F McKinney和J.Breebaart.“Features for audio and music classification”,Proc.ISMIR,2003，其全部内容通过引用合并到本文中)。

[0526] 中断/静音特性可以包括语音中断、突然下降、静音长度、不自然安静、不自然安静的平均值、不自然安静的总能量等。

[0527] 短期音频质量特征是关于短期片段的音频质量特征，其与从音频帧提取的音频质量特征类似，将在以下论述。

[0528] 可替代地或者附加地，如图28所示，音频分类器200可以包括帧级特征提取器2012，用于从短期片段中所包括的音频帧序列的每个帧中提取帧级特征，并且短期特征提取器2022可以被配置成基于从音频帧序列中提取的帧级特征来计算短期特征。

[0529] 作为预处理，输入音频信号可以被下混合(down-mix)为单声道音频信号。如果音频信号已经是单声道信号则不需要该预处理。然后以预定义的长度(通常10毫秒到25毫秒)将其划分为帧。相应地，从每个帧提取帧级特征。

[0530] 帧级特征提取器2012可以被配置成提取以下特征中的至少一个特征:以各种短期音频类型的属性表征的特征、截止频率、静态信噪比(static SNR)特性、分段信噪比(segmental SNR)特性、基本语音描述特征(basic speech descrptor)和声道(vocal tract)特性。

[0531] 以各种短期音频类型(尤其是语音、短期音乐、背景声音和噪声)的属性表征的特征可以包括以下特征中的至少一个特征:帧能量、子带谱分布、谱通量(spectral flux)、梅尔倒谱系数(Mel-frequency Cepstral Coefficient,MFCC)、低音(bass)、残留(residual)信息、纯度(Chroma) 特征和过零率(zero-crossing rate)。

[0532] 关于MFCC的细节,可以参考L.Lu,H.-J.Zhang和S.Li, "Content-based Audio Classification and Segmentation by Using Support Vector Machines",ACM Multimedia Systems Journal 8(6),pp.482-492 (2003.3),其全部内容通过引用合并到本文中。关于音调色度信息的细节,可以参考G.H.Wakefield,"Mathematical representation of joint time Chroma distributions"in SPIE,1999,其全部内容通过引用合并到本文中。

[0533] 截止频率表示音频信号的最高频率,高于该最高频率的内容的能量接近于零。截止频率被设计用来检测频带受限的内容(band limited content),在本申请中,频带受限的内容对于音频上下文分类是有用的。截止频率通常是由编码导致的,因为大部分编码器在低比特率或者中等比特率时丢弃高频成分。例如,MP3编解码器在128kbps时具有16kHz的截止频率;再例如,很多流行的VoIP编解码器具有8kHz或者16kHz的截止频率。

[0534] 除了截止频率,还将音频编码处理期间的信号退化当作另一个特性,其用于区分各种音频上下文,例如VoIP上下文与非VoIP上下文,高质量音频上下文与低质量音频上下文。还可以在多个级别中进一步提取代表音频质量的特征,例如用于客观语音质量评估的特征(参见Ludovic Malfait,Jens Berger,和Martin Kastner,"P.563-The ITU-T Standard for Single-Ended Speech Quality Assessment",IEEE Transaction on Audio, Speech,and Language Processing,VOL.14,NO.6(2006.11),其全部内容通过引用合并到本文中),以获取更丰富的特征。音频质量特征的示例包括:

[0535] a)静态SNR特性,包括估计的背景噪声级别、谱清晰性等。

[0536] b)分段SNR特性,包括谱级别偏差、谱级别范围、相对最低噪声(relative noise floor)等。

[0537] c)基本语音描述特征,包括音高平均值(pitch average)、语音分段声级变化(speech section level variation)、语音声级等。

[0538] d)声道特性,包括类机器人声化(robotization)、音高互功率(pitch cross power)等。

[0539] 为了从帧级特征得出短期特性,短期特征提取器2022可以被配置成计算帧级特征的统计数据,作为短期特征。

[0540] 帧级特征的统计数据的示例包括平均值和标准偏差,其捕捉了节奏特征以区分各种音频类型,例如短期音乐、语音、背景声音和噪声。例如,语音通常以音节速率在浊音和清音之间交替但是音乐则不这样,表示语音的帧级特征的变化通常比音乐的帧级特征的变化更大。

[0541] 统计数据的另一个示例是帧级特征的加权平均值。例如，针对截止频率，使用每个帧的能量或者响度作为权重，从短期片段中的每个音频帧提取的截止频率间的加权平均值将是针对该短期片段的截止频率。

[0542] 可替代地或者附加地，如图29所示，音频分类器200可以包括：帧级特征提取器2012，用于从音频帧提取帧级特征；以及帧级分类器2014，用于使用相应的帧级特征来将该音频帧序列中的每个帧分类到帧级音频类型中，其中，短期特征提取器2022可以被配置成基于帧级分类器2014 关于该音频帧序列的结果来计算短期特征。

[0543] 换言之，除了音频内容分类器202和音频上下文分类器204之外，音频分类器200还可以包括帧分类器201。在这样的架构中，音频内容分类器202基于帧分类器201的帧级分类结果来对短期片段进行分类，并且音频上下文分类器204基于音频内容分类器202的短期分类结果来对长期片段进行分类。

[0544] 帧级分类器2014可以被配置成将该音频帧序列中的每个音频帧分类到任何的可以被称为“帧级音频类型”的类中。在一个实施方式中，帧级音频类型可以具有与在上文中所论述的内容类型的架构类似的架构，并且也具有与内容类型类似的含义，唯一的差异是帧级音频类型和内容类型是在音频信号的不同级别——即，帧级别和短期片段级别——分类的。例如，帧级分类器2014可以被配置成将该音频帧序列的每个帧分类到以下帧级音频类型中的至少一个类型中：语音、音乐、背景声音和噪声。另一方面，帧级音频类型还可以具有与内容类型的架构部分地不同或者完全不同的架构，更适于帧级分类，并且更适于用作针对短期分类的短期特征。例如，帧级分类器2014可以被配置成将该音频帧序列的每个帧分类到以下帧级音频类型中的至少一个帧级音频类型中：浊音、清音和停顿。

[0545] 关于如何从帧级分类的结果提取短期特征，可以通过参考小节6.2中的描述来采用类似的方案。

[0546] 作为替代方式，短期分类器2024既可以使用基于帧级分类器2014的结果的短期特征，也可以使用直接基于从帧级特征提取器2012获得的帧级特征的短期特征。因此，短期特征提取器2022可以被配置成基于从该音频帧序列提取的帧级特征以及帧级分类器关于该音频帧序列的结果两者来计算短期特征。

[0547] 换言之，帧级特征提取器2012可以被配置成计算与在小节6.2中论述的统计数据类似的统计数据以及结合图28描述的短期特征两者，其包括以下特征中的至少一个特征：表征各种短期音频类型的属性的特征、截止频率、静态信噪比特性、分段信噪比特性、基本语音描述特征和声道特性。

[0548] 为了实时地工作，在所有的实施方式中，短期特征提取器2022可以被配置成在使用以预定步长在长期音频片段的时间维度上滑行的滑窗所形成的短期音频片段上工作。关于用于短期音频片段的滑窗，以及音频帧和用于长期音频片段的滑窗，其细节可以参考小节1.1。

[0549] 6.4实施方式和应用场景的组合

[0550] 与第1部分类似，以上所述的所有实施方式和实施方式的变型可以以其任意的组合来实现，并且在不同的部分/实施方式中提到的但是具有相同或者类似功能的任何组件可以作为同一组件或者单独的组件来实现。

[0551] 例如，在小节6.1至小节6.3中所述的任何两个或者更多个解决方案可以彼此组

合。并且这些组合还可以与在第1部分至第5部分以及在稍后将要描述的其他部分中所描述的和所暗示的任何实施方式进行组合。特别地，在第1部分中所描述的类型平滑单元712可以用于本部分以作为音频分类器200的组件，用于使帧分类器2014或者音频内容分类器202或者音频上下文分类器204的结果平滑。另外，计时器916也可以用作音频分类器200的组件，以避免音频分类器200的输出的突变。

[0552] 6.5音频分类方法

[0553] 与第1部分类似，在描述上文实施方式中的音频分类器的过程中，显然也公开了一些过程和方法。在下文中，在不重复已经论述的细节的情况下给出这些方法的概要。

[0554] 如图30所示，在一个实施方式中，提供了音频分类方法。为了识别包括短期音频片段序列(彼此重叠或者不重叠)的长期音频片段的长期音频类型(即，上下文类型)，首先将短期音频片段分类到短期音频类型，即内容类型，并且通过计算针对该长期音频片段中的短期片段序列的分类操作的结果的统计数据(操作3006)来获得长期特征。然后，可以使用长期特征来进行长期分类(操作3008)。短期音频片段可以包括音频帧序列。当然，为了识别短期片段的短期音频类型，需要从短期片段提取短期特征 (操作3002)。

[0555] 短期音频类型(内容类型)可以包括但不限于语音、短期音乐、背景声音和噪声。

[0556] 长期特征可以包括但不限于：短期音频类型的置信度值的平均值和方差、由短期片段的重要度进行加权的上述平均值和方差、每个短期音频类型的出现频率和在不同短期音频类型之间转换的频率。

[0557] 如图31所示，在变型中，可以直接基于长期音频片段中的短期片段序列的短期特征获得另外的长期特征。这种另外的长期特征可以包括但不限于以下短期特征统计数据：平均值、方差、加权平均值、加权方差、高平均、低平均、以及高平均与低平均的比率(对比度)。

[0558] 有不同的方式来提取短期特征。一个方式是从待分类的短期音频片段直接提取短期特征。这种特征包括但不限于：节奏特性、中断/静音特性和短期音频质量特征。

[0559] 第二种方式是从每个短期片段所包括的音频帧提取帧级特征(图32 中的操作3201)，然后基于帧级特征来计算短期特征，例如计算帧级特征的统计作为短期特征。帧级特征可以包括但不限于：表征各种短期音频类型的属性的特征、截止频率、静态信噪比特性、分段信噪比特性、基本语音描述特征和声道特性。表征各种短期音频类型的属性的特征还可以包括：帧能量、子带谱分布、谱通量、梅尔倒谱系数、低音、残余信息、音调色度特征和过零率。

[0560] 第三种方式是以与提取长期特征类似的方式来提取短期特征：在从待分类的短期片段中的音频帧提取帧级特征(操作3201)之后，使用相应的帧级特征将每个音频帧分类到帧级音频类型中(图33中的操作32011)；以及可以通过基于帧级音频类型(可选地包括置信度值)计算短期特征来提取短期特征(操作3002)。帧级音频类型可以具有与短期音频类型(内容类型)类似的属性和架构，并且也可以包括语音、音乐、背景声音和噪声。

[0561] 第二种方式和第三种方式可以组合在一起，如图33中的虚线箭头所示。

[0562] 如在第1部分所论述的，短期音频片段饿长期音频片段都可以用滑窗取样。也就是说，提取短期特征的操作(操作3002)可以在使用以预定步长在长期音频片段的时间维度上滑行的滑窗所形成的短期音频片段上进行，并且提取长期特征的操作(操作3107)和计算短

期音频类型的统计数据的操作(操作3006)也可以在使用以预定步长在音频信号的时间维度上滑行的滑窗所形成的长期音频片段上进行。

[0563] 与音频处理设备的实施方式类似地,一方面,音频处理方法的实施方式与实施方式的变型的任何组合都是可行的;另一方面,音频处理方法的实施方式和实施方式的变型的每个方面也可以是单独的解决方案。另外,在本小节中所描述的任何两个或者更多个解决方案可以彼此组合,并且这些组合还可以与在本公开内容的其他部分中所描述的和所暗示的任何实施方式进行组合。特别地,如已经在小节6.4中所论述的,音频类型的平滑方案和转换方案可以是这里所论述的音频分类方法的一部分。

[0564] 第7部分:VoIP分类器和分类方法

[0565] 在第6部分中提出了一种新颖的音频分类器,用于至少部分地基于内容类型分类器的结果将音频信号分类到音频上下文类型中。在第6部分所论述的实施方式中,从长度为数秒至数十秒的长期片段中提取长期特征,因此,音频上下文分类可能造成长的延迟时间。期望也可以实时地或近乎实时地,例如在短期片段级别,来分类音频上下文。

[0566] 7.1基于短期片段的上下文分类

[0567] 因此,如图34所示,提供了音频分类器200A,包括:音频内容分类 202A,用于识别音频信号的短期片段的内容类型;以及音频上下文分类器 204A,用于至少部分地基于由音频内容分类器识别的内容类型来识别短期片段的上下文类型。

[0568] 这里,音频内容分类器202A可以采取第6部分中已经提到的技术,但也可以采用不同的技术,如下面将要在小节7.2中讨论的。而且,音频上下文分类器204A可以采用第6部分已经提到的技术,不同之处在于上下文分类器204A可以直接使用音频内容分类器202A的结果,而不是使用音频内容分类器202A的结果的统计数据,因为音频上下文分类器204A 和音频内容的分类器202A都对同一短期片段进行分类。另外,与第6部分类似,除了来自音频内容分类器202A的结果之外,音频上下文分类器 204A可以使用从短期片段直接提取的其他特征。也就是说,音频上下文分类器204A可以被配置成基于通过使用短期片段内容类型置信度值作为特征以及从短期片段提取的其他特征的机器学习模型来对短期片段进行分类。关于从短期片段提取的特征,可以参考第6部分。

[0569] 音频内容分类器200A可以同时将短期片段标记为除了VoIP语音/噪声和/或非VoIP语音/噪声之外的更多的音频类型(VoIP语音/噪声和非 VoIP语音/噪声将在下文第7.2节讨论),并且多个音频类型中的每个音频类型都可以具有其自身的置信度值,如在第1.2节中所论述的。这可以实现更好的分类准确度,因为可以捕捉更丰富的信息。例如,语音和短期音乐的置信度值的联合信息显示出音频内容在何种程度上可能是语音和背景音乐的混合,以使得它能够与纯VoIP内容进行区分。

[0570] 7.2使用VoIP语音和VoIP噪声的分类

[0571] 本申请的这一方面在VoIP/非VoIP分类系统中尤其有用,在以短的决策延迟对当前短期片段进行分类时将需要该分类系统。

[0572] 为了这个目的,如图34所示,音频分类器200A是专门为VoIP/非 VoIP分类而设计的。为了对VoIP/非VoIP进行分类,开发了VoIP语音分类器2026和/或VoIP噪声分类器,以生成用于音频上下文分类器204A 的最终鲁棒的VoIP/非VoIP分类的中间结果。

[0573] VoIP短期片段要么包括VoIP语音要么包括VoIP噪声。观察到,将语音的短期片段

分类为VoIP语音或非VoIP语音是可以达到高准确度的，但是将噪声的短期片段分类为VoIP噪声或非VoIP噪声却并非如此。因此，可以得出结论：通过直接将短期片段分类为VoIP(包括VoIP语音和VoIP 噪声，但没有具体识别出VoIP语音和VoIP噪声)和非VoIP，而不考虑语音和噪声之间的区别，从而将这两种内容类型(语音和噪声)的特征混合在一起，将会降低鉴别力。

[0574] 对分类器来说，实现VoIP语音/非VoIP语音分类的准确度比VoIP噪声/非VoIP噪声分类的准确度更高是合理的，因为语音比噪声包含更多的信息，并且诸如截止频率等特征对语音的分类更为有效。根据从AdaBoost 训练过程获得的权重等级，针对VoIP/非VoIP语音分类的加权短期特征的前几位是：对数能量的标准偏差、截止频率、节奏强度的标准偏差和谱通量的标准偏差。VoIP语音的对数能量的标准偏差、节奏强度的标准偏差和谱通量的标准偏差通常高于非VoIP语音。一个可能的原因是，非VoIP上下文例如电影类媒体或游戏中的许多短期语音片段通常与上述特征的值较低的其他声音例如背景音乐或音效相混合。同时，VoIP语音的截止频率通常比非VoIP语音的截止频率低，这说明由许多流行的VoIP编解码器引入了较低的截止频率。

[0575] 因此，在一种实施方式中，音频内容分类器202A可以包括：VoIP语音分类器2026，用于将短期片段分类成VoIP语音内容类型或非VoIP语音内容类型；以及音频上下文分类器204A，可以被配置成基于VoIP语音和非VoIP语音的置信度值将短期片段分类到VoIP上下文类型或非VoIP 上下文类型中。

[0576] 在另一个实施方式中，音频内容分类器202A还可以包括：VoIP噪声分类器2028，用于将短期片段分类到VoIP噪声内容类型或非VoIP噪声内容类型中；以及音频上下文分类器204A，可以被配置成基于VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声的置信度值将短期片段分类到VoIP 上下文类型或非VoIP上下文类型中。

[0577] 如第6部分、小节1.2和小节7.1中论述的，VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声的内容类型可以用现有技术进行识别。

[0578] 或者，音频内容分类器202A可以具有如图35所示的层次结构。也就是说，利用语音/噪声分类器2025的结果以首先将短期片段分类到语音或噪声/背景中。

[0579] 在仅使用VoIP语音分类器2026的实施方式的基础上，如果一个短期片段被语音/噪声分类器2025(在这种情况下它只是语音分类器)确定为语音，则VoIP语音分类器2026继续区分其是VoIP语音还是非VoIP语音，并且计算出二元分类结果；否则，可以认为VoIP语音的置信度值低或者认为对VoIP语音的判定不确定。

[0580] 在仅使用VoIP噪声分类器2028的实施方式的基础上，如果一个短期片段被语音/噪声分类器2025(在这种情况下它只是噪声(背景)分类器) 确定为噪声，则VoIP噪声分类器2028继续区分其是VoIP噪声还是非VoIP 噪声，并且计算出二元分类结果。否则，可以认为VoIP噪声的置信度值低或者认为对VoIP噪声的判定不确定。

[0581] 这里，因为通常语音是信息性的内容类型而噪声/背景是干扰性内容类型，即使短期片段不是噪声，在前面段落中的实施方式中也无法绝对确定该短期片段不是VoIP上下文类型。而如果一个短期片段不是语音，在仅使用VolP语音分类器2026的实施方式中，其可能不是VolP上下文类型。因此，通常仅使用VoIP语音分类器2026的实施方式可以独立地实现，而仅使用VoIP噪声分类器2028的其他实施方式可能被用作补充实施方式，与例如使用VoIP

语音分类器2026的实施方式协作。

[0582] 也就是说，可以使用VoIP语音分类器2026和VoIP噪声分类器2028 两者。如果短期片段被语音/噪声分类器2025确定为语音，则VoIP语音分类器2026继续区分是VoIP语音还是非VoIP语音，并且计算出二元分类结果。如果短期片段被语音/噪声分类器2025确定为噪声，则VoIP噪声分类器2028继续区分是VoIP噪声还是非VoIP噪声，并且计算出二元分类结果。否则，可以认为短期片段可以被分类为非VoIP。

[0583] 语音/噪声分类器2025、VoIP语音分类器2026和VoIP噪声分类器2028 的实现可以采用任何现有技术，也可以是第1部分至第6部分中讨论的音频内容分类器202。

[0584] 如果根据以上描述实现的音频内容分类器202A最终将一个短期片段没有分类到语音、噪声和背景中，或者没有分类到VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声中，意味着所有的相关置信度值都低，则音频内容分类器202A(和音频上下文分类器204A)将该短期片段分类为非 VoIP。

[0585] 为了基于VoIP语音分类器2026和VoIP噪声分类器2028的结果将短期片段分类到VoIP或非VoIP的上下文类型中，音频上下文分类器204A 可以采用如小节7.1中讨论的基于机器学习的技术，并且作为一种改进，可以使用更多特征，包括从短期片段直接提取的短期特特征和/或针对除了 VoIP相关内容类型之外的其他内容类型的其它音频内容分类器的结果，如小节7.1中所论述的。

[0586] 除了上述基于机器学习的技术之外，VoIP/非VoIP分类的可替代途径可以是利用领域知识并且利用与VoIP语音和VoIP噪声相关的分类结果的启发式规则。这种启发式规则的一个示例如下所示。

[0587] 如果时间t的当前短期片段被确定为VoIP语音或非VoIP语音，该分类结果直接被当作VoIP/非VoIP分类结果，因为VoIP/非VoIP语音分类是鲁棒的，如前面所讨论的。也就是说，如果短期片段被确定为VoIP语音，那么它是VoIP上下文类型；如果短期片段被确定为非VoIP语音，那么它是非VoIP上下文类型。

[0588] 当VoIP语音分类器2026针对如上面提到的由语音/噪声分类器2025 确定的语音作出关于VoIP语音/非VoIP语音的二元决策时，VoIP语音和非VoIP的语音置信度值可能是互补的，即其总和为1(如果0表示100％不是，1表示100％是)，并且用于区分VoIP语音和非VoIP语音的置信度值的阈值可以实际上表示同一点。如果VoIP语音分类器2026不是二元分类器，则VoIP语音和非VoIP语音的置信度值可能不是互补的，并且用于区分VoIP语音和非VoIP语音的置信度值的阈值可以不一定表示同一点。

