



(21) 申请号 202110227616.2

(22) 申请日 2021.03.01

(65) 同一申请的已公布的文献号  
申请公布号 CN 112951259 A

(43) 申请公布日 2021.06.11

(73) 专利权人 杭州网易云音乐科技有限公司  
地址 310052 浙江省杭州市萧山区钱江世  
纪城奔竞大道353号杭州国际博览中  
心A座1201室

(72) 发明人 潘颂声 曹偲 李鹏 赵翔宇  
刘华平

(74) 专利代理机构 北京律智知识产权代理有限  
公司 11438  
专利代理师 王辉 阚梓瑄

(51) Int.Cl.

G10L 21/0208 (2013.01)

G10L 21/028 (2013.01)

(56) 对比文件

CN 106971741 A, 2017.07.21

CN 108630221 A, 2018.10.09

审查员 陈文爽

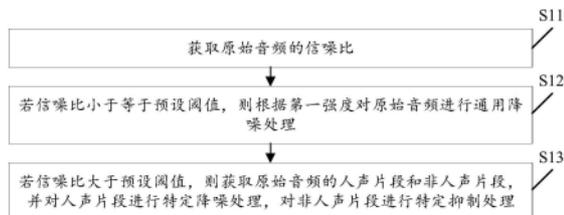
权利要求书4页 说明书21页 附图6页

#### (54) 发明名称

音频降噪方法、装置、电子设备及计算机可读存储介质

#### (57) 摘要

本申请的实施方式提供了一种音频降噪方法、音频降噪装置、电子设备以及计算机可读存储介质,涉及音频处理技术领域。该方法包括:获取原始音频的信噪比;若信噪比小于等于预设阈值,则根据第一强度对原始音频进行通用降噪处理;若信噪比大于预设阈值,则获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。可见,实施本申请实施例可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。



1. 一种音频降噪方法,其特征在于,包括:
  - 获取原始音频的信噪比;
  - 若所述信噪比小于等于预设阈值,则根据第一强度对所述原始音频进行通用降噪处理;
  - 若所述信噪比大于所述预设阈值,则获取所述原始音频的人声片段和非人声片段,并对所述人声片段进行特定降噪处理,对所述非人声片段进行特定抑制处理;
  - 其中,对所述非人声片段进行特定抑制处理,包括:
    - 按照预设衰减比例对各非人声片段的前边界音量和中间音频音量进行衰减淡出处理,直到所述前边界音量和所述中间音频音量衰减至目标阈值为止;
    - 对各非人声片段的后边界音量进行音量淡入处理,直到后边界音量达到相邻语音片段的前边界音量为止。
2. 根据权利要求1所述的方法,其特征在于,若所述信噪比大于所述预设阈值,还包括:
  - 根据第二强度对所述人声片段和/或所述非人声片段进行通用降噪处理;其中,所述第二强度小于所述第一强度。
3. 根据权利要求1所述的方法,其特征在于,获取原始音频的信噪比,包括:
  - 计算所述原始音频中各人声片段的能量均值以及所述原始音频中各非人声片段的能量均值;
  - 对所述各人声片段的能量均值进行求和,得到第一参数;
  - 对所述各非人声片段的能量均值进行求和,得到第二参数;
  - 根据所述第一参数和所述第二参数计算所述信噪比。
4. 根据权利要求3所述的方法,其特征在于,根据所述第一参数和所述第二参数计算所述信噪比,包括:
  - 计算所述第一参数与第一预设权重的第一乘积;
  - 计算所述第二参数与第二预设权重的第二乘积;
  - 根据所述第一乘积和所述第二乘积计算所述信噪比。
5. 根据权利要求1所述的方法,其特征在于,对所述人声片段进行特定降噪处理,包括:
  - 按照预处理窗长对所述原始音频进行加窗,得到多帧音频构成的参考音频;其中,所述多帧音频中各帧音频均对应于所述预处理窗长;
  - 通过特定噪声检测窗长对参考音频进行加窗,得到多段音频;其中,所述多段音频中各段音频均对应于所述特定噪声检测窗长,所述特定噪声检测窗长大于所述预处理窗长;
  - 若检测到所述多段音频中存在至少一个包含特定噪声的目标段音频,则对所述人声片段进行特定降噪处理。
6. 根据权利要求5所述的方法,其特征在于,所述多帧音频中相邻帧音频之间存在预设时长的音频重叠。
7. 根据权利要求5所述的方法,其特征在于,对所述人声片段进行特定降噪处理之前,还包括:
  - 确定所述多段音频中各段音频对应的音频特征;
  - 通过降维模型对所述各段音频对应的音频特征进行降维处理;
  - 通过分类模型确定降维处理后所述各段音频的音频特征对应的检测结果;其中,所述