[0589] 但是，在VoIP语音或非VoIP语音置信度接近阈值并在阈值附近上下波动的情况下，VoIP/非VoIP分类结果可能过于频繁地切换。为了避免这种波动，可以提供缓冲方案：VoIP语音的阈值和非VoIP语音的阈值两者都可以设定得更大，使得不那么容易从当前内容类型切换到另一内容类型。为了便于描述，可以将非VoIP语音的置信度值转换成VoIP语音的置信度值。也就是，如果置信度值高，则认为短期片段是接近VoIP语音，而如果置信度值低，则认为短期片段接近非VoIP语音。虽然对于上述非二元分类器来说，非VoIP语音的高置信度值不一定意味着VoIP语音的低置信度值，但是这种简化可以很好地反映该解决方案的本质，并且以二元分类器的语言描述的相关权利要求应被解释为涵盖非二元分类器的等同解决方案。

[0590] 缓冲方案如图36所示。在两个阈值Th1和Th2(Th1>=Th2)之间有缓冲区域。当VoIP语音的置信值v(t)落在该区域中时,上下文分类不会改变,如图36中的左右两侧箭头所示。仅当置信值v(t)大于较大的阈值 Th1时,短期片段将被分类为VoIP(如图36中的下部箭头所示);并且仅当置信度值不大于较小的阈值Th2时,短期片段将被分类为非VoIP(如图36中上部箭头所示)。

[0591] 如果替代地使用VoIP噪声分类器2028,则情况类似。为了使解决方案更具鲁棒性,可以结合使用VoIP语音分类器2026和VoIP噪声分类器 2028。然后,音频上下文分类器204A可以被配置成:如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值,则将短期片段分类为VoIP上下文类型;如果VoIP语音的置信度值不大于第二阈值,其中第二阈值不大于第一阈值,或者如果VoIP噪声的置信度值不大于第四阈值,其中第四阈值不大于第三阈值,则将短期片段分类为非 VoIP上下文类型;否则,将短期片段分类为上一个短期片段的上下文类型。

[0592] 这里,第一阈值可以等于第二阈值,并且第三阈值可以等于第四阈值,尤其是针对但不限于二元VoIP语音分类器和二元VoIP噪声分类器。但是,由于VoIP噪声分类结果通常鲁棒性不佳,所以如果第三和第四阈值彼此不相等将会更好,并且二者应远离0.5(0表示是非VoIP噪声的高置信度, 1表示是VoIP噪声的高置信度)。

[0593] 7.3使波动平滑

[0594] 为了避免快速波动,另一种解决方案是对音频内容分类器所确定的置信度值进行平滑。因此,如图37所示,类型平滑单元203A可被包含在音频分类器200A中。对于前面所讨论的4个VoIP相关的内容类型中任意一个内容类型的置信度值,可以采用1.3节讨论的平滑方案。

[0595] 或者,与小节7.2类似,VoIP语音和非VoIP语音可被视为具有互补置信度值的对;VoIP噪声和非VoIP噪声也可被视为具有互补置信度值的对。在这种情况下,每对中只有一个需要进行平滑,可以采用小节1.3中讨论的平滑方案。

[0596] 以VoIP语音的置信度值为例,公式(3)可重写为:

[0597] $v(t)=\beta \cdot v(t-1)+(1-\beta) \cdot voipSpeechConf(t)$ (3")

[0598] 其中,v(t)是时刻t(当次)的经平滑的VoIP语音置信度值,v(t-1)是上一个时刻(上次)的经平滑的VoIP语音置信度值,而VoipSpeechConf是当前时刻t的在平滑之前的VoIP语音置信度,α是加权系数。

[0599] 在一个变型中,如果有上述的语音/噪声分类器2025,如果短片段的语音置信度值低,则该短期片段不能被鲁棒地分类成VoIP语音,并且可以直接设定VoipSpeechConf(t)=v(t-1)而无需使VoIP语音分类器2026实际工作。

[0600] 或者,在上述情形中,可以设定VoipSpeechConf(t)=0.5(或其他不大于0.5的值,比如0.4-0.5),表示不确定的情况(这里置信度=1表示其 VoIP的高置信度,置信度=0表示非VoIP的高置信度)。

[0601] 因此,根据如图37所示的变型,音频内容分类器200A还可以包括:语音/噪声分类器2025,用于识别短期片段的语音内容类型;以及类型平滑单元203A,可以被配置成,在由语音/噪声分类器分类的语音内容类型的置信度值低于第五阈值的情况下,将当前短期片段的平滑之前的VoIP 语音置信度值设定为预定的置信度值(如0.5或其他值,如0.4-0.5)

或上一个短期片段的经平滑的置信度值。在这种情况下，VoIP语音分类器2026 可以工作也可以不工作。或者，可以通过VoIP语音分类器2026来设定置信度值，这等同于通过类型平滑单元203A来设定置信度值的解决方案，并且权利要求应该被解释为涵盖这两种情况。此外，在这里使用语句“由语音/噪声分类器分类的语音内容类型的置信度值低于第五阈值”，但保护的范围不限于此，并且它等同于短期片段被分类到除语音之外的其他内容类型中的情况。

[0602] 对于VoIP噪声的置信度值，情况类似并且这里省略详细描述。

[0603] 为了避免快速波动，又一种解决方案是对音频上下文分类器204A所确定的置信度值进行平滑，可以采用小节1.3中所讨论的平滑方案。

[0604] 为了避免快速波动，再一种解决方案是延迟上下文类型在VoIP和非 VoIP之间的转换，可以使用与小节1.6中所述的方案相同的方案。如小节 1.6所述，计时器916可在音频分类器的外部或在音频分类器的内部作为其一部分。因此，如图38所示，音频分类器200A还可以包括计时器916。并且音频分类器被配置成连续地输出当前上下文类型直至新的上下文类型的持续时间的长度达到第六阈值(上下文类型是音频类型的一个实例)。通过参考小节1.6，在此省略详细描述。

[0605] 附加地或可替代地，作为延迟VoIP和非VoIP之间的转换的另一种方案，如前所述的针对VoIP/非VoIP分类的第一和/或第二阈值可以取决于上一个短期片段的上下文类型而不同。也就是说，当新的短期片段的上下文类型不同于上一短期片段的上下文类型时，第一阈值和/或第二阈值变得更大；当新的短期片段的上下文类型与上一短期片段的上下文类型相同时，第一阈值和/或第二阈值变得更小。通过这种方式，上下文类型趋于保持在当前上下文类型，从而可以在一定程度上抑制上下文类型的突然波动。

[0606] 7.4实施方式和应用场景的组合

[0607] 与第1部分类似，以上所述的所有实施方式和实施方式的变型可以以其任意的组合来实现，并且在不同的部分/实施方式中所提到的但是具有相同或者类似功能的任何组件可以作为同一组件或者单独的组件来实现。

[0608] 例如，在小节7.1至小节7.3中所述的任何两个或者更多个解决方案可以彼此组合。并且这些组合还可以与在第1部分至第6部分所描述的和所暗示的任何实施方式进行组合。特别地，在本部分所论述的实施方式和其任意组合可以与音频处理设备/方法或者与在第4部分论述的音量校平器控制器/控制方法进行组合。

[0609] 7.5VoIP分类方法

[0610] 与第1部分类似，在描述上文实施方式中的音频分类器的过程中，显然也公开了一些过程和方法。在下文中，在不重复已经论述的细节的情况下给出这些方法的概要。

[0611] 在图39所示的一个实施方式中，音频分类方法包括：识别音频信号的短期片段的内容类型(操作4004)，然后至少部分地基于所识别的内容类型来识别短期片段的上下文类型(操作4008)。

[0612] 为了动态地并且快速地识别音频信号的上下文类型，本部分中的音频分类方法对于识别VoIP上下文类型和非VoIP上下文类型特别有用。在这样的情形下，首先可以将短期片段分类到VoIP语音内容类型或者非VoIP 语音内容类型中，并且识别上下文类型的操作被配置成基于VoIP语音和非VoIP语音的置信度值，将短期片段分类到VoIP上下文类型或者

非VoIP 上下文类型中。

[0613] 或者，可以首先将短期片段分类到VoIP噪声内容类型或者非VoIP噪声内容类型中，并且识别上下文类型的操作被配置成基于VoIP噪声和非 VoIP噪声的置信度值，将短期片段分类到VoIP上下文类型或者非VoIP 上下文类型中。

[0614] 可以联合考虑语音和噪声。在这种情形下，识别上下文类型的操作可以被配置成基于VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声的置信度值，将短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

[0615] 为了识别短期片段的上下文类型，可以使用机器学习模型，将短期片段的内容类型的置信度值和从短期片段提取出来的其他特征都作为特征。

[0616] 也可以基于启发式规则来实现识别上下文类型的操作。当只涉及VoIP 语音和非VoIP语音时，启发式规则是这样的：如果VoIP语音的置信度值大于第一阈值，则将短期片段分类成VoIP上下文类型；如果VoIP语音的置信度值不大于第二阈值，则将短期片段分类成非VoIP上下文类型，其中第二阈值不大于第一阈值；否则，将短期片段分类成上一个短期片段的上下文类型。

[0617] 用于只涉及VoIP噪声和非VoIP噪声的情形的启发式规则是类似的。

[0618] 当既涉及语音和噪声两者时，启发式规则是这样的：如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值，则将短期片段分类成VoIP上下文类型；如果VoIP语音的置信度值不大于第二阈值，其中第二阈值不大于第一阈值，或者如果VoIP噪声的置信度值不大于第四阈值，其中第四阈值不大于第三阈值，则将短期片段分类成非VoIP上下文类型；否则，将短期片段分类成上一个短期片段的上下文类型。

[0619] 这里可以采用在小节1.3和小节1.8中所论述的平滑方案并且省略详细描述。作为在小节1.3中所述的平滑方案的修改，在平滑操作4106之前，该方法还可以包括根据短期片段识别语音内容类型(图40中的操作 40040)，其中，在语音内容类型的置信度值低于第五阈值(操作40041中的“N”)的情况下，当前短期片段的平滑之前的VoIP语音置信度值被设定成预定的置信度值或者上一个短期片段的经平滑的置信度值(图40中的操作40044)。

[0620] 否则，如果识别语音内容类型的操作鲁棒地判断出该短期片段是语音 (操作40041中的“Y”)，则在平滑操作4106之前，该短期片段进一步被分成VoIP语音或者非VoIP语音(操作40042)。

[0621] 实际上，即使不使用平滑方案，该方法也可以首先识别出语音内容类型和/或噪声内容类型，当短期片段被分类成语音或者噪声时，实现进一步的分类以将短期片段分类到VoIP语音和非VoIP语音之一，或者分类到 VoIP噪声和非VoIP噪声之一。然后进行识别上下文类型的操作。

[0622] 如在小节1.6和小节1.8中所提到的，其中所论述的转换方案可以作为这里所描述的音频分类方法的一部分，并且这里省略其细节。简言之，该方法还可以包括对识别上下文类型的操作连续地输出同一上下文类型的持续时间进行测量，其中音频分类方法被配置成继续输出当前上下文类型直到新的上下文类型的持续时间达到第六阈值。

[0623] 类似地，可以针对不同的从一个上下文类型到另一个上下文类型的转换对来设定不同的第六阈值。此外，可以使第六阈值与新的上下文类型的置信度值负相关。

[0624] 作为特别针对VoIP/非VoIP分类的音频分类方法中的转换方案的改进，可以将针

对当前短期片段的第一阈值至第四阈值的一个或者更多个设置成取决于上一个短期片段的上下文类型而不同。

[0625] 与音频处理设备的实施方式类似，一方面，音频处理方法的实施方式与实施方式的变型的任何组合都是可行的；另一方面，音频处理方法的实施方式和实施方式的变型的每个方面也可以是单独的解决方案。另外，在本小节中所描述的任何两个或者更多个解决方案可以彼此组合，并且这些组合还可以与在本公开内容的其他部分中所描述的和所暗示的任何实施方式进行组合。特别地，这里所描述的音频分类方法可以用于之前所描述的音频处理方法，尤其是音量校平器控制方法。

[0626] 如在本申请的“具体实施方式”的开始所描述的，本申请的实施方式可以实施为硬件或者软件，或者两者。图41是示出了用于实现本申请的各方面的示例性系统的框图。

[0627] 在图41中，中央处理单元(CPU)4201根据存储在只读存储器(ROM) 4202中的程序或者从存储部分4208加载到随机存取存储器(RAM)4203 的程序来进行各种处理。在RAM 4203中，也根据需要存储了当CPU 4201 进行各种处理等时所需的数据。

[0628] CPU 4201、ROM 4202和RAM 4203通过总线4204彼此连接。输入/ 输出接口4205也连接至总线4204。

[0629] 以下组件连接至输入/输出接口4205：包括键盘、鼠标等的输入部分 4206；包括例如阴极射线管(CRT)、液晶显示器(LCD)等的显示器和扬声器等的输出部分4207；包括硬盘等的存储部分4208；以及包括例如 LAN卡、调制解调器等的网络接口卡的通信部分4209。通信部分4209通过网络例如因特网来进行通信处理。

[0630] 根据需要，驱动器4210也连接至输入/输出接口4205。可移除介质4211 例如磁盘、光盘、磁光盘、半导体存储器等根据需要而安装到驱动器4210，以使得从其上读取的计算机程序根据需要被安装到存储部分4208中。

[0631] 在通过软件实现上述组件的情况下，从网络例如因特网或者从存储介质例如可移除介质4211来安装构成该软件的程序。

[0632] 请注意，这里所用的术语仅仅是为了描述具体实施方式的目的，而并非意图限制本申请。如这里所用到的，除非上下文已经明确地另外指定，单数形式的“一”和“该”意指也包括复数形式。还要理解的是，当在本申请中使用术语“包括”时，说明存在所指的特征、整体、操作、步骤、元素和/或组件，但是不排除存在或者增加一个或者更多个其他特征、整体、操作、步骤、元素、组件和/或他们的组合。

[0633] 以下权利要求中的相应结构、材料、操作以及所有“装置或操作加功能”元素的等同替换，旨在包括任何用于与在权利要求中具体指出的其它元素相组合地执行该功能的结构、材料或操作。仅出于图解和描述的目的而给出对本申请的描述，而并非穷进行的的或限于所公开的应用。对于所属技术领域的普通技术人员而言，在不偏离本发明范围和精神的情况下，显然可以作出许多修改和变型。对实施例的选择和说明，是为了最好地解释本发明的原理和实际应用，并使所属技术领域的普通技术人员能够理解本申请，可以有适合所设想的特定用途的具有各种改变的各种实施例。

[0634] 根据以上，可以看出描述了以下示例性的实施方式(每个均用“EE”表示)。

[0635] 设备实施方式：

[0636] EE.1.一种音量校平器控制器，包括：

[0637] 音频内容分类器，用于实时地识别音频信号的内容类型；以及

[0638] 调整单元，用于基于所识别的内容类型来以连续的方式调整音量校平器；

[0639] 其中，所述调整单元被配置成使所述音量校平器的动态增益与所述音频信号的信息性内容类型正相关，并且使所述音量校平器的动态增益与所述音频信号的干扰性内容类型负相关。

[0640] EE.2.根据EE 1所述的音量校平器控制器，其中，所述音频信号的所述内容类型包括语音、短期音乐、噪声和背景声音中的一个。

[0641] EE.3.根据EE 1所述的音量校平器控制器，其中，噪声被视为干扰性内容类型。

[0642] EE.4.根据EE 1所述的音量校平器控制器，其中，所述调整单元被配置成基于所述内容类型的置信度值来调整所述音量校平器的动态增益。

[0643] EE.5.根据EE 4所述的音量校平器控制器，其中，所述调整单元被配置成通过所述内容类型的置信度值的传递函数来调整所述动态增益。

[0644] EE.6.根据EE 1所述的音量校平器控制器，其中，所述音频内容分类器被配置成将所述音频信号分类到具有相应置信度值的多个内容类型中，并且所述调整单元被配置成通过基于所述多个内容类型的重要性对所述多个内容类型的置信度值进行加权来考虑所述多个内容类型中的至少一些内容类型。

[0645] EE.7.根据EE 1所述的音量校平器控制器，其中，所述音频内容分类器被配置成将所述音频信号分类到具有相应置信度值的多个内容类型中，并且所述调整单元被配置成使用至少一个其他内容类型的置信度值来修改一个内容类型的权重。

[0646] EE.8.根据EE 1所述的音量校平器控制器，其中，所述音频内容分类器被配置成将所述音频信号分类到具有相应置信度值的多个内容类型中，并且所述调整单元被配置成通过基于所述置信度值对所述多个内容类型的影响进行加权来考虑所述多个内容类型中的至少一些内容类型。

[0647] EE.9.根据EE 8所述的音量校平器控制器，其中，所述调整单元被配置成基于所述置信度值来考虑至少一个主导的内容类型。

[0648] EE.10.根据EE 9所述的音量校平器控制器，其中，所述音频内容分类器被配置成将所述音频信号分类到具有相应置信度值的多个干扰性内容类型中和/或具有相应置信度值的多个信息性内容类型中，并且所述调整单元被配置成基于所述置信度值来考虑至少一个主导的干扰性内容类型和/ 或至少一个主导的信息性内容类型。

[0649] EE.11.根据EE 1至EE 10中任一项所述的音量校平器控制器，还包括类型平滑单元，用于针对每个内容类型，基于所述音频信号的过去的置信度值来对所述音频信号的当次置信度值进行平滑。

[0650] EE.12.根据EE 11所述的音量校平器控制器，其中，所述类型平滑单元被配置成通过计算当前的实际置信度值与上一次的经平滑的置信度值的加权和来确定所述音频信号当次的经平滑的置信度值。

[0651] EE.13.根据EE 1至EE 10中任一项所述的音量校平器控制器，还包括音频上下文分类器，用于识别所述音频信号的上下文类型，其中，所述调整单元被配置成基于所述上下文类型的置信度值来调整所述动态增益的范围。

[0652] EE.14.根据EE 1至EE 10中任一项所述的音量校平器控制器，还包括音频上下文

分类器，用于识别所述音频信号的上下文类型，其中，所述调整单元被配置成基于所述音频信号的所述上下文类型来将所述音频信号的所述内容类型视为信息性的或者是干扰性的。

[0653] EE.15.根据EE 14所述的音量校平器控制器，其中，所述音频信号的所述上下文类型包括VoIP、电影类媒体、长期音乐和游戏中的一个。

[0654] EE.16.根据EE 14所述的音量校平器控制器，其中，在VoIP上下文类型的音频信号中，所述背景声音被视为干扰性内容类型；而在非VoIP上下文类型的音频信号中，所述背景声音和/或语音和/或音乐被视为信息性内容类型。

[0655] EE.17.根据EE 14所述的音量校平器控制器，其中，所述音频信号的所述上下文类型包括高质量音频或低质量音频。

[0656] EE.18.根据EE 14所述的音量校平器控制器，其中，取决于音频信号的上下文类型，给不同上下文类型的音频信号中的内容类型分配不同的权重。

[0657] EE.19.根据EE 14所述的音量校平器控制器，其中，所述音频上下文分类器被配置成将所述音频信号分类到具有相应置信度值的多个上下文类型中，并且所述调整单元被配置成通过基于所述多个上下文类型的重要性对所述多个上下文类型的置信度值进行加权来考虑所述多个上下文类型中的至少一些上下文类型。

[0658] EE.20.根据EE 14所述的音量校平器控制器，其中，所述音频上下文分类器被配置成将所述音频信号分类到具有相应置信度值的多个上下文类型中，并且所述调整单元被配置成通过基于所述置信度值对所述多个上下文类型的影响进行加权来考虑所述多个上下文类型中的至少一些上下文类型。

[0659] EE.21.根据EE 14所述的音量校平器控制器，其中，

[0660] 所述音频内容分类器被配置成按所述音频信号的短期片段来识别所述内容类型；

[0661] 所述音频上下文分类器被配置成至少部分地基于由所述音频内容分类器识别的所述内容类型来按所述音频信号的短期片段识别所述上下文类型。

[0662] EE.22.根据EE 21所述的音量校平器控制器，其中，所述音频内容分类器包括VoIP语音分类器，用于将短期片段分类到VoIP语音内容类型或者非VoIP语音内容类型中；并且

[0663] 所述音频上下文分类器被配置成基于VoIP语音和非VoIP语音的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

[0664] EE.23.根据EE 22所述的音量校平器控制器，其中，所述音频内容分类器还包括：

[0665] VoIP噪声分类器，用于将所述短期片段分类到VoIP噪声内容类型或非VoIP噪声内容类型中；以及

[0666] 所述音频上下文分类器被配置成基于VoIP语音、非VoIP语音、VoIP 噪声和非VoIP噪声的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

[0667] EE.24.根据EE 22所述的音量校平器控制器，其中，所述音频上下文分类器被配置成：

[0668] 如果VoIP语音的置信度值大于第一阈值，则将所述短期片段分类成所述VoIP上下文类型；

[0669] 如果VoIP语音的置信度值不大于第二阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值；否则

[0670] 将所述短期片段分类成上一个短期片段的上下文类型。

[0671] EE.25.根据EE 23所述的音量校平器控制器，其中，所述音频上下文分类器被配置成：

[0672] 如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值，则将所述短期片段分类成所述VoIP上下文类型，

[0673] 如果VoIP语音的置信度值不大于第二阈值，或者如果VoIP噪声的置信度值不大于第四阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值，所述第四阈值不大于所述第三阈值；否则

[0674] 将所述短期片段分类成上一个短期片段的上下文类型。

[0675] EE.26.根据EE 21至EE 25中任一项所述的音量校平器控制器，还包括类型平滑单元，用于基于所述内容类型的过去的置信度值来对所述内容类型的当次置信度值进行平滑。

[0676] EE.27.根据EE 26所述的音量校平器控制器，其中，所述类型平滑单元被配置成通过计算当前短期片段的置信度值与上一个短期片段的经平滑的置信度值的加权和来确定当前短期片段的经平滑的置信度值。

[0677] EE.28.根据EE 26所述的音量校平器控制器，其中，

[0678] 所述音频内容分类器还包括语音/噪声分类器，用于识别所述短期片段的语音内容类型，并且所述类型平滑单元被配置成，在由所述语音/噪声分类器分类的所述语音内容类型的置信度值低于第五阈值的情况下，将平滑之前的当前短期片段的VoIP语音的置信度值设定为预定的置信度值或者上一个短期片段的经平滑的置信度值。

[0679] EE.29.根据EE 22或EE 23所述的音量校平器控制器，其中，所述音频上下文分类器被配置成通过使用所述短期片段的所述内容类型的置信度值以及从短期片段提取的其他特征作为特征，基于机器学习模型对所述短期片段进行分类。

[0680] EE.30.根据EE 14至EE 29中任一项所述的音量校平器控制器，还包括计时器，用于测量所述音频上下文分类器连续地输出同一上下文类型的持续时间，其中，所述调整单元被配置成继续使用当前的上下文类型直到新的上下文类型的持续时间的长度达到第六阈值为止。

[0681] EE.31.根据EE 30所述的音量校平器控制器，其中，针对不同的从一个上下文类型到另一个上下文类型的转换对来设定不同的所述第六阈值。

[0682] EE.32.根据EE 30所述的音量校平器控制器，其中，所述第六阈值与所述新的上下文类型的置信度值负相关。

[0683] EE.33.根据EE 24或25所述的音量校平器控制器，其中，所述第一阈值和/或所述第二阈值取决于上一个短期片段的上下文类型而不同。

[0684] EE.34.一种音频处理设备，包括根据EE1至EE33中任一项所述的音量校平器控制器。

[0685] EE.35.一种音频分类器，包括：

[0686] 音频内容分类器，用于识别音频信号的短期片段的内容类型；以及

[0687] 音频上下文分类器，用于至少部分地基于由所述音频内容分类器识别的所述内容类型来识别所述短期片段的上下文类型。

[0688] EE.36.根据EE 35所述的音频分类器，其中，所述音频内容分类器包括 VoIP语音

分类器，用于将所述短期片段分类到VoIP语音内容类型或者非 VoIP语音内容类型中；并且

[0689] 所述音频上下文分类器被配置成基于VoIP语音和非VoIP语音的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。EE.37.根据EE 36所述的音频分类器，其中，所述音频内容分类器还包括：

[0690] VoIP噪声分类器，用于将所述短期片段分类到VoIP噪声内容类型或者非VoIP噪声内容类型中；并且

[0691] 所述音频上下文分类器被配置成基于VoIP语音、非VoIP语音、VoIP 噪声和非VoIP噪声的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

[0692] EE.38.根据EE 37所述的音频分类器，其中，所述音频上下文分类器被配置成：

[0693] 如果VoIP语音的置信度值大于第一阈值，则将所述短期片段分类成所述VoIP上下文类型；

[0694] 如果VoIP语音的置信度值不大于第二阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值；否则

[0695] 将所述短期片段分类成上一个短期片段的所述上下文类型。

[0696] EE.39.根据EE 37所述的音频分类器，其中，所述音频上下文分类器被配置成：

[0697] 如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值，则将所述短期片段分类为VoIP上下文类型；

[0698] 如果VoIP语音的置信度值不大于第二阈值，或者如果VoIP噪声的置信度值不大于第四阈值，则将所述短期片段分类为非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值，所述第四阈值不大于所述第三阈值；否则

[0699] 将所述短期片段分类为上一个短期片段的上下文类型。

[0700] EE.40.根据EE 35至EE 39中任一项所述的音频分类器，还包括类型平滑单元，用于基于所述内容类型的过去的置信度值来对所述内容类型的当次置信度值进行平滑。

[0701] EE.41.根据EE 40所述的音频分类器，其中，所述类型平滑单元被配置成通过计算当前短期片段的置信度值与上一个短期片段的经平滑的置信度值的加权和来确定所述当前短期片段的经平滑的置信度值。

[0702] EE.42.根据EE 41所述的音频分类器，其中，所述音频内容分类器还包括语音/噪声分类器，用于从所述短期片段识别语音内容类型，并且所述类型平滑单元被配置成，在由所述语音/噪声分类器分类的所述语音内容类型的置信度值低于第五阈值的情况下，将平滑之前的当前短期片段的 VoIP语音的置信度值设定为预定的置信度值或者上一个短期片段的经平滑的置信度值。

[0703] EE.43.根据EE 36或者EE 37所述的音频分类器，其中，所述音频上下文分类器被配置成：通过使用所述短期片段的所述内容类型的置信度值以及从所述短期片段提取的其他特征作为特征，基于机器学习模型对所述短期片段进行分类。

[0704] EE.44.根据EE 38或者EE 39所述的音频分类器，还包括计时器，用于测量所述音频上下文分类器连续地输出同一上下文类型的持续时间，其中，所述调整单元被配置成继续使用当前的上下文类型直到新的上下文类型的持续时间的长度达到第六阈值为止。

[0705] EE.45.根据EE 44所述的音频分类器，其中，针对不同的从一个上下文类型到另一个上下文类型的转换对来设定不同的所述第六阈值。

[0706] EE.46.根据EE 44所述的音频分类器，其中，所述第六阈值与所述新的上下文类型的置信度值负相关。

[0707] EE.47.根据EE 38或39所述的音频分类器，其中，所述第一阈值和/或所述第二阈值取决于上一个短期片段的上下文类型而不同。

[0708] EE.48.一种音频处理设备，包括根据EE 35至EE 47中任一项所述的音频分类器。

[0709] 方法实施方式：

[0710] EE.1.一种音量校平器控制方法，包括：

[0711] 实时地识别音频信号的内容类型；以及

[0712] 通过使所述音量校平器的动态增益与所述音频信号的信息性内容类型正相关，并且使所述音量校平器的动态增益与所述音频信号的干扰性内容类型负相关，来基于所识别的内容类型以连续的方式调整音量校平器。

[0713] EE.2.根据EE 1所述的音量校平器控制方法，其中，所述音频信号的所述内容类型包括语音、短期音乐、噪声和背景声音中的一个。

[0714] EE.3.根据EE 1所述的音量校平器控制方法，其中，噪声被视为干扰性内容类型。

[0715] EE.4.根据EE 1所述的音量校平器控制方法，其中，所述调整的操作被配置成基于所述内容类型的置信度值来调整所述音量校平器的动态增益。

[0716] EE.5.根据EE 4所述的音量校平器控制方法，其中，所述调整的操作被配置成通过所述内容类型的置信度值的传递函数来调整所述动态增益。

[0717] EE.6.根据EE 1所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个内容类型中，并且所述调整的操作被配置成通过基于所述多个内容类型的重要性对所述多个内容类型的置信度值进行加权来考虑所述多个内容类型中的至少一些内容类型。

[0718] EE.7.根据EE 1所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个音频内容中，并且所述调整的操作被配置成使用至少一个其他内容类型的置信度值来修改一个内容类型的权重。

[0719] EE.8.根据EE 1所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个内容类型中，并且所述调整的操作被配置成通过基于所述置信度值对所述多个内容类型的影响进行加权来考虑所述多个内容类型中的至少一些内容类型。

[0720] EE.9.根据EE 8所述的音量校平器控制方法，其中，所述调整的操作被配置成基于所述置信度值来考虑至少一个主导的内容类型。

[0721] EE.10.根据EE 8所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个干扰性内容类型中和/或具有相应置信度值的多个信息性内容类型中，并且所述调整的操作被配置成基于所述置信度值来考虑至少一个主导的干扰性内容类型和/或至少一个主导的信息性内容类型。

[0722] EE.11.根据EE 1至EE 10中任一项所述的音量校平器控制方法，还包括，针对每个内容类型，基于所述音频信号的过去的置信度值来对所述音频信号的当次置信度值进行平滑。

[0723] EE.12.根据EE 11所述的音量校平器控制方法，其中，类型平滑操作被配置成通过计算当前的实际置信度值与上一次的经平滑的置信度值的加权和来确定所述音频信号当

次的经平滑的置信度值。

[0724] EE.13.根据EE 1至EE 10中任一项所述的音量校平器控制方法，还包括识别所述音频信号的上下文类型，其中，所述调整的操作被配置成基于所述上下文类型的置信度值来调整所述动态增益的范围。

[0725] EE.14.根据EE 1至EE 10中任一项所述的音量校平器控制方法，还包括识别所述音频信号的上下文类型，其中，所述调整的操作被配置成基于所述音频信号的所述上下文类型来将所述音频信号的所述内容类型视为信息性的或者是干扰性的。

[0726] EE.15.根据EE 14所述的音量校平器控制方法，其中，所述音频信号的所述上下文类型包括VoIP、电影类媒体、长期音乐和游戏中的一个。

[0727] EE.16.根据EE 14所述的音量校平器控制方法，其中，在VoIP上下文类型的音频信号中，所述背景声音被视为干扰性内容类型；而在非VoIP上下文类型的音频信号中，所述背景声音和/或语音和/或音乐被视为信息性内容类型。

[0728] EE.17.根据EE 14所述的音量校平器控制方法，其中，所述音频信号的所述上下文类型包括高质量音频或低质量音频。

[0729] EE.18.根据EE 14所述的音量校平器控制方法，其中，取决于音频信号的上下文类型，给不同上下文类型的音频信号中的所述内容类型分配不同的权重。

[0730] EE.19.根据EE 14所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个上下文类型中，并且所述调整的操作被配置成通过基于所述多个上下文类型的重要性对所述多个上下文类型的置信度值进行加权来考虑所述多个上下文类型中的至少一些上下文类型。

[0731] EE.20.根据EE 14所述的音量校平器控制方法，其中，将所述音频信号分类到具有相应置信度值的多个上下文类型中，并且所述调整的操作被配置成通过基于所述置信度值对所述多个上下文类型的影响进行加权来考虑所述多个上下文类型中的至少一些上下文类型。

[0732] EE.21.根据EE 14所述的音量校平器控制方法，其中，

[0733] 识别所述内容类型的操作被配置成按所述音频信号的短期片段来识别所述内容类型；并且

[0734] 识别所述上下文类型的操作被配置成至少部分地基于所识别的所述内容类型来按所述音频信号的短期片段识别所述上下文类型。

[0735] EE.22.根据EE 21所述的音量校平器控制方法，其中，识别所述内容类型的操作包括将短期片段分类到VoIP语音内容类型或者非VoIP语音内容类型中；并且

[0736] 所述识别所述上下文类型的操作被配置成基于VoIP语音和非VoIP 语音的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

[0737] EE.23.根据EE 22所述的音量校平器控制方法，其中，识别所述内容类型的操作还包括：

[0738] 将短期片段分类到VoIP噪声内容类型和非VoIP噪声内容类型中；并且

[0739] 识别所述上下文类型的操作被配置成基于VoIP语音、非VoIP语音、VoIP噪声和非VoIP噪声的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

[0740] EE.24.根据EE 22所述的音量校平器控制方法，其中，识别所述上下文类型的操作

被配置成：

[0741] 如果VoIP语音的置信度值大于第一阈值，则将所述短期片段分类成所述VoIP上下文类型；

[0742] 如果VoIP语音的置信度值不大于第二阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值；否则

[0743] 将所述短期片段分类成上一个短期片段的上下文类型。

[0744] EE.25.根据EE 23所述的音量校平器控制方法，其中，识别所述上下文类型的操作被配置成：

[0745] 如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值，则将所述短期片段分类成所述VoIP上下文类型；

[0746] 如果VoIP语音的置信度值不大于第二阈值，或者如果VoIP噪声的置信度值不大于第四阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值，所述第四阈值不大于所述第三阈值；否则

[0747] 将所述短期片段分类成上一个短期片段的上下文类型。

[0748] EE.26.根据EE 21至EE 25中任一项所述的音量校平器控制方法，还包括基于所述内容类型的过去的置信度值来对所述内容类型的当次置信度值进行平滑。

[0749] EE.27.根据EE 26所述的音量校平器控制方法，其中，所述平滑的操作被配置成通过计算当前短期片段的置信度值与上一个短期片段的经平滑的置信度值的加权和来对当前短期片段的置信度值进行平滑。

[0750] EE.28.根据EE 27所述的音量校平器控制方法，还包括识别所述短期片段的语音内容类型，其中，在所述语音内容类型的置信度值低于第五阈值的情况下，将平滑之前的当前短期片段的VoIP语音的置信度值设定为预定的置信度值或者上一个短期片段的经平滑的置信度值。

[0751] EE.29.根据EE 22或EE 23所述的音量校平器控制方法，其中，通过使用所述短期片段的所述内容类型的置信度值以及从所述短期片段提取的其他特征作为特征，基于机器学习模型对所述短期片段进行分类。

[0752] EE.30.根据EE 14至EE 29中任一项所述的音量校平器控制方法，还包括测量所述识别所述上下文类型的操作连续地输出同一上下文类型的持续时间，其中，所述调整的操作被配置成继续使用当前的上下文类型直到新的上下文类型的持续时间的长度达到第六阈值为止。

[0753] EE.31.根据EE 30所述的音量校平器控制方法，其中，针对不同的从一个上下文类型到另一个上下文类型的转换对来设定不同的所述第六阈值。

[0754] EE.32.根据EE 30所述的音量校平器控制方法，其中，所述第六阈值与所述新的上下文类型的置信度值负相关。

[0755] EE.33.根据EE24或25所述的音量校平器控制方法，其中，所述第一阈值和/或所述第二阈值取决于上一个短期片段的上下文类型而不同。

[0756] EE.34.一种音频分类方法，包括：

[0757] 识别音频信号的短期片段的内容类型；以及

[0758] 至少部分地基于所识别的所述内容类型来识别所述短期片段的上下文类型。

[0759] EE.35.根据EE 34所述的音频分类方法，其中，对所述内容类型进行分类的操作包括将所述短期片段分类到VoIP语音内容类型或者非VoIP语音内容类型中；并且

[0760] 对所述上下文类型进行分类的操作被配置成基于VoIP语音和非VoIP 语音的置信度值将所述短期片段分类到VoIP上下文类型或者非VoIP上下文类型中。

[0761] EE.36.根据EE 35所述的音频分类方法，其中，对所述内容类型进行分类的操作还包括：

[0762] 将所述短期片段分类到VoIP噪声内容类型或者非VoIP噪声内容类型中；并且

[0763] 对所述上下文类型进行分类的操作被配置成基于VoIP语音、非VoIP 语音、VoIP噪声和非VoIP噪声的置信度值将所述短期片段分类到VoIP 上下文类型或者非VoIP上下文类型中。

[0764] EE.37.根据EE 35所述的音频分类方法，其中，对所述上下文类型进行分类的操作被配置成：

[0765] 如果VoIP语音的置信度值大于第一阈值，则将所述短期片段分类成所述VoIP上下文类型；

[0766] 如果VoIP语音的置信度值不大于第二阈值，则将所述短期片段分类成所述非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值；否则

[0767] 将所述短期片段分类成上一个短期片段的上下文类型。

[0768] EE.38.根据EE 36所述的音频分类方法，其中，对所述上下文类型进行分类的操作被配置成：

[0769] 如果VoIP语音的置信度值大于第一阈值或者如果VoIP噪声的置信度值大于第三阈值，则将所述短期片段分类为VoIP上下文类型；

[0770] 如果VoIP语音的置信度值不大于第二阈值，或者如果VoIP噪声的置信度值不大于第四阈值，则将所述短期片段分类为非VoIP上下文类型，其中所述第二阈值不大于所述第一阈值，所述第四阈值不大于所述第三阈值；否则

[0771] 将所述短期片段分类为上一个短期片段的上下文类型。

[0772] EE.39.根据EE 34至EE 38中任一项所述的音频分类方法，还包括基于所述内容类型的过去的置信度值来对所述内容类型的当次置信度值进行平滑。

[0773] EE.40.根据EE 39所述的音频分类方法，其中，平滑操作被配置成通过计算当前短期片段的置信度值与上一个短期片段的经平滑的置信度值的加权和来确定所述当前短期片段的经平滑的置信度值。

[0774] EE.41.根据EE 40所述的音频分类方法，还包括从所述短期片段识别语音内容类型，其中，在所述语音内容类型的置信度值低于第五阈值的情况下，将平滑之前的当前短期片段的VoIP语音的置信度值设定为预定的置信度值或者上一个短期片段的经平滑的置信度值。

[0775] EE.42.根据EE 35或者EE 36所述的音频分类方法，其中，对所述上下文类型进行分类的操作被配置成：通过使用所述短期片段的所述内容类型的置信度值以及从所述短期片段提取的其他特征作为特征，基于机器学习模型对所述短期片段进行分类。

[0776] EE.43.根据EE 37或者EE 38所述的音频分类方法，还包括测量所述识别上下文类型的操作连续地输出同一上下文类型的持续时间，其中，所述音频分类方法被配置成继续

使用当前的上下文类型直到新的上下文类型的持续时间的长度达到第六阈值为止。

[0777] EE.44.根据EE 43所述的音频分类方法，其中，针对不同的从一个上下文类型到另一个上下文类型的转换对来设定不同的所述第六阈值。

[0778] EE.45.根据EE 43所述的音频分类方法，其中，所述第六阈值与所述新的上下文类型的置信度值负相关。

[0779] EE.46.根据EE 37或38所述的音频分类方法，其中，所述第一阈值和/ 或所述第二阈值取决于上一个短期片段的上下文类型而不同。

[0780] EE.47.一种其上记录有计算机程序指令的计算机可读介质，当被处理器执行时，所述指令使所述处理器能够执行音量校平器控制方法，所述音量校平器控制方法包括：

[0781] 实时地识别音频信号的内容类型；以及

[0782] 通过使所述音量校平器的动态增益与所述音频信号的信息性内容类型正相关，并且使所述音量校平器的动态增益与所述音频信号的干扰性内容类型负相关，来基于所识别的内容类型以连续的方式调整音量校平器。