检测结果用于表征音频中存在特定噪声的概率。

8. 根据权利要求7所述的方法,其特征在于,通过分类模型确定降维处理后所述各段音频的音频特征对应的检测结果,包括:

通过所述分类模型计算降维处理后所述各段音频的音频特征分别对应的特定噪声发生概率;

根据所述特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后所述各段音频的音频特征对应的检测结果。

9. 根据权利要求7所述的方法,其特征在于,确定所述多段音频中各段音频对应的音频特征,包括:

对所述各帧音频进行特征提取,得到所述各帧音频分别对应的音频特征;

对所述各段音频中分别包含的帧音频进行特征合并,得到所述各段音频对应的音频特征。

10. 根据权利要求7所述的方法,其特征在于,通过降维模型对所述各段音频对应的音频特征进行降维处理之前,还包括:

按照所述特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合;其中,所述特定噪声单元样本集合包括正样本和负样本;

通过所述正样本和所述负样本训练所述降维模型;

通过降维处理后的正样本和降维处理后的负样本训练所述分类模型。

11. 根据权利要求10所述的方法,其特征在于,通过降维处理后的正样本和降维处理后的负样本训练所述分类模型之后,还包括:

通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率;

根据所述特定噪声发生概率与所述测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

12. 根据权利要求1所述的方法,其特征在于,根据第一强度对所述原始音频进行通用降噪处理,包括:

对所述原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域;

确定频域中的各帧音频在每个频带的幅值和相位;

通过预训练的掩蔽模型对所述幅值进行掩蔽处理;

将所述相位和掩蔽处理后的幅值对应的原始音频由所述频域变换为所述时域,得到通用降噪处理后的原始音频。

13. 一种音频降噪装置,其特征在于,包括:

参数获取单元,用于获取原始音频的信噪比;

降噪处理单元,用于在所述信噪比小于等于预设阈值时,根据第一强度对所述原始音频进行通用降噪处理;

所述降噪处理单元,还用于在所述信噪比大于所述预设阈值时,获取所述原始音频的人声片段和非人声片段,并对所述人声片段进行特定降噪处理,对所述非人声片段进行特定抑制处理;

其中,所述降噪处理单元对所述非人声片段进行特定抑制处理,包括:

按照预设衰减比例对各非人声片段的前边界音量和中间音频音量进行衰减淡出处理,直到所述前边界音量和所述中间音频音量衰减至目标阈值为止;

对各非人声片段的后边界音量进行音量淡入处理,直到后边界音量达到相邻语音片段的前边界音量为止。

14. 根据权利要求13所述的装置,其特征在于,若所述信噪比大于所述预设阈值,所述降噪处理单元,还用于根据第二强度对所述人声片段和/或所述非人声片段进行通用降噪处理;其中,所述第二强度小于所述第一强度。

15. 根据权利要求13所述的装置,其特征在于,所述参数获取单元获取原始音频的信噪比,包括:

计算所述原始音频中各人声片段的能量均值以及所述原始音频中各非人声片段的能量均值;

对所述各人声片段的能量均值进行求和,得到第一参数;

对所述各非人声片段的能量均值进行求和,得到第二参数;

根据所述第一参数和所述第二参数计算所述信噪比。

16. 根据权利要求15所述的装置,其特征在于,所述参数获取单元根据所述第一参数和所述第二参数计算所述信噪比,包括:

计算所述第一参数与第一预设权重的第一乘积;

计算所述第二参数与第二预设权重的第二乘积;

根据所述第一乘积和所述第二乘积计算所述信噪比。

17. 根据权利要求13所述的装置,其特征在于,所述降噪处理单元对所述人声片段进行特定降噪处理,包括:

按照预处理窗长对所述原始音频进行加窗,得到多帧音频构成的参考音频;其中,所述多帧音频中各帧音频均对应于所述预处理窗长;

通过特定噪声检测窗长对参考音频进行加窗,得到多段音频;其中,所述多段音频中各段音频均对应于所述特定噪声检测窗长,所述特定噪声检测窗长大于所述预处理窗长;

若检测到所述多段音频中存在至少一个包含特定噪声的目标段音频,则对所述人声片段进行特定降噪处理。

18. 根据权利要求17所述的装置,其特征在于,所述多帧音频中相邻帧音频之间存在预设时长的音频重叠。

19. 根据权利要求17所述的装置,其特征在于,还包括:

特征提取单元,用于在所述降噪处理单元对所述人声片段进行特定降噪处理之前,确定所述多段音频中各段音频对应的音频特征;