[0783] EE.48.一种其上记录有计算机程序指令的计算机可读介质，当被处理器执行时，所述指令使所述处理器能够执行音频分类方法，所述音频分类方法包括：

[0784] 识别音频信号的短期片段的内容类型；以及

[0785] 至少部分地基于所识别的内容类型来识别所述短期片段的上下文类型。

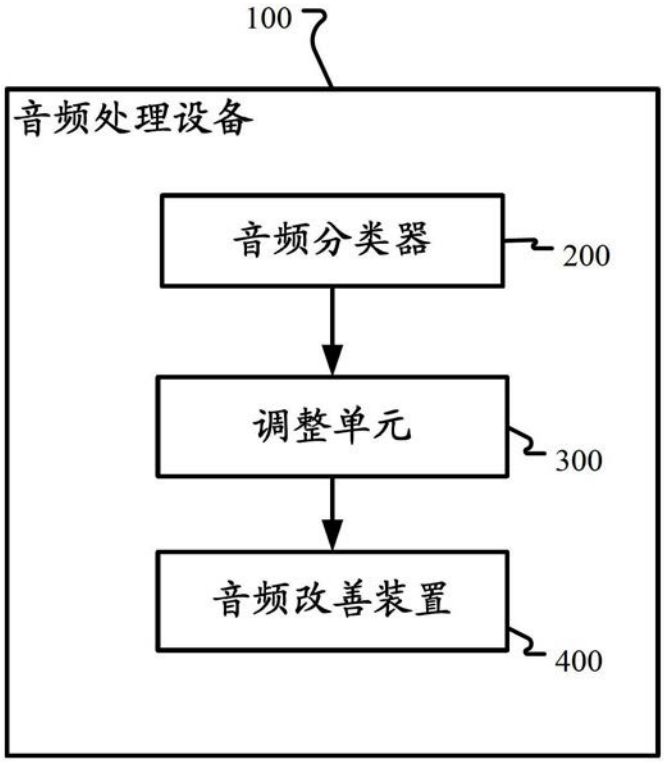

100
音频处理设备
音频分类器
200
调整单元
300
音频改善装置
400

图1

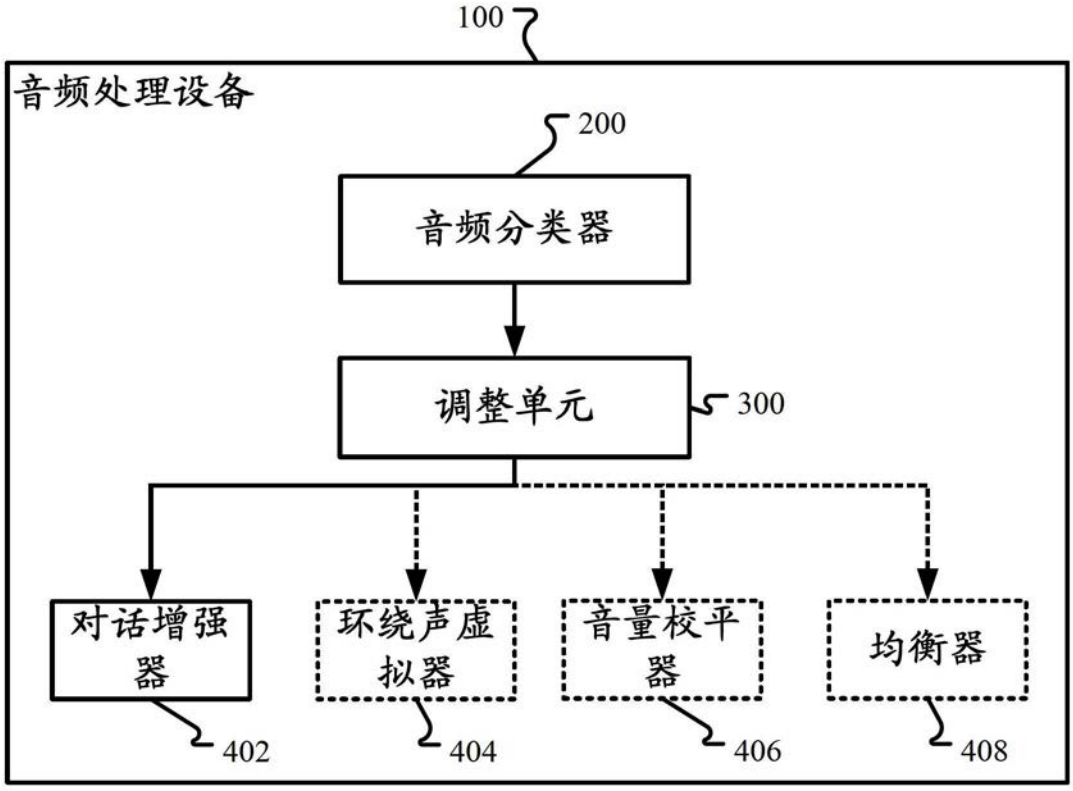

100
音频处理设备
200
音频分类器
调整单元
300
对话增强器
环绕声虚拟器
音量校平器
均衡器
402
404
406
408

图2

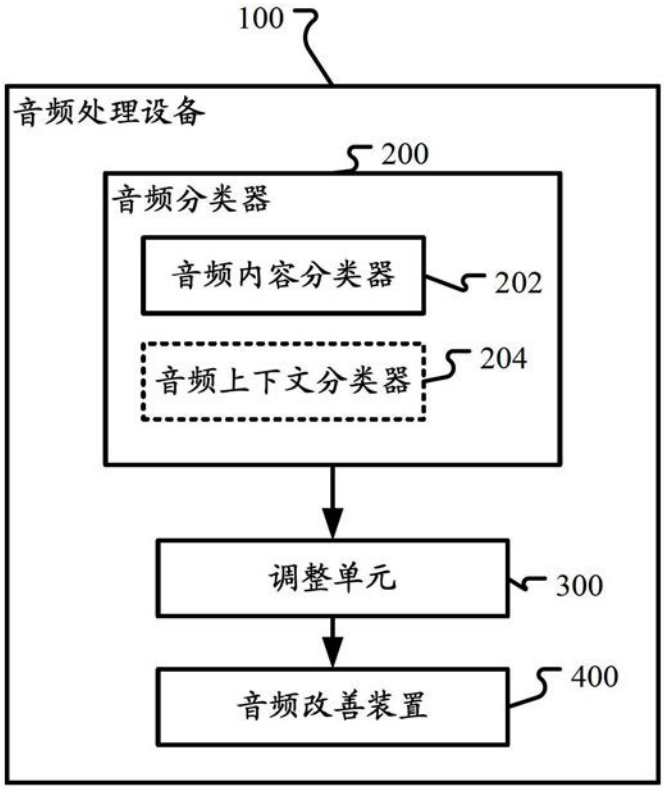

100
音频处理设备
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
调整单元
300
音频改善装置
400

图3

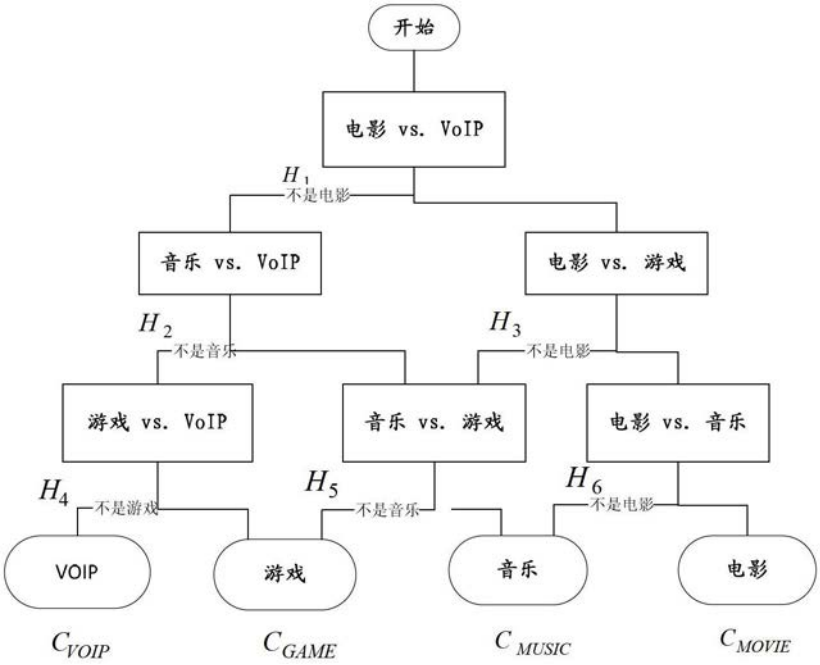

开始
电影 vs. VoIP
H1
不是电影
音乐 vs. VoIP
电影 vs. 游戏
H2
不是音乐
H3
不是电影
游戏 vs. VoIP
音乐 vs. 游戏
电影 vs. 音乐
H4
不是游戏
H5
不是音乐
H6
不是电影
VOIP
游戏
音乐
电影
CVOIP
CGAME
CMUSIC
CMOVIE

图4

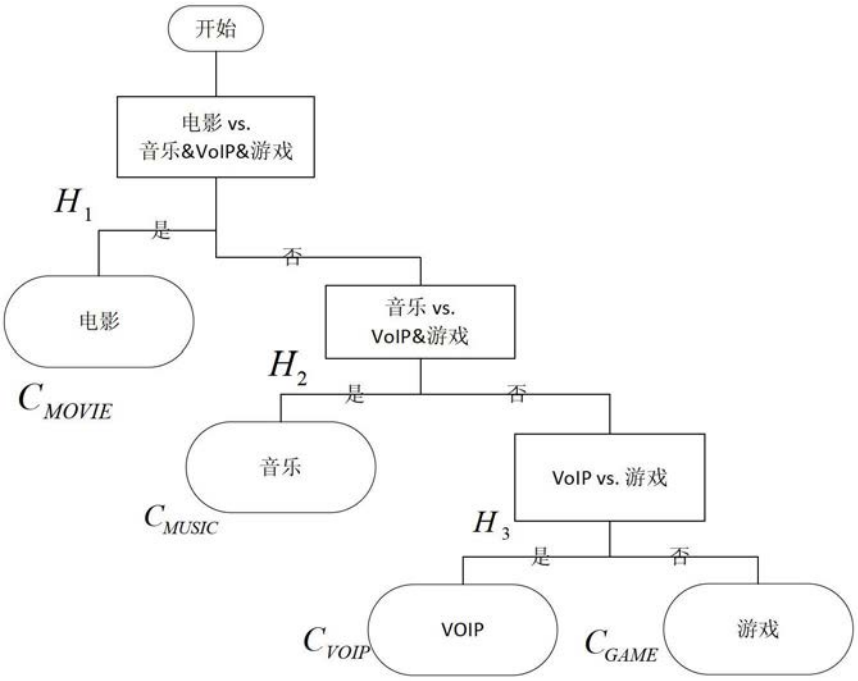

开始
电影 vs.
音乐&VoIP&游戏
H_1
是
否
电影
C_{MOVIE}
音乐 vs.
VoIP&游戏
H_2
是
否
音乐
C_{MUSIC}
VoIP vs. 游戏
H_3
是
否
C_{VOIP}
VOIP
C_{GAME}
游戏

图5

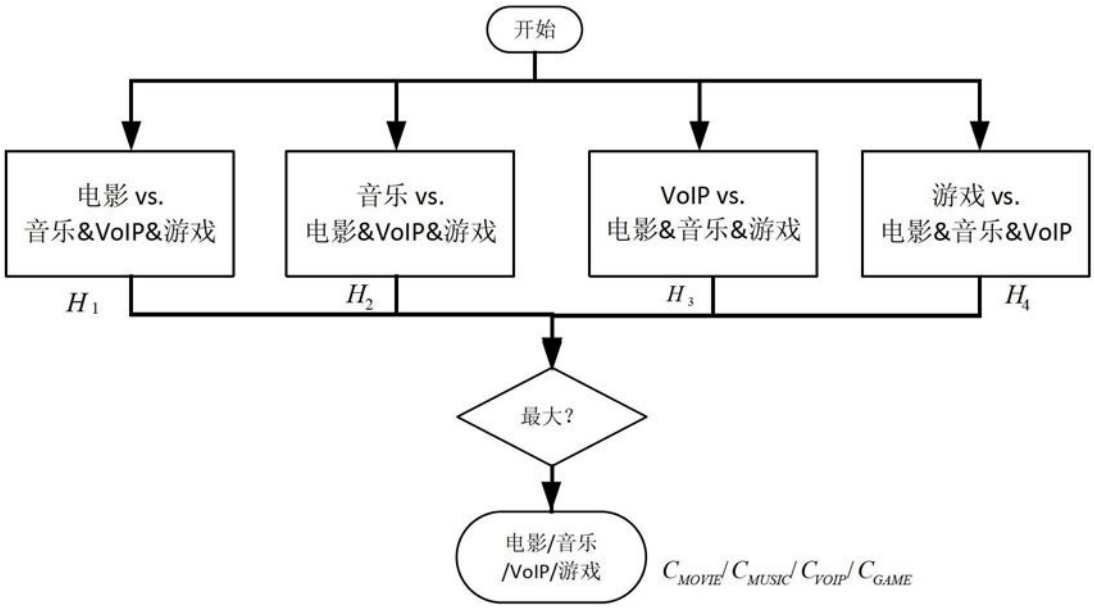

开始
电影 vs.
音乐&VoIP&游戏
音乐 vs.
电影&VoIP&游戏
VoIP vs.
电影&音乐&游戏
游戏 vs.
电影&音乐&VoIP
H_1
H_2
H_3
H_4
最大?
电影/音乐
/VoIP/游戏
$C_{MOVIE}/C_{MUSIC}/C_{VOIP}/C_{GAME}$

图6

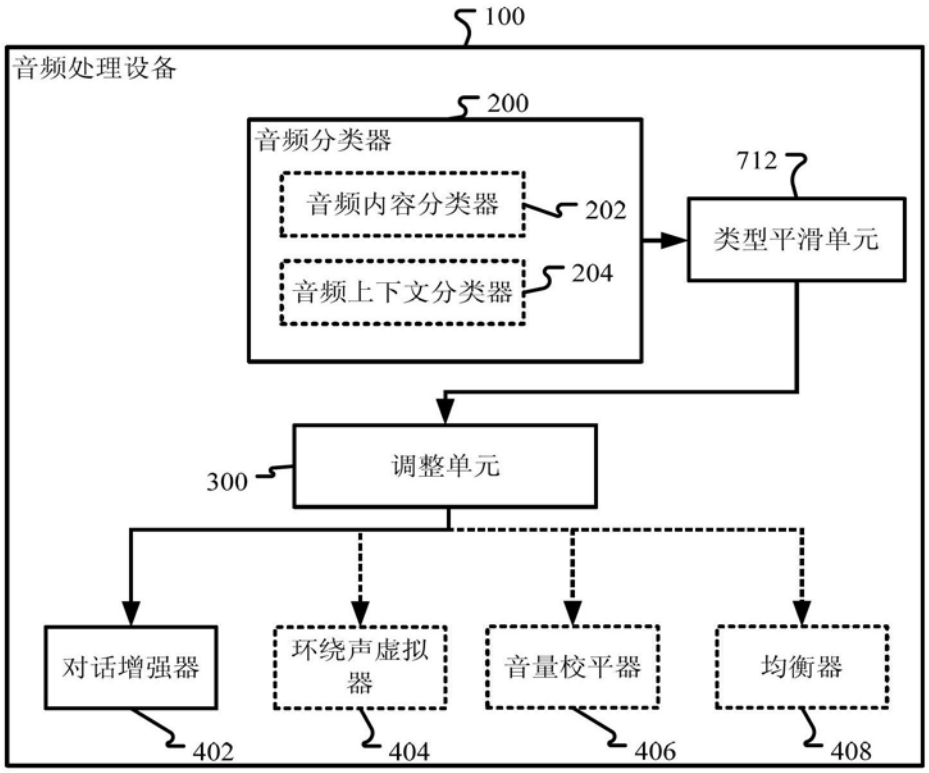

100
音频处理设备
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
712
类型平滑单元
300
调整单元
对话增强器
环绕声虚拟器
音量校平器
均衡器
402
404
406
408

图7

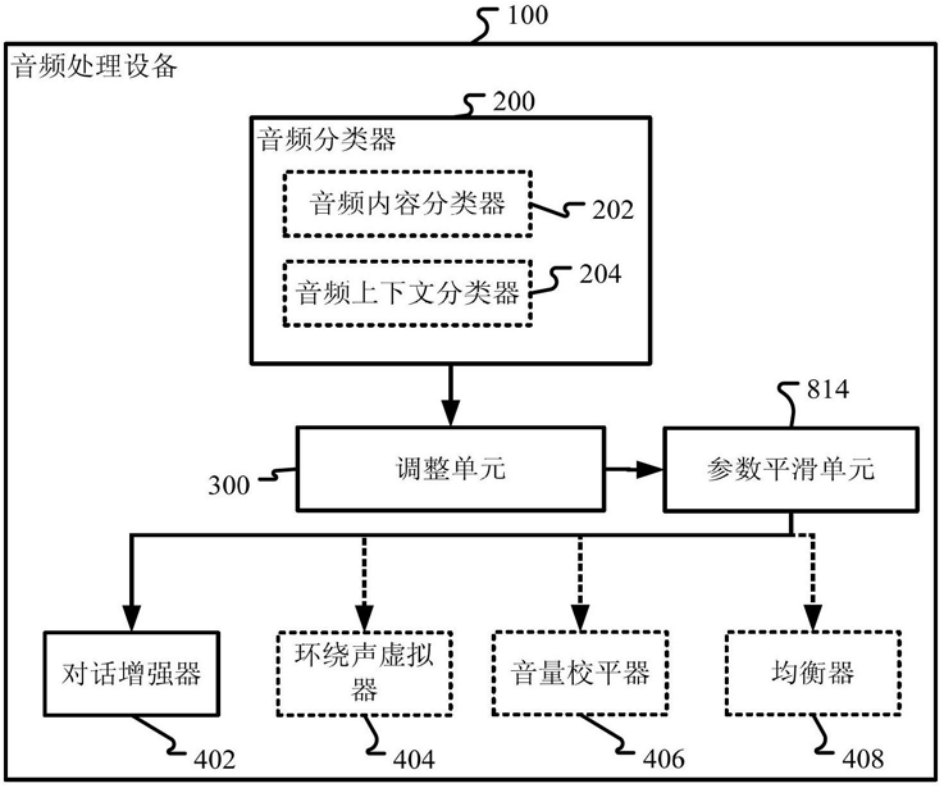

100
音频处理设备
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
814
300
调整单元
参数平滑单元
对话增强器
环绕声虚拟器
音量校平器
均衡器
402
404
406
408

图8

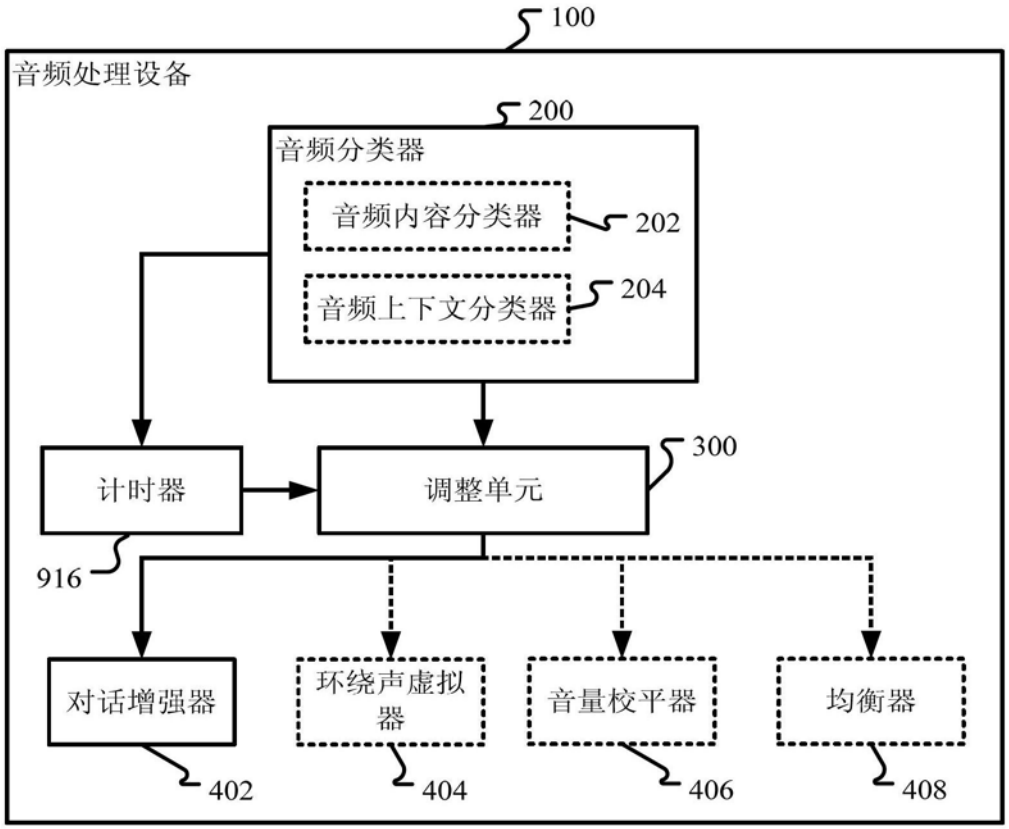

100
音频处理设备
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
计时器
调整单元
300
916
对话增强器
402
环绕声虚拟器
404
音量校平器
406
均衡器
408

图9

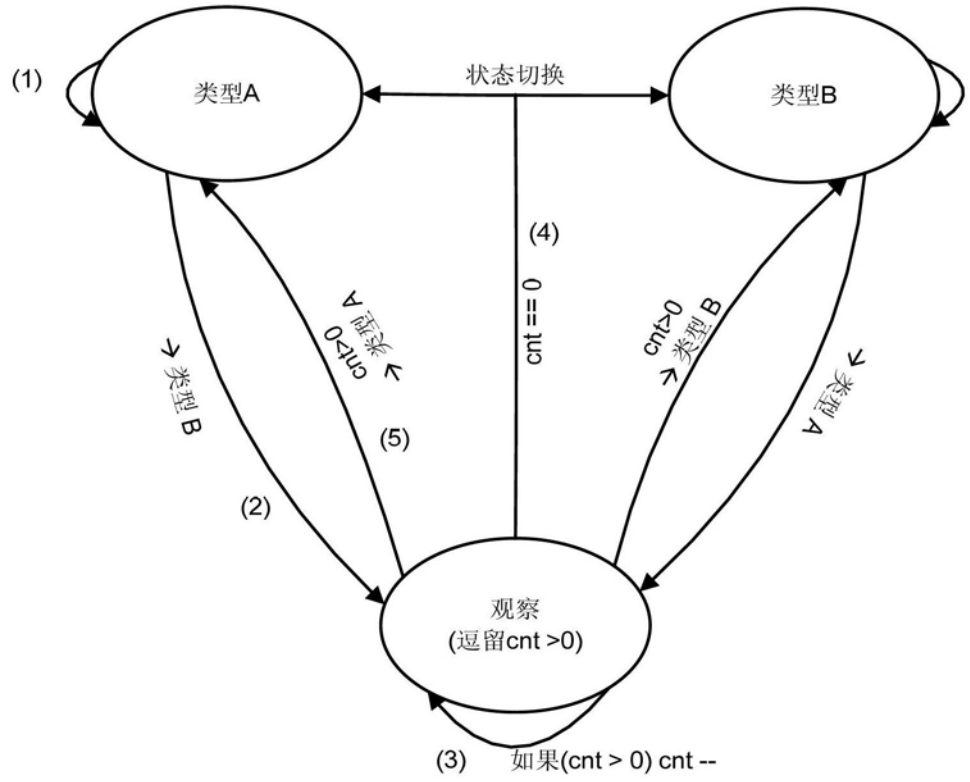

(1)
类型A
状态切换
类型B
(4)
cnt == 0
类型 B
(2)
cnt>0
类型 A
(5)
cnt>0
类型 B
类型 A
观察
(逗留cnt >0)
(3)
如果(cnt > 0) cnt --

图10

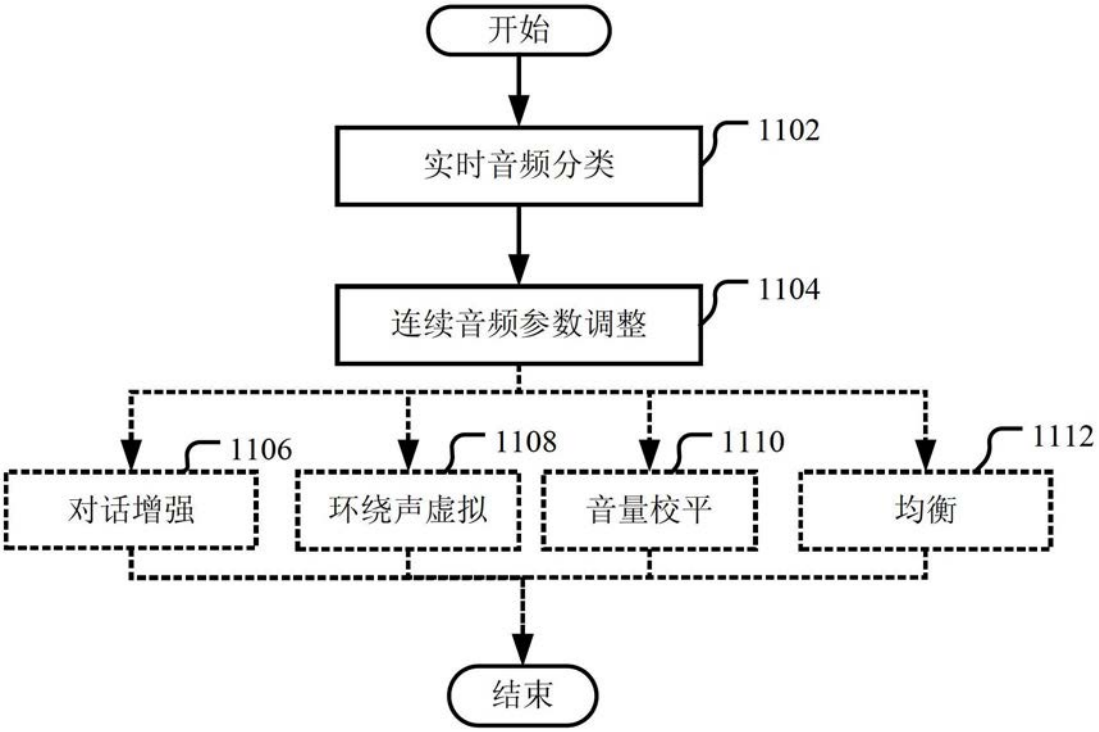

开始
实时音频分类
1102
连续音频参数调整
1104
对话增强
1106
环绕声虚拟
1108
音量校平
1110
均衡
1112
结束

图11

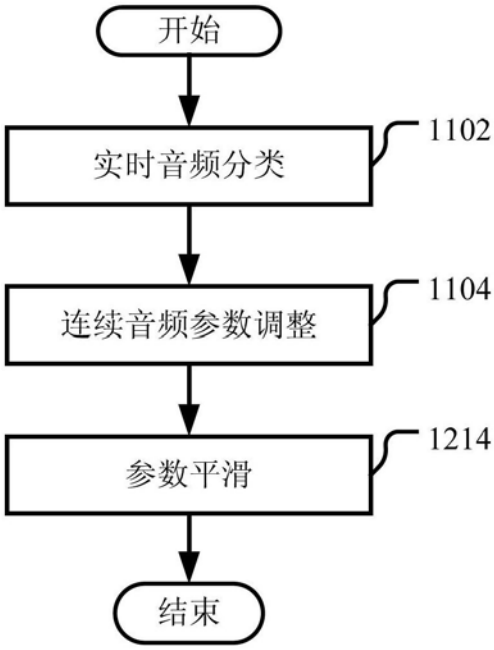

开始
实时音频分类
1102
连续音频参数调整
1104
参数平滑
1214
结束

图12

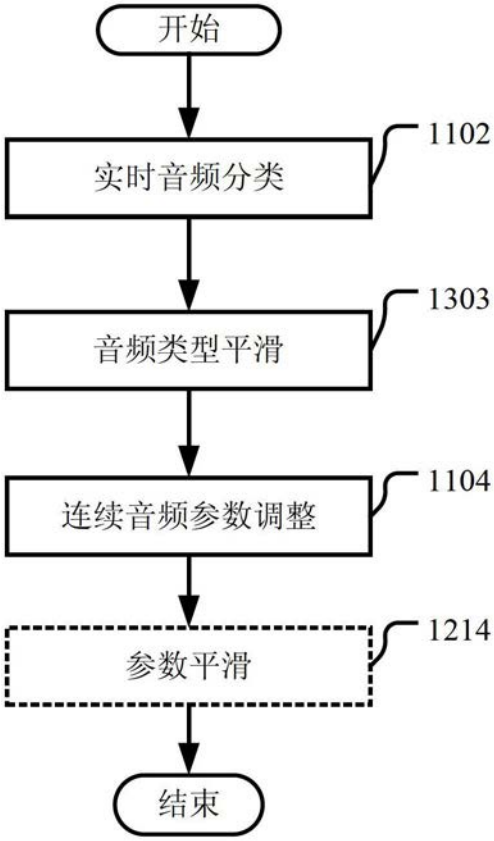

开始
实时音频分类
1102
音频类型平滑
1303
连续音频参数调整
1104
参数平滑
1214
结束

图13

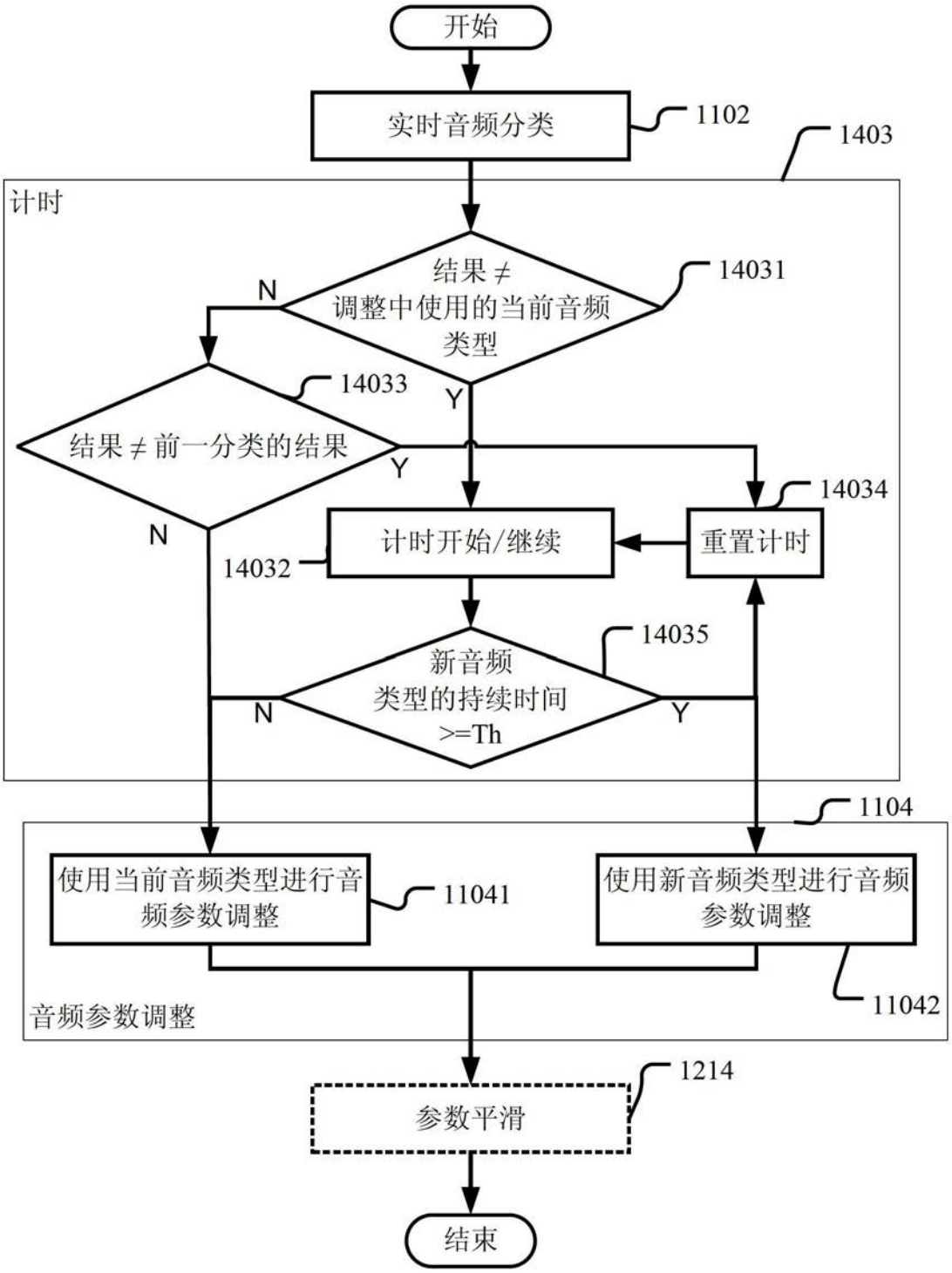

开始
实时音频分类
1102
1403
计时
结果≠
调整中使用的当前音频
类型
14031
N
Y
14033
结果≠前一分类的结果
Y
N
14034
计时开始/继续
重置计时
14032
14035
新音频
类型的持续时间
>=Th
N
Y
1104
使用当前音频类型进行音
频参数调整
11041
使用新音频类型进行音频
参数调整
11042
音频参数调整
1214
参数平滑
结束

图14

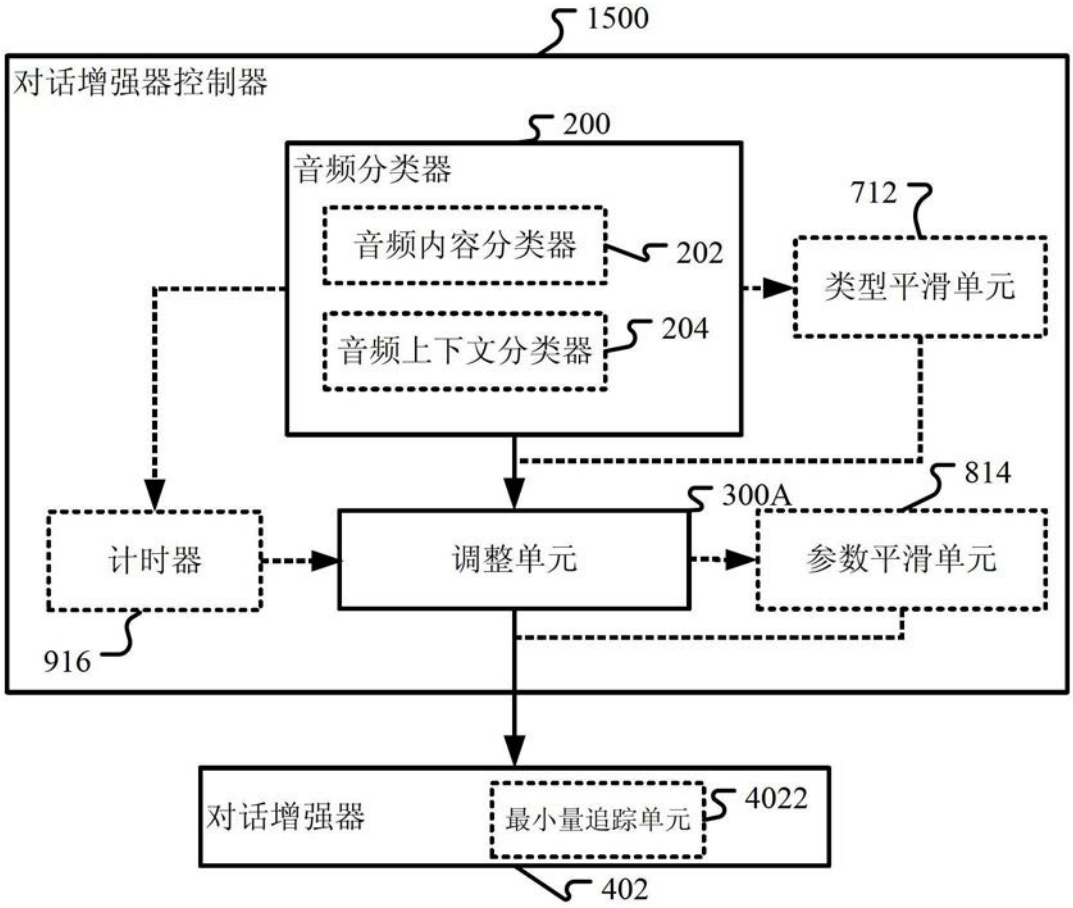

1500
对话增强器控制器
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
712
类型平滑单元
814
300A
计时器
调整单元
参数平滑单元
916
对话增强器
最小量追踪单元
4022
402

图15

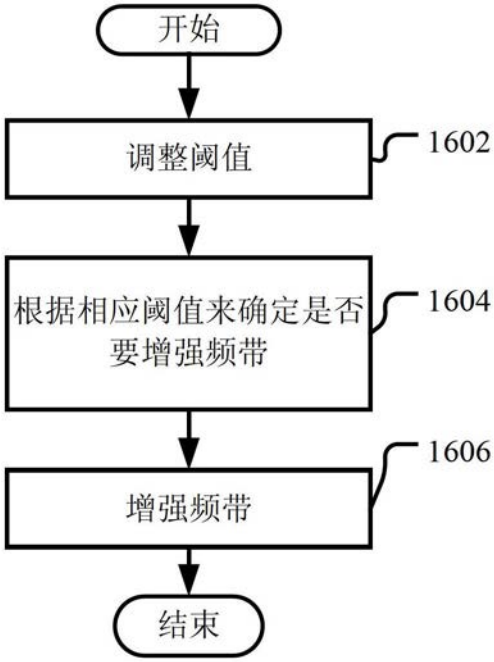

开始
调整阈值
1602
根据相应阈值来确定是否要增强频带
1604
增强频带
1606
结束

图16

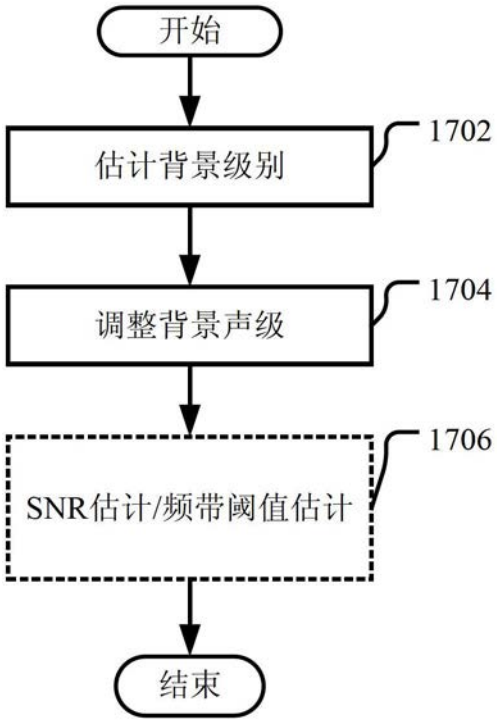

开始
估计背景级别
1702
调整背景声级
1704
SNR估计/频带阈值估计
1706
结束

图17

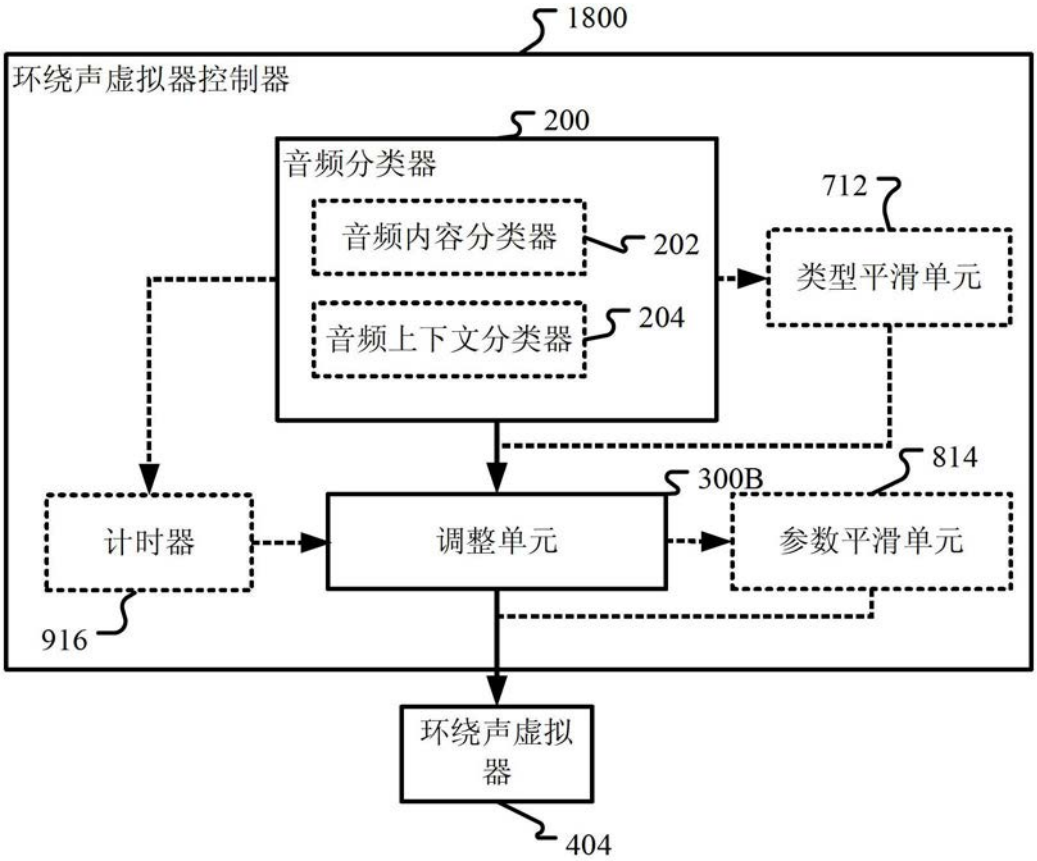

1800
环绕声虚拟器控制器
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
712
类型平滑单元
814
300B
计时器
调整单元
参数平滑单元
916
环绕声虚拟器
404