特征降维单元,用于通过降维模型对所述各段音频对应的音频特征进行降维处理;

噪声检测单元,用于通过分类模型确定降维处理后所述各段音频的音频特征对应的检测结果;其中,所述检测结果用于表征音频中存在特定噪声的概率。

20. 根据权利要求19所述的装置,其特征在于,所述噪声检测单元通过分类模型确定降维处理后所述各段音频的音频特征对应的检测结果,包括:

通过所述分类模型计算降维处理后所述各段音频的音频特征分别对应的特定噪声发生概率;

根据所述特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后所述各段音频的音频特征对应的检测结果。

21. 根据权利要求19所述的装置,其特征在於,所述特征提取单元确定所述多段音频中各段音频对应的音频特征,包括:

对所述各帧音频进行特征提取,得到所述各帧音频分别对应的音频特征;

对所述各段音频中分别包含的帧音频进行特征合并,得到所述各段音频对应的音频特征。

22. 根据权利要求19所述的装置,其特征在於,还包括:

模型训练单元,用于在所述特征降维单元通过降维模型对所述各段音频对应的音频特征进行降维处理之前,按照所述特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合;其中,所述特定噪声单元样本集合包括正样本和负样本;通过所述正样本和所述负样本训练所述降维模型;通过降维处理后的正样本和降维处理后的负样本训练所述分类模型。

23. 根据权利要求22所述的装置,其特征在於,还包括:

模型测试单元,用于在所述模型训练单元通过降维处理后的正样本和降维处理后的负样本训练所述分类模型之后,通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率;根据所述特定噪声发生概率与所述测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

24. 根据权利要求13所述的装置,其特征在於,所述降噪处理单元根据第一强度对所述原始音频进行通用降噪处理,包括:

对所述原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域;

确定频域中的各帧音频在每个频带的幅值和相位;

通过预训练的掩蔽模型对所述幅值进行掩蔽处理;

将所述相位和掩蔽处理后的幅值对应的原始音频由所述频域变换为所述时域,得到通用降噪处理后的原始音频。

25. 一种电子设备,其特征在於,包括:

处理器;以及

存储器,所述存储器上存储有计算机可读指令,所述计算机可读指令被所述处理器执行时实现如权利要求1至12中任一项所述的音频降噪方法。

26. 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如权利要求1至12中任一项所述的音频降噪方法。

## 音频降噪方法、装置、电子设备及计算机可读存储介质

### 技术领域

[0001] 本申请的实施方式涉及音频处理技术领域,更具体地,本申请的实施方式涉及音频降噪方法、音频降噪装置、电子设备以及计算机可读存储介质。

### 背景技术

[0002] 随着计算机技术的不断发展,唱歌形式不再拘泥于线下KTV唱歌,而是应运而生了线上唱歌的软件,用户可以通过该软件满足线上唱歌的需求。

[0003] 一般来说,由于用户线上唱歌的环境不同,所录制的音频或多或少会存在一些噪声,而噪声容易导致音频播放效果不佳。因此,对音频降噪是非常重要的,现有的降噪方式通常为:在检测到用户完成录制后,通过降噪算法直接对音频进行降噪。

[0004] 但是,不同的用户录制时的环境可能是不同的,有的用户在较为安静的环境下录制音频,而有的用户在较为嘈杂的环境下录制音频。可以理解的是,不同环境下录制的音频的主导噪声是不同的,通过一概而论的方式进行音频降噪,容易造成降噪效果不佳的问题。

[0005] 需要说明的是,在上述背景技术部分公开的信息仅用于加强对本申请的背景的理解,因此,不构成对本领域普通技术人员已知的现有技术的信息。

### 发明内容

[0006] 基于上述问题,发明人进行了相应的思考,做出了有针对性的改进,提供了音频降噪方法、音频降噪装置、电子设备以及计算机可读存储介质,可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。

[0007] 根据本申请实施例的第一方面,公开了一种音频降噪方法,包括:

[0008] 获取原始音频的信噪比;

[0009] 若信噪比小于等于预设阈值,则根据第一强度对原始音频进行通用降噪处理;

[0010] 若信噪比大于预设阈值,则获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。

[0011] 在一个实施例中,基于前述方案,若信噪比大于预设阈值,上述方法还包括:

[0012] 根据第二强度对人声片段和/或非人声片段进行通用降噪处理;其中,第二强度小于第一强度。

[0013] 在一个实施例中,基于前述方案,获取原始音频的信噪比,包括:

[0014] 计算原始音频中各人声片段的能量均值以及原始音频中各非人声片段的能量均值;

[0015] 对各人声片段的能量均值进行求和,得到第一参数;

[0016] 对各非人声片段的能量均值进行求和,得到第二参数;