图18

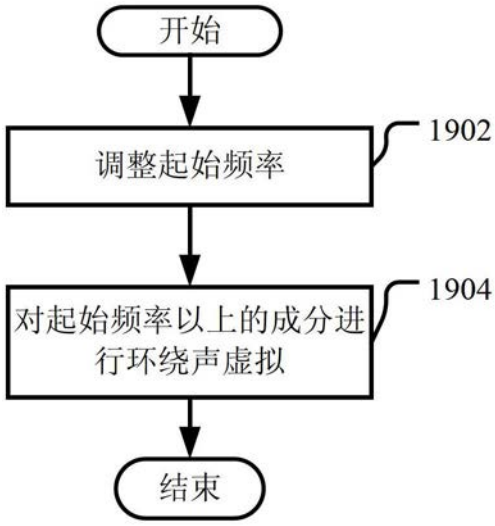

开始
1902
调整起始频率
1904
对起始频率以上的成分进行环绕声虚拟
结束

图19

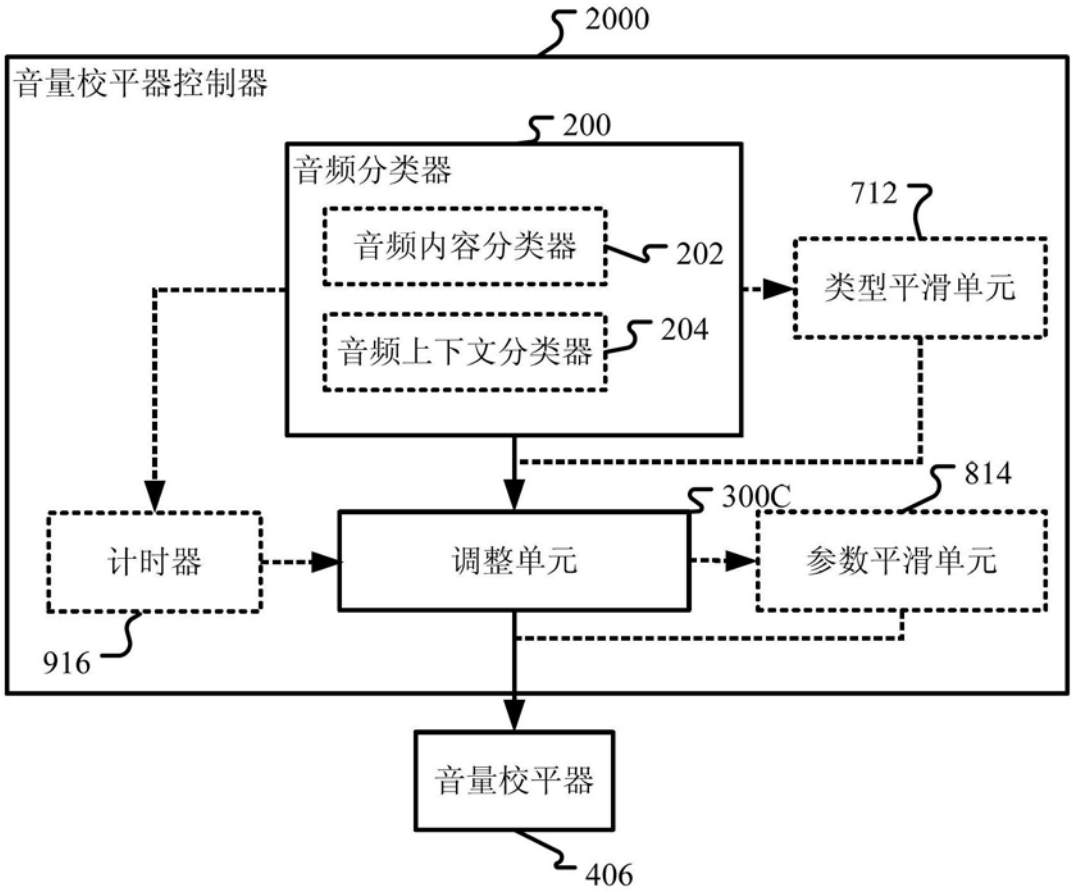

2000
音量校平器控制器
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
712
类型平滑单元
814
300C
计时器
调整单元
参数平滑单元
916
音量校平器
406

图20

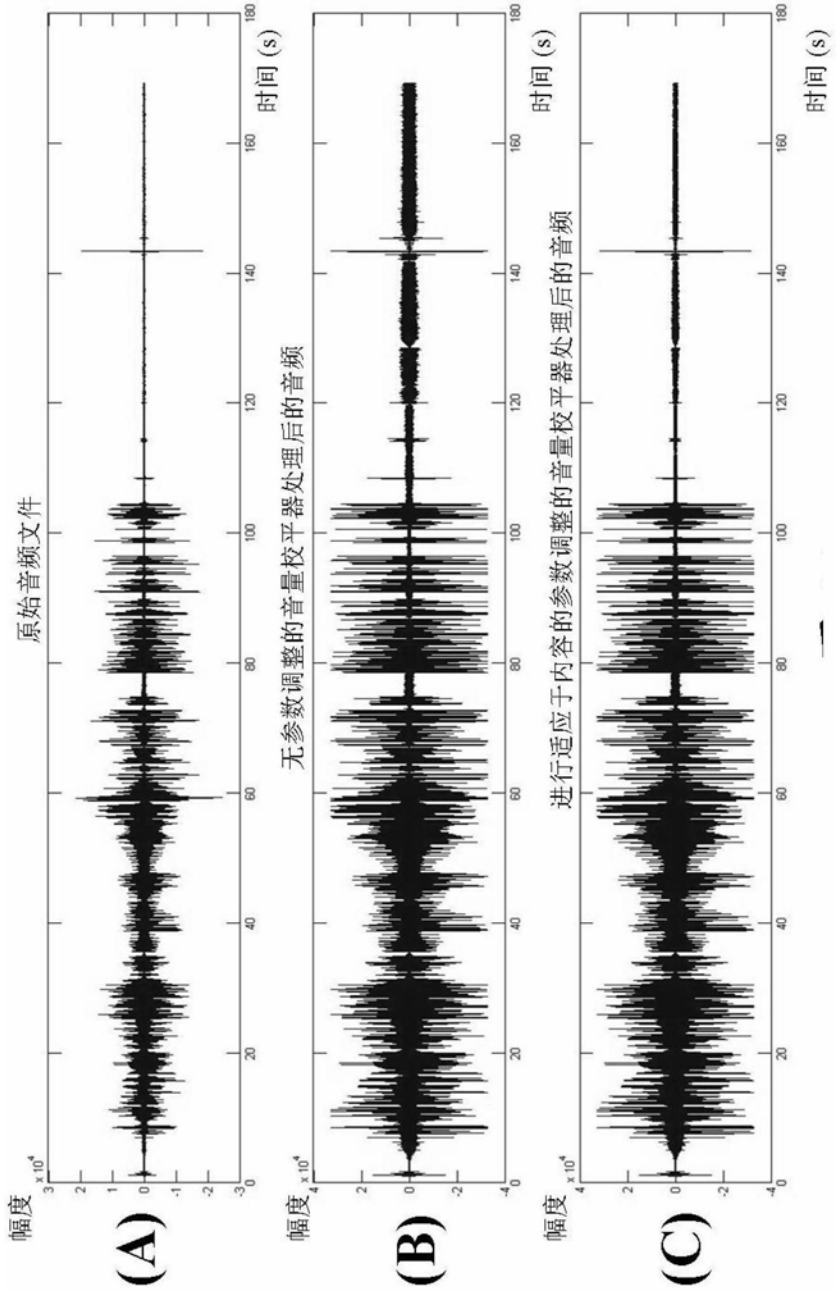

原始音频文件
幅度
时间 (s)
(A)
无参数调整的音量校平器处理后的音频
幅度
时间 (s)
(B)
进行适应于内容的参数调整的音量校平器处理后的音频
幅度
时间 (s)
(C)

图21

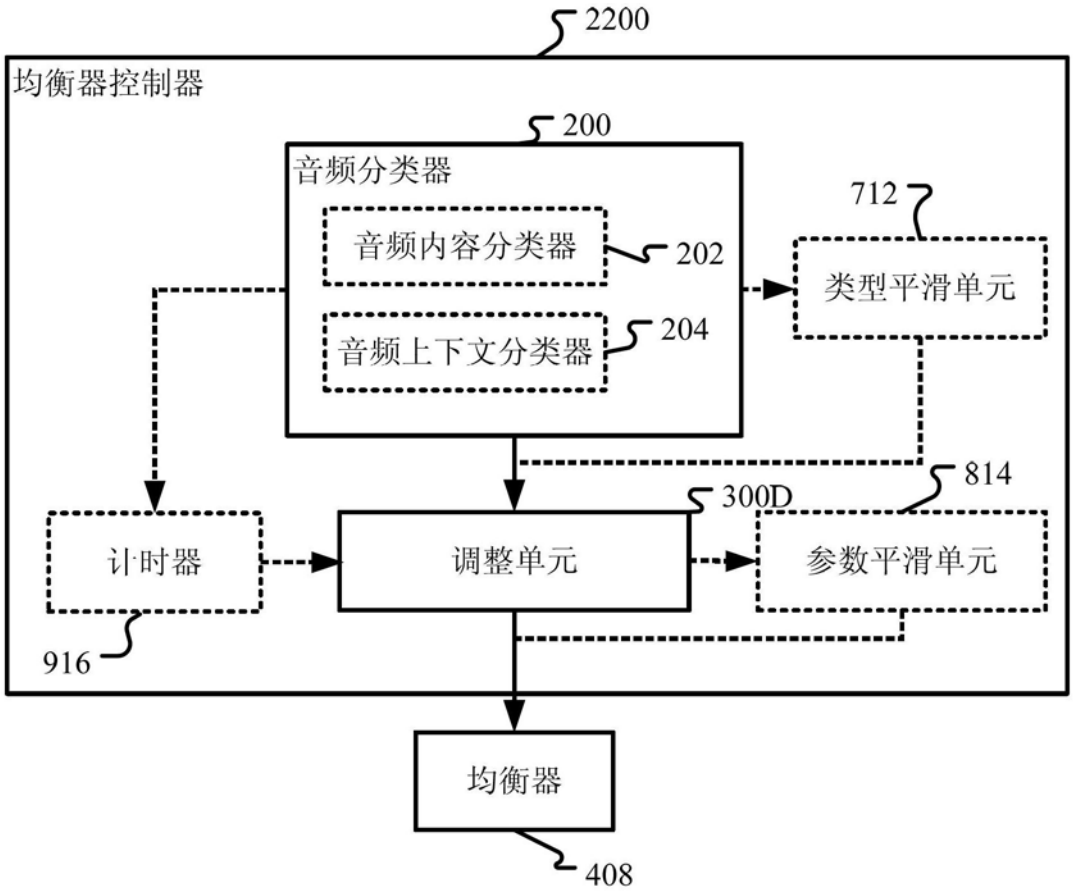

2200
均衡器控制器
200
音频分类器
音频内容分类器
202
音频上下文分类器
204
712
类型平滑单元
814
300D
计时器
调整单元
参数平滑单元
916
均衡器
408

图22

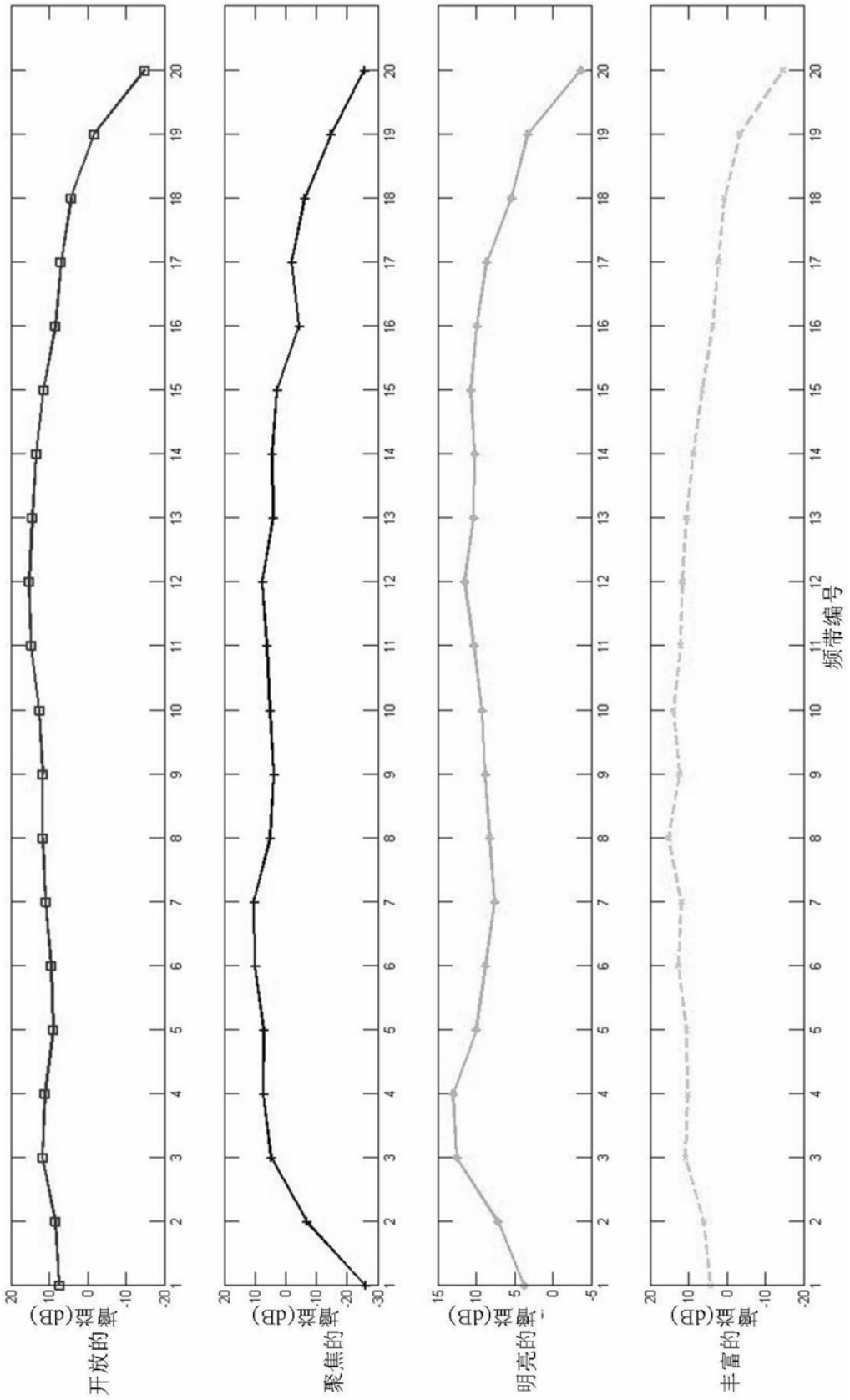


图23

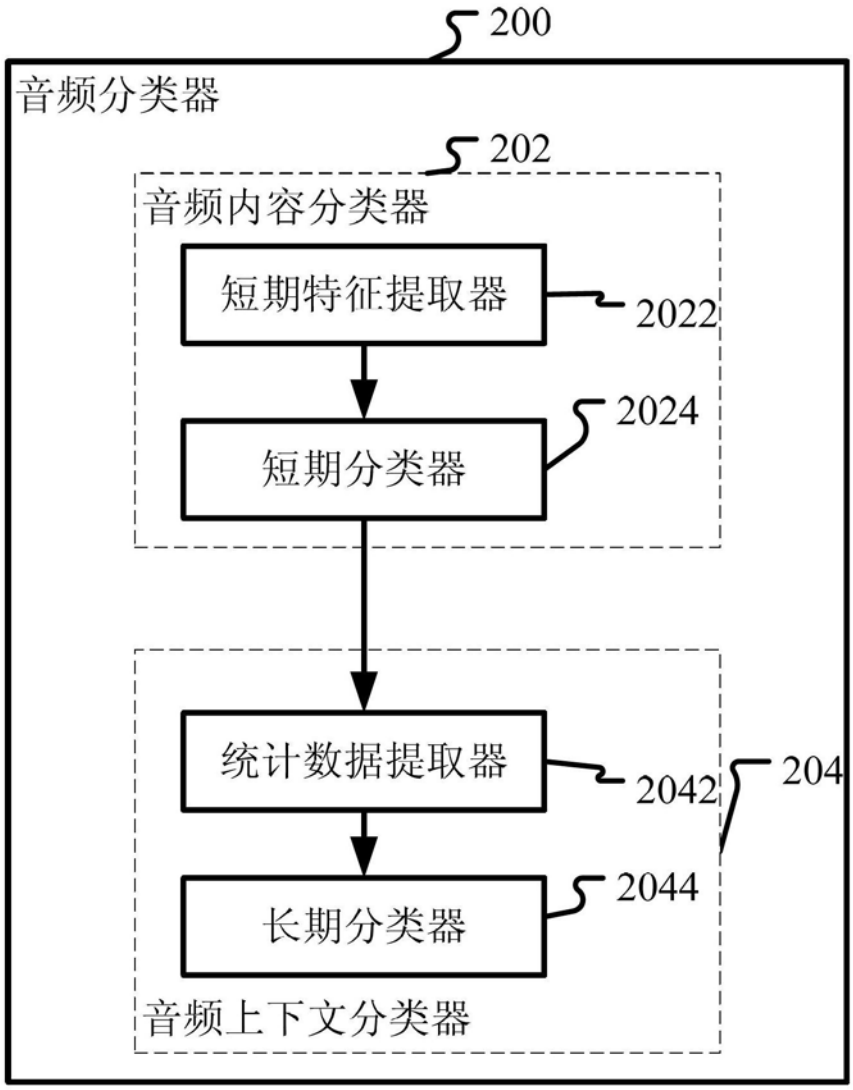

200
音频分类器
202
音频内容分类器
短期特征提取器
2022
2024
短期分类器
统计数据提取器
2042
204
2044
长期分类器
音频上下文分类器

图24

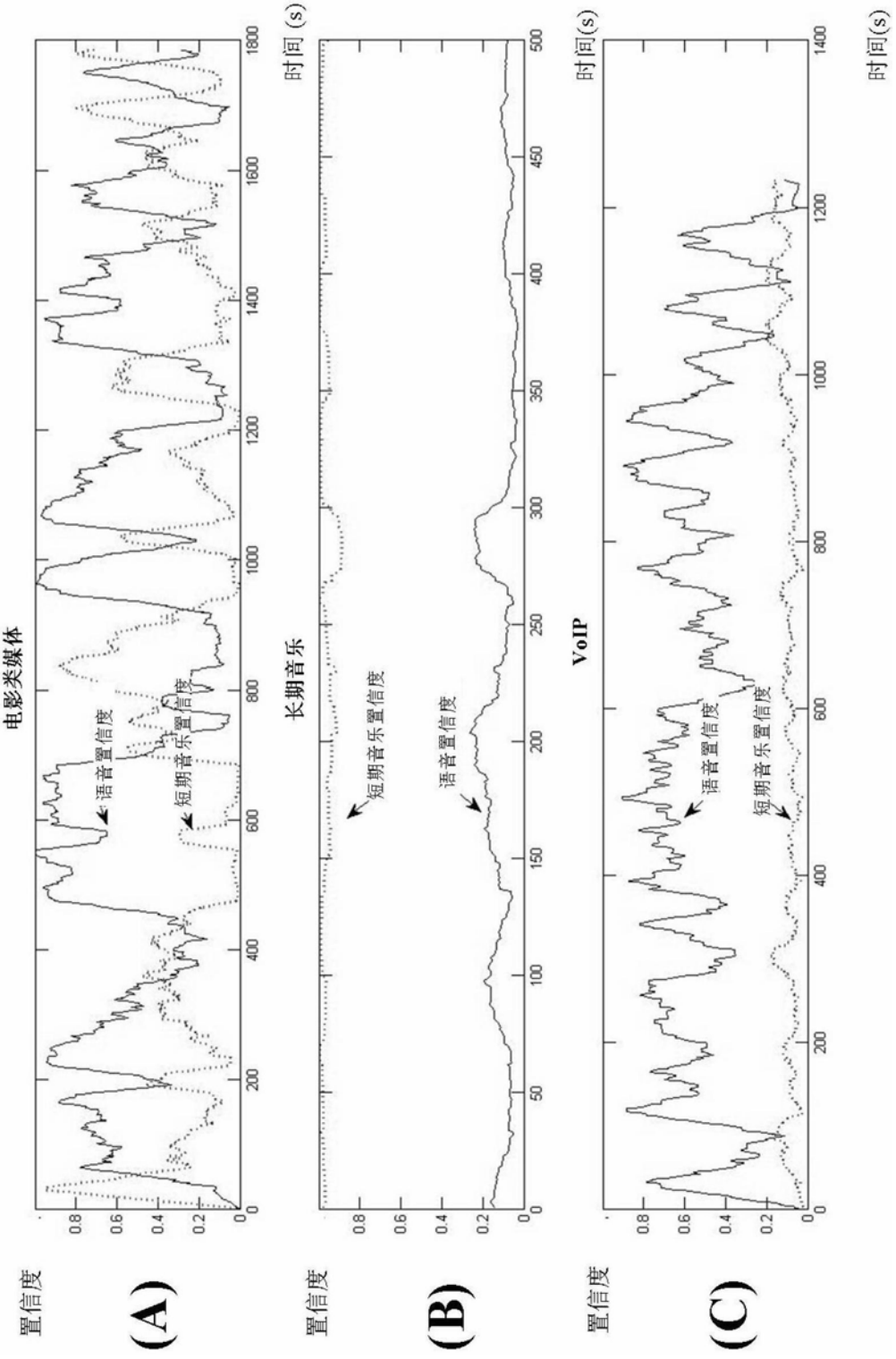

电影类媒体
语音置信度
短期音乐置信度
置信度
(A)
长期音乐
短期音乐置信度
语音置信度
时间 (s)
置信度
(B)
VoIP
时间(s)
语音置信度
短期音乐置信度
置信度
(C)
时间(s)

图25

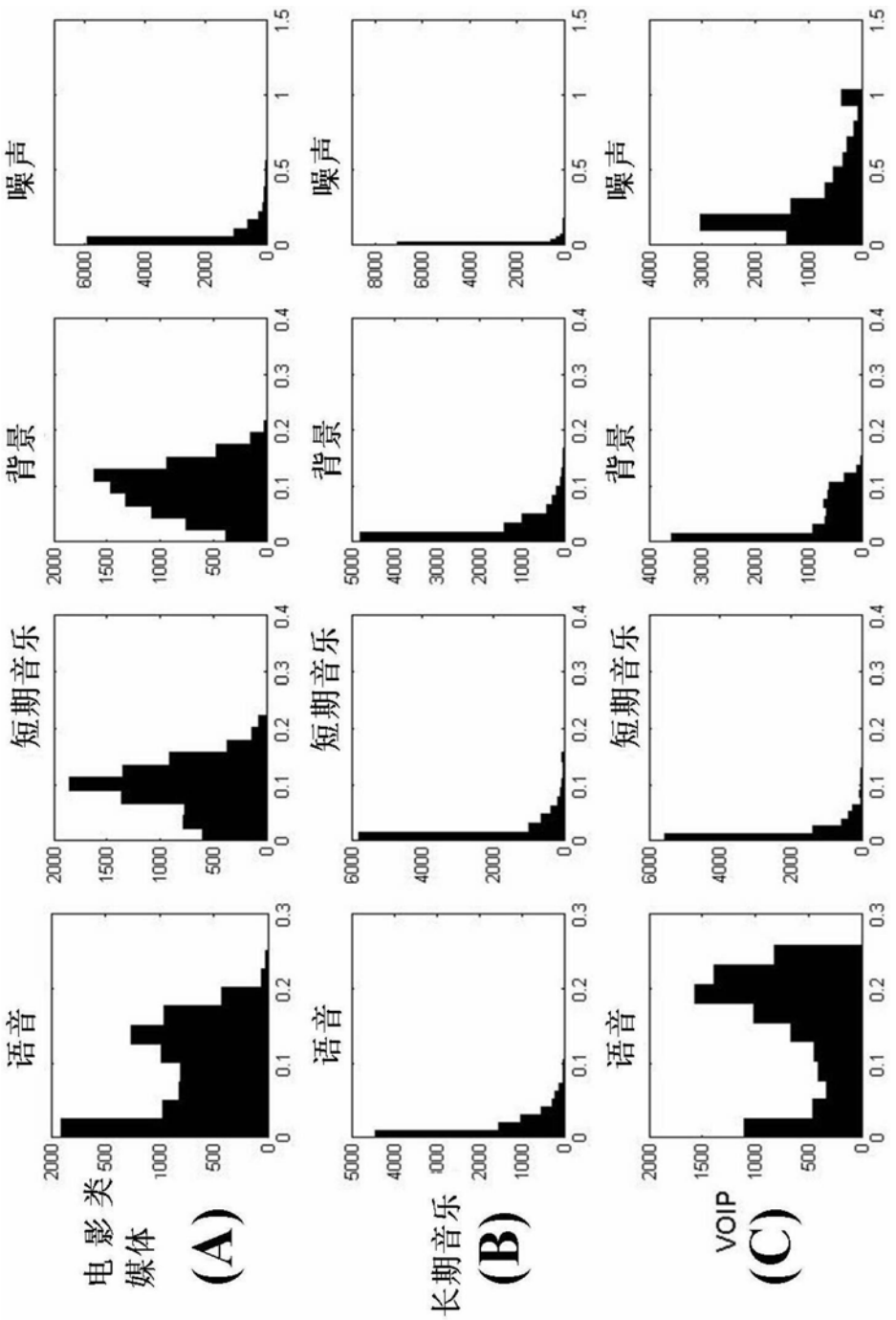


图26

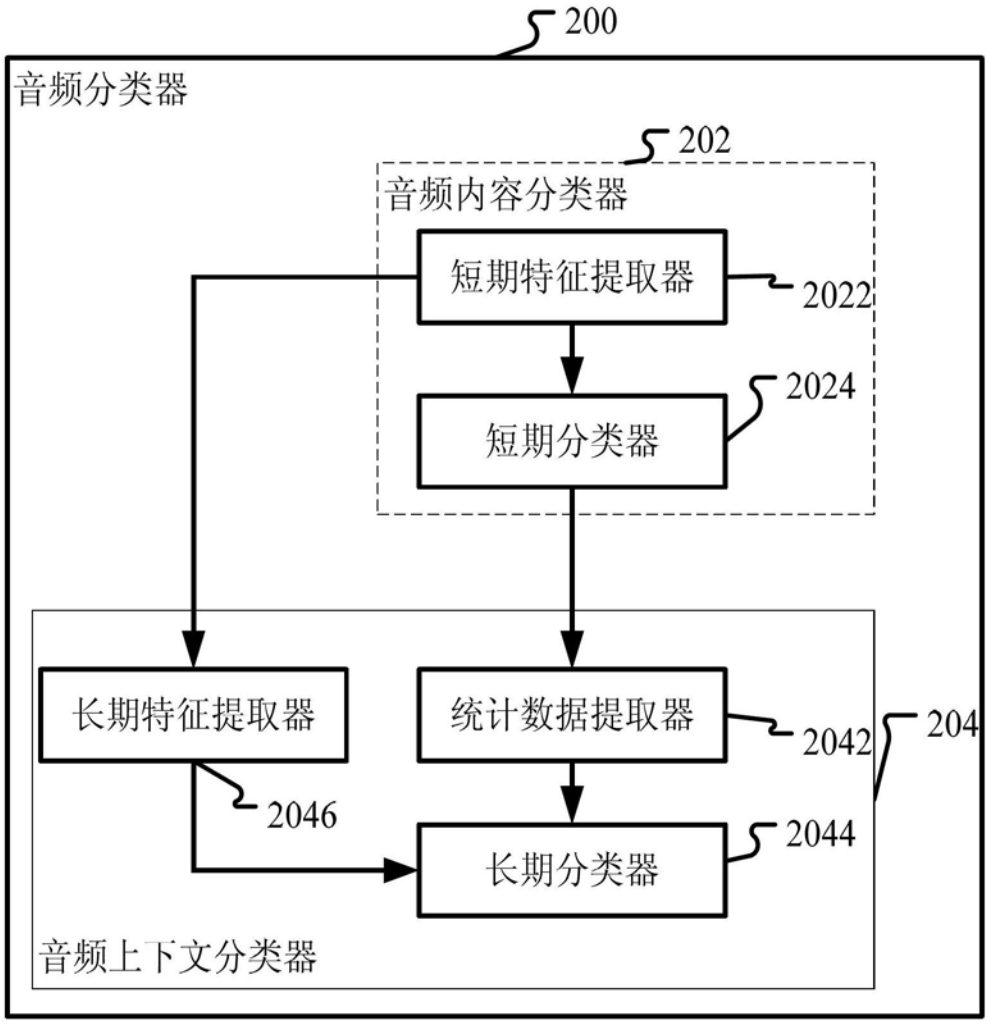

200
音频分类器
202
音频内容分类器
短期特征提取器
2022
2024
短期分类器
长期特征提取器
统计数据提取器
2042
204
2046
2044
长期分类器
音频上下文分类器

图27

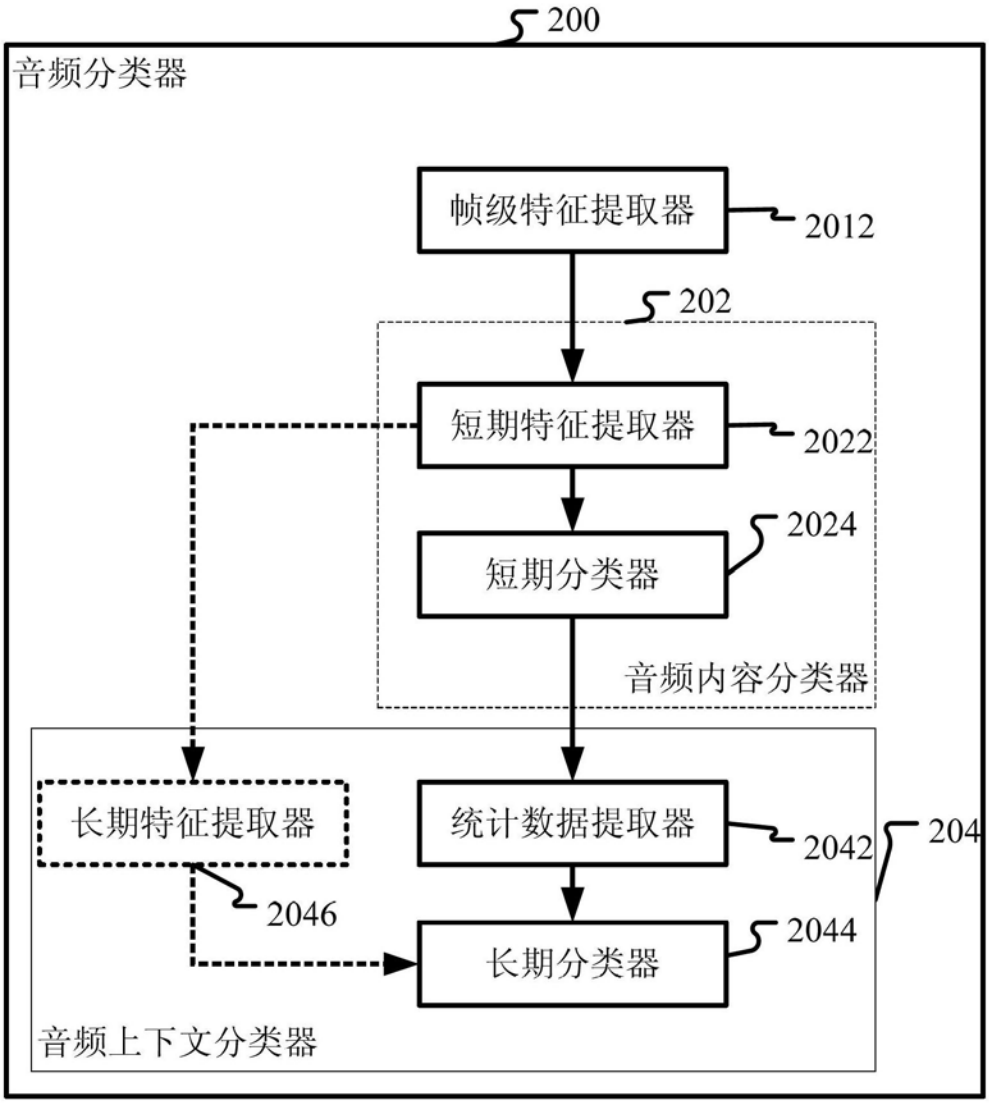

200
音频分类器
帧级特征提取器
2012
202
短期特征提取器
2022
2024
短期分类器
音频内容分类器
长期特征提取器
统计数据提取器
2042
204
2046
2044
长期分类器
音频上下文分类器

图28

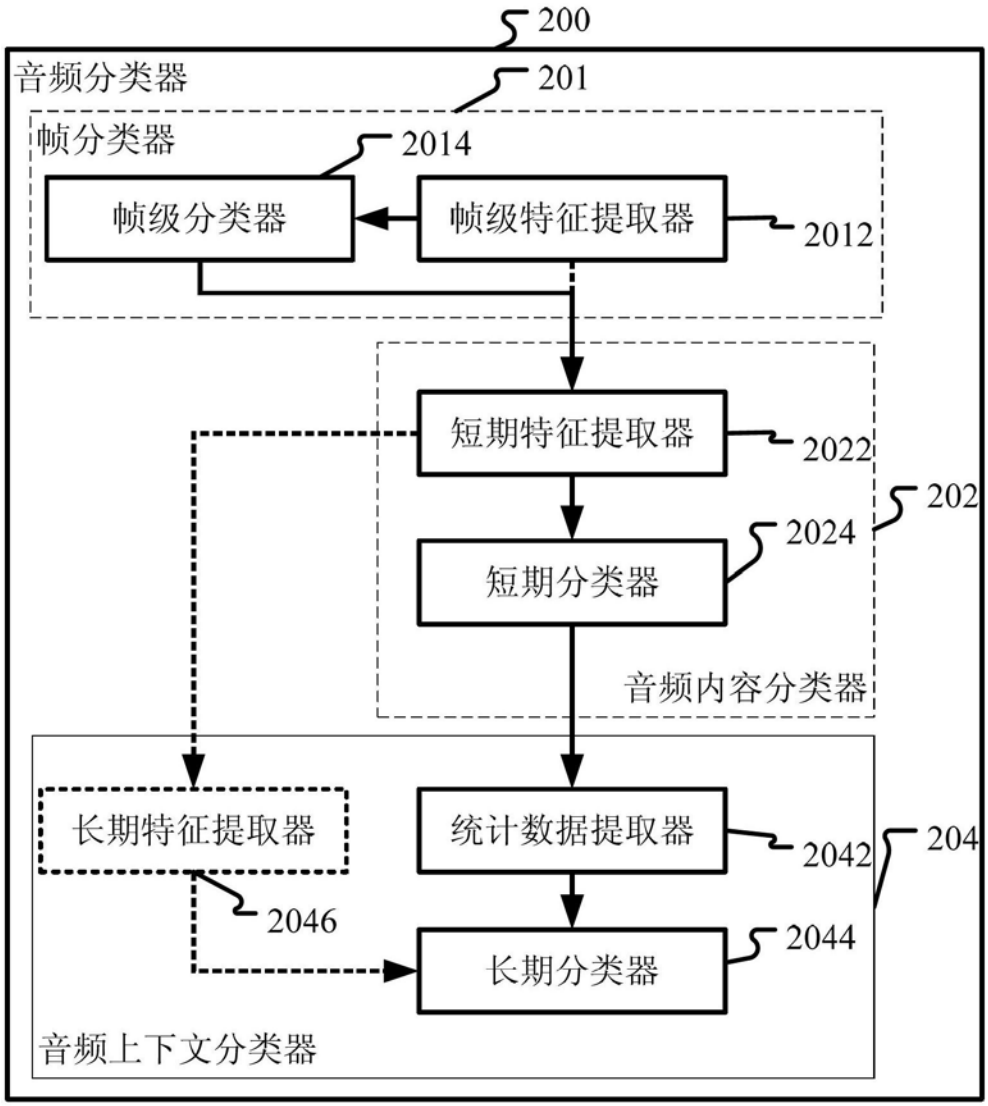

200
音频分类器
201
帧分类器
2014
帧级分类器
帧级特征提取器
2012
短期特征提取器
2022
202
2024
短期分类器
音频内容分类器
长期特征提取器
统计数据提取器
2042
204
2046
2044
长期分类器
音频上下文分类器

图29

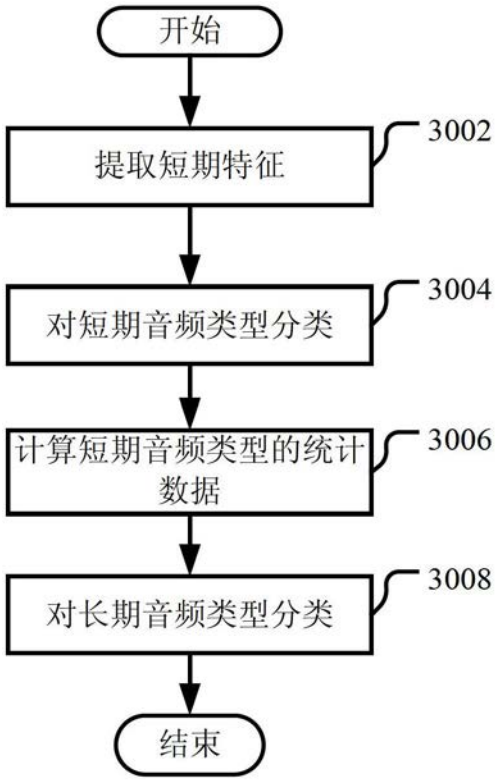

开始
提取短期特征
3002
对短期音频类型分类
3004
计算短期音频类型的统计数据
3006
对长期音频类型分类
3008
结束

图30

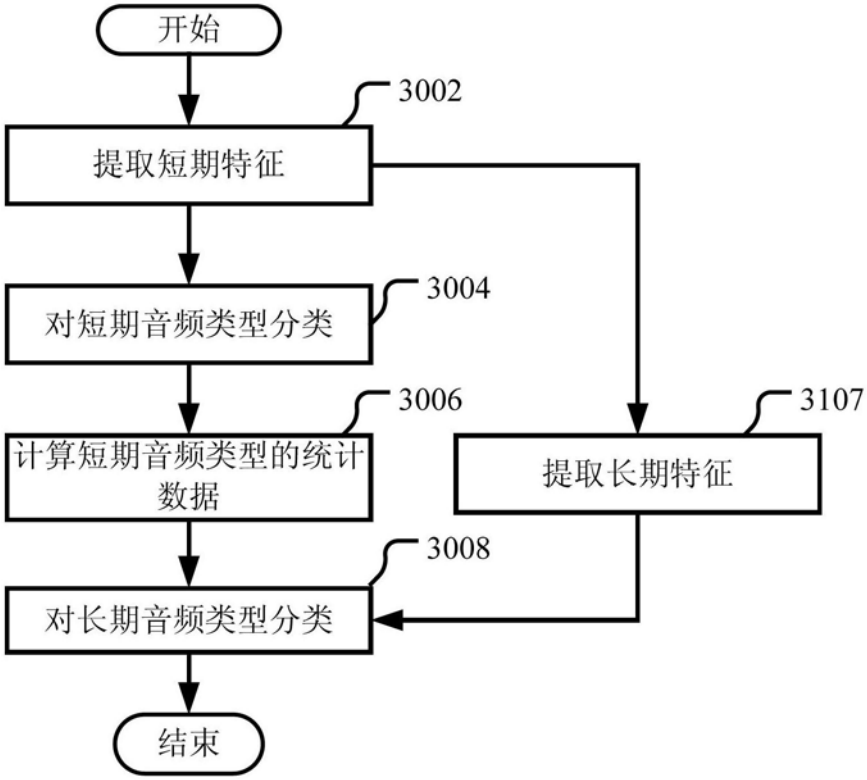

开始
提取短期特征
3002
对短期音频类型分类
3004
计算短期音频类型的统计数据
3006
提取长期特征
3107
对长期音频类型分类
3008
结束

图31

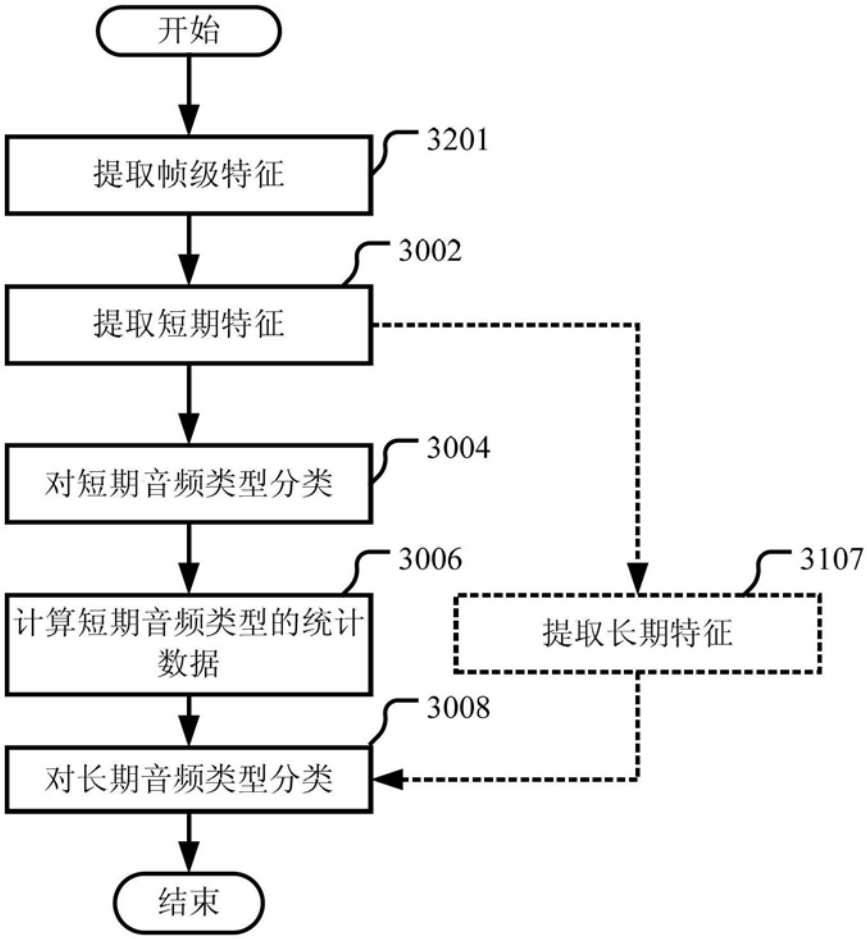

开始
提取帧级特征
3201
3002
提取短期特征
对短期音频类型分类
3004
3006
计算短期音频类型的统计数据
3107
提取长期特征
3008
对长期音频类型分类
结束

图32

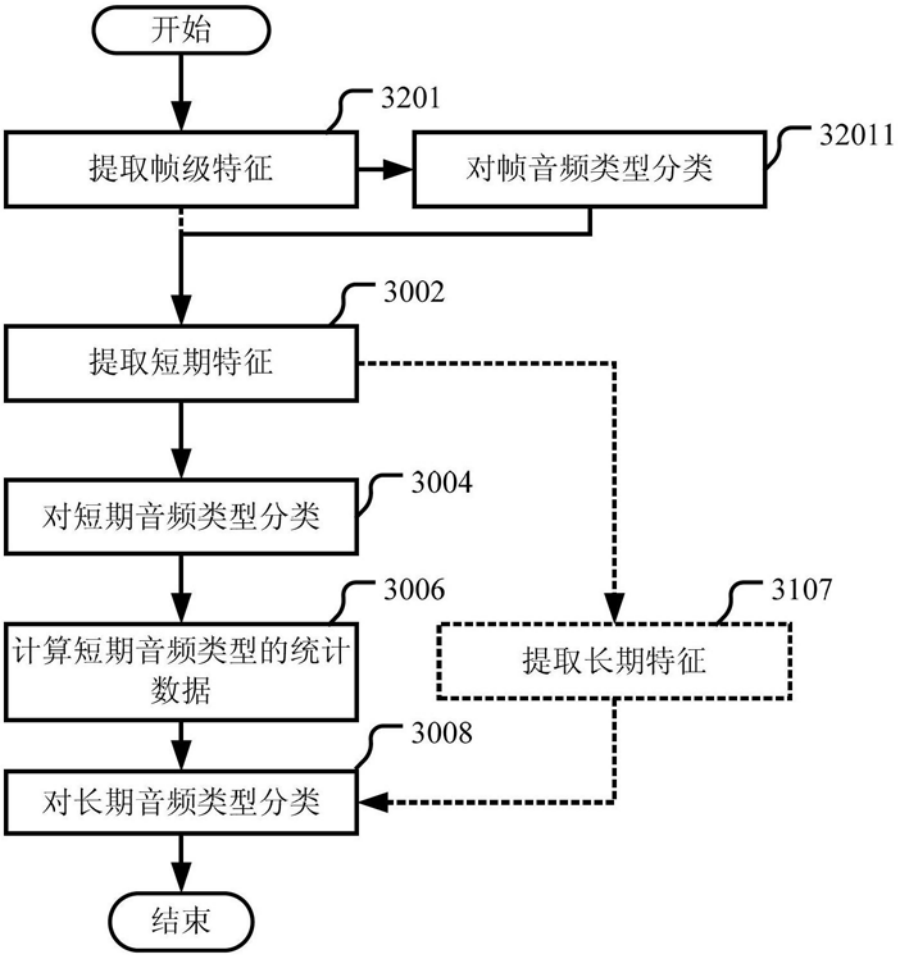

开始
3201
提取帧级特征
对帧音频类型分类
32011
3002
提取短期特征
3004
对短期音频类型分类
3006
计算短期音频类型的统计数据
3107
提取长期特征
3008
对长期音频类型分类
结束

图33

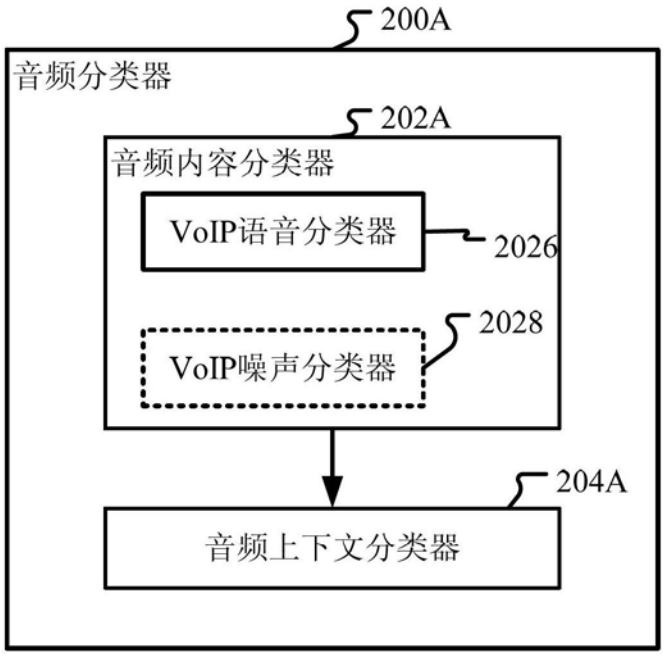

200A
音频分类器
202A
音频内容分类器
VoIP语音分类器
2026
2028
VoIP噪声分类器
204A
音频上下文分类器

图34

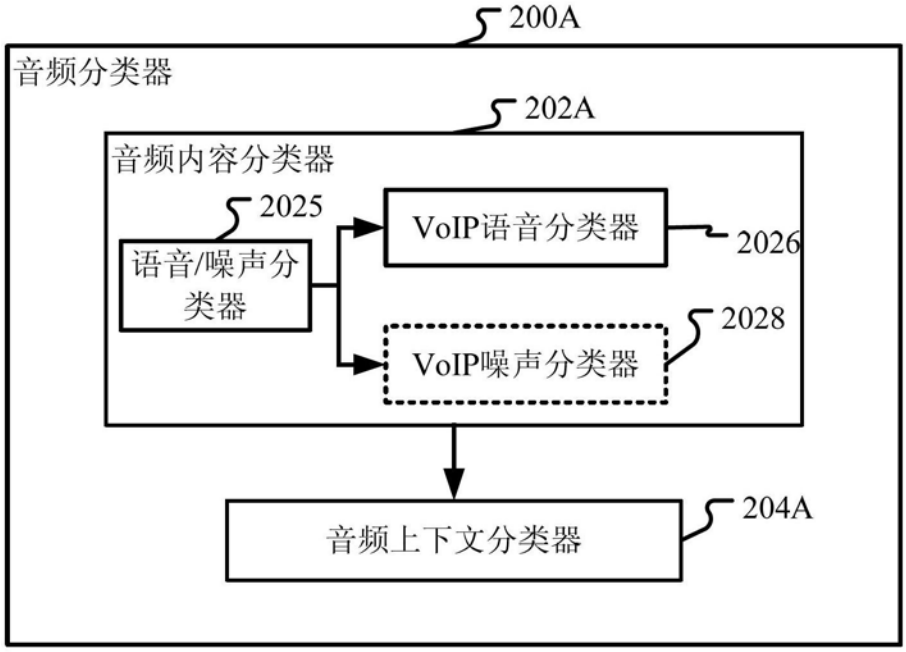

200A
音频分类器
202A
音频内容分类器
2025
语音/噪声分类器
VoIP语音分类器
2026
2028
VoIP噪声分类器
204A
音频上下文分类器

图35

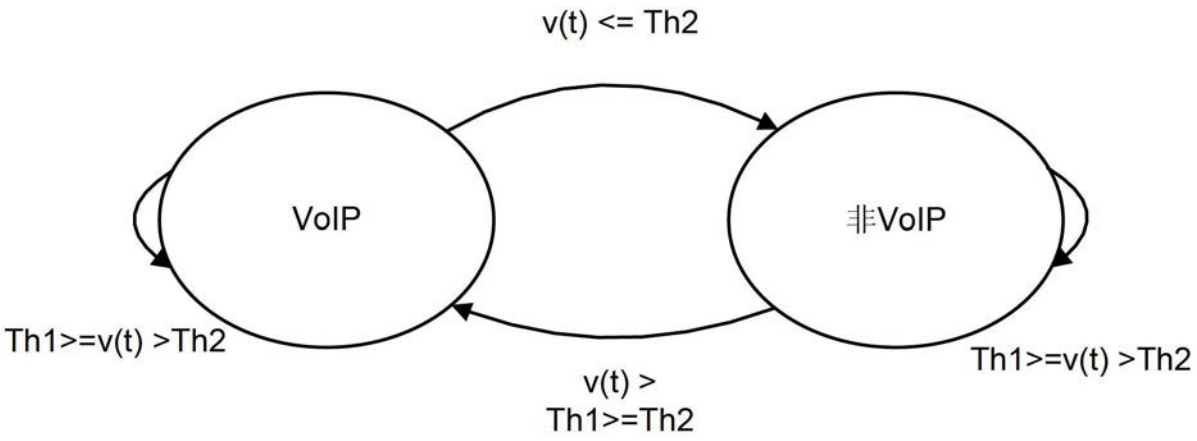

v(t) <= Th2
VoIP
非VoIP
Th1>=v(t) >Th2
Th1>=v(t) >Th2
v(t) >
Th1>=Th2

图36

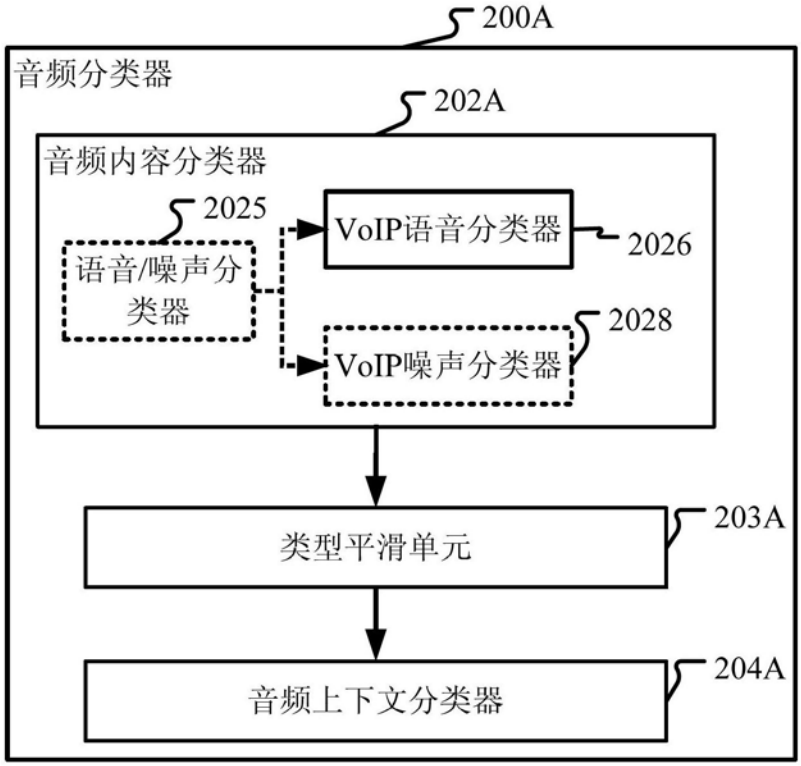

200A
音频分类器
202A
音频内容分类器
2025
语音/噪声分类器
VoIP语音分类器
2026
2028
VoIP噪声分类器
类型平滑单元
203A
音频上下文分类器
204A

图37

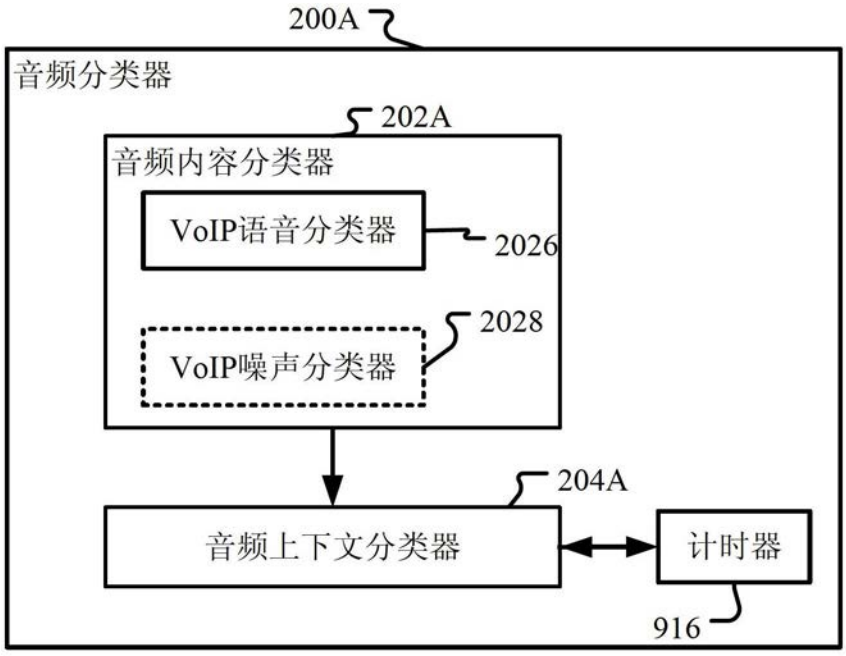

200A
音频分类器
202A
音频内容分类器
VoIP语音分类器
2026
2028
VoIP噪声分类器
204A
音频上下文分类器
计时器
916

图38

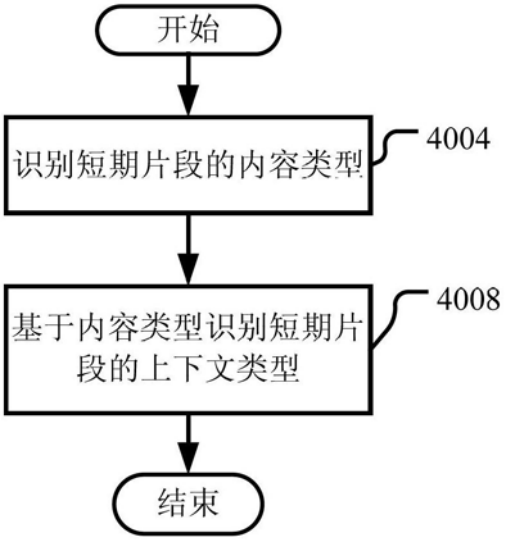

开始
识别短期片段的内容类型
4004
基于内容类型识别短期片段的上下文类型
4008
结束

图39

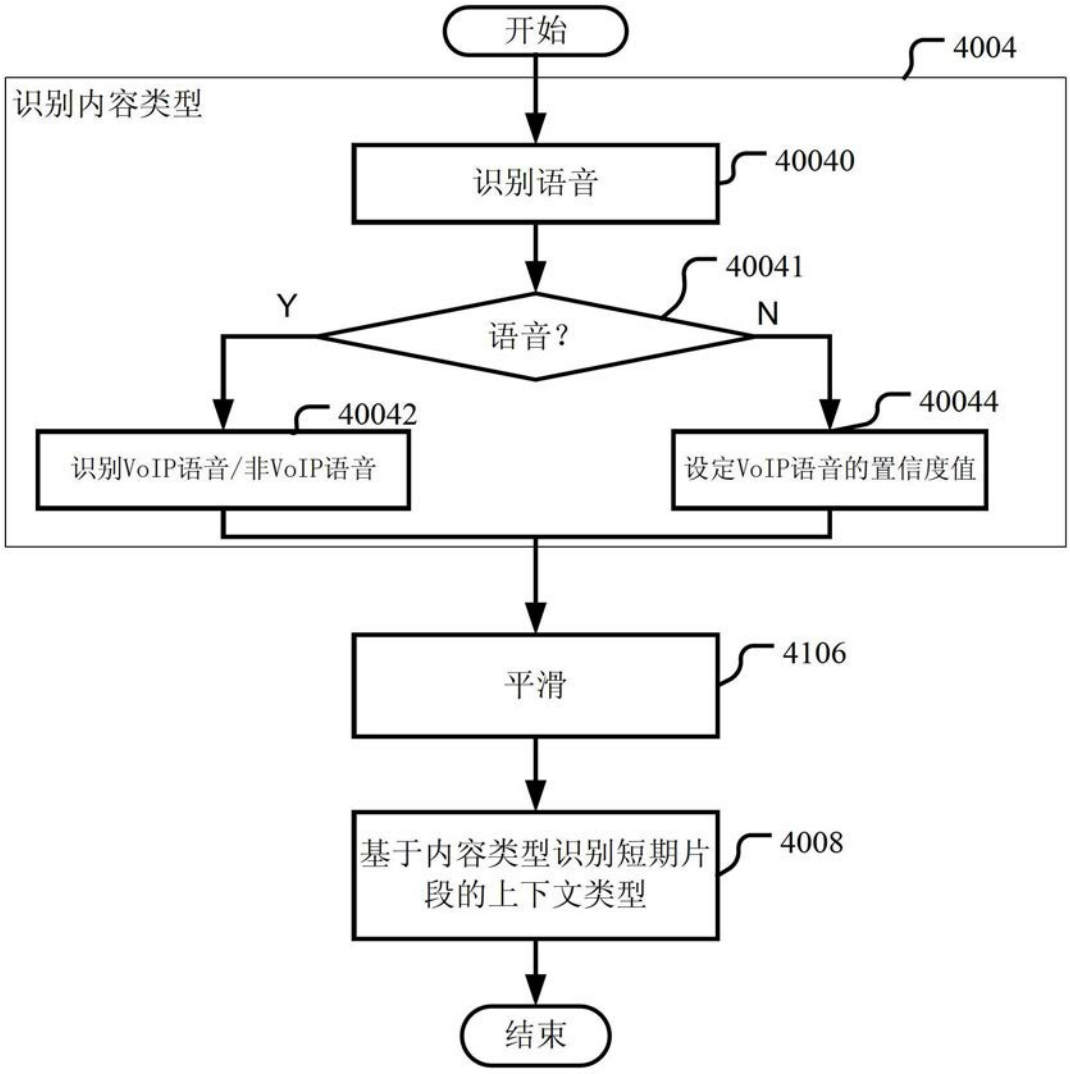

开始
4004
识别内容类型
识别语音
40040
40041
Y
语音？
N
40042
识别VoIP语音/非VoIP语音
40044
设定VoIP语音的置信度值
平滑
4106
基于内容类型识别短期片段的上下文类型
4008
结束

图40

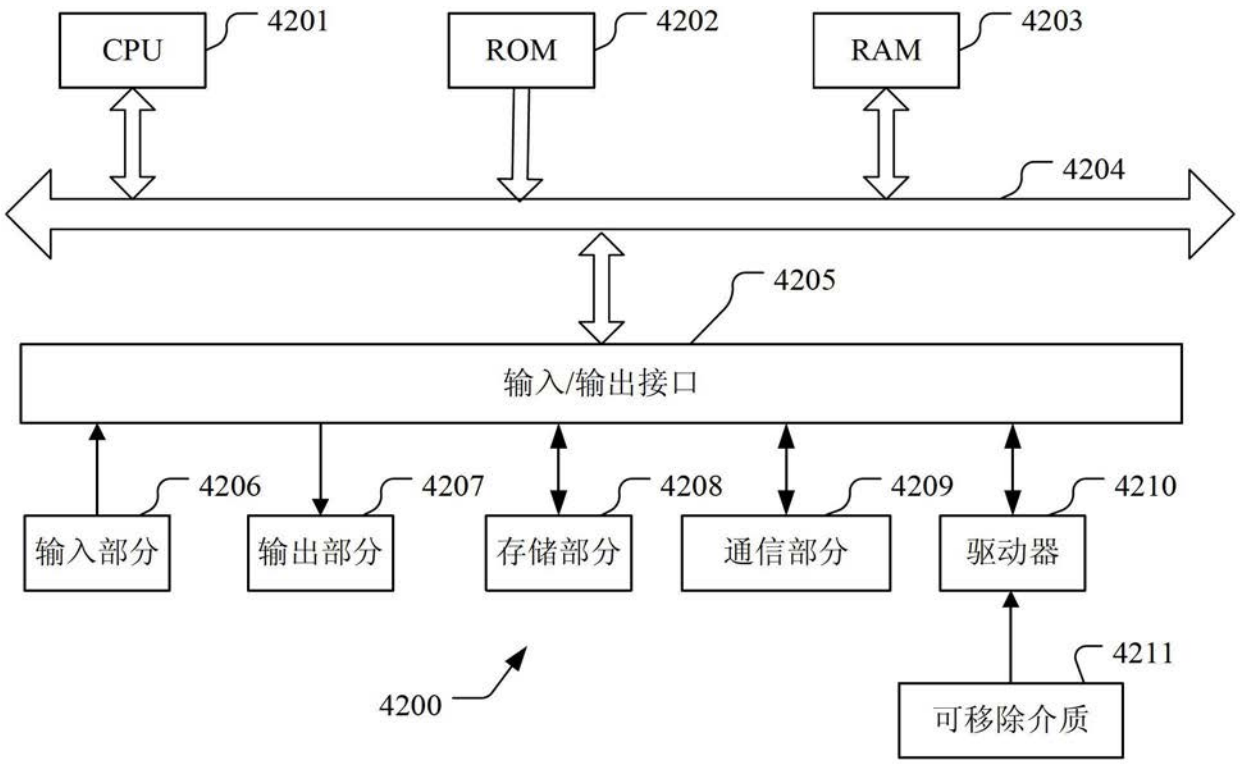

CPU
4201
ROM
4202
RAM
4203
4204
4205
输入/输出接口
4206
4207
4208
4209
4210
输入部分
输出部分
存储部分
通信部分
驱动器
4211
4200
可移除介质

图41