[0017] 根据第一参数和第二参数计算信噪比。

- [0018] 在一个实施例中,基于前述方案,根据第一参数和第二参数计算信噪比,包括:
- [0019] 计算第一参数与第一预设权重的第一乘积;
- [0020] 计算第二参数与第二预设权重的第二乘积;
- [0021] 根据第一乘积和第二乘积计算信噪比。
- [0022] 在一个实施例中,基于前述方案,对人声片段进行特定降噪处理,包括:
- [0023] 按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频;其中,多帧音频中各帧音频均对应于预处理窗长;
- [0024] 通过特定噪声检测窗长对参考音频进行加窗,得到多段音频;其中,多段音频中各段音频均对应于特定噪声检测窗长,特定噪声检测窗长大于预处理窗长;
- [0025] 若检测到多段音频中存在至少一个包含特定噪声的目标段音频,则对人声片段进行特定降噪处理。
- [0026] 在一个实施例中,基于前述方案,多帧音频中相邻帧音频之间存在预设时长的音频重叠。
- [0027] 在一个实施例中,基于前述方案,对人声片段进行特定降噪处理之前,上述方法还包括:
- [0028] 确定多段音频中各段音频对应的音频特征;
- [0029] 通过降维模型对各段音频对应的音频特征进行降维处理;
- [0030] 通过分类模型确定降维处理后各段音频的音频特征对应的检测结果;其中,检测结果用于表征音频中存在特定噪声的概率。
- [0031] 在一个实施例中,基于前述方案,通过分类模型确定降维处理后各段音频的音频特征对应的检测结果,包括:
- [0032] 通过分类模型计算降维处理后各段音频的音频特征分别对应的特定噪声发生概率;
- [0033] 根据特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后各段音频的音频特征对应的检测结果。
- [0034] 在一个实施例中,基于前述方案,确定多段音频中各段音频对应的音频特征,包括:
- [0035] 对各帧音频进行特征提取,得到各帧音频分别对应的音频特征;
- [0036] 对各段音频中分别包含的帧音频进行特征合并,得到各段音频对应的音频特征。
- [0037] 在一个实施例中,基于前述方案,通过降维模型对各段音频对应的音频特征进行降维处理之前,上述方法还包括:
- [0038] 按照特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合;其中,特定噪声单元样本集合包括正样本和负样本;
- [0039] 通过正样本和负样本训练降维模型;
- [0040] 通过降维处理后的正样本和降维处理后的负样本训练分类模型。
- [0041] 在一个实施例中,基于前述方案,通过降维处理后的正样本和降维处理后的负样本训练分类模型之后,上述方法还包括:
- [0042] 通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率;

[0043] 根据特定噪声发生概率与测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

[0044] 在一个实施例中,基于前述方案,对非人声片段进行特定抑制处理,包括:

[0045] 衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止;

[0046] 增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。

[0047] 在一个实施例中,基于前述方案,根据第一强度对原始音频进行通用降噪处理,包括:

[0048] 对原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域;

[0049] 确定频域中的各帧音频在每个频带的幅值和相位;

[0050] 通过预训练的掩蔽模型对幅值进行掩蔽处理;

[0051] 将相位和掩蔽处理后的幅值对应的原始音频由频域变换为所述时域,得到通用降噪处理后的原始音频。

[0052] 根据本申请实施例的第二方面,公开了一种音频降噪装置,包括:参数获取单元和降噪处理单元,其中:

[0053] 参数获取单元,用于获取原始音频的信噪比;

[0054] 降噪处理单元,用于在信噪比小于等于预设阈值时,根据第一强度对原始音频进行通用降噪处理;

[0055] 降噪处理单元,还用于在信噪比大于预设阈值时,获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。

[0056] 在一个实施例中,基于前述方案,若信噪比大于预设阈值,降噪处理单元,还用于根据第二强度对人声片段和/或非人声片段进行通用降噪处理;其中,第二强度小于第一强度。

[0057] 在一个实施例中,基于前述方案,参数获取单元获取原始音频的信噪比,包括:

[0058] 计算原始音频中各人声片段的能量均值以及原始音频中各非人声片段的能量均值;

[0059] 对各人声片段的能量均值进行求和,得到第一参数;

[0060] 对各非人声片段的能量均值进行求和,得到第二参数;

[0061] 根据第一参数和第二参数计算信噪比。

[0062] 在一个实施例中,基于前述方案,参数获取单元根据第一参数和第二参数计算信噪比,包括:

[0063] 计算第一参数与第一预设权重的第一乘积;

[0064] 计算第二参数与第二预设权重的第二乘积;

[0065] 根据第一乘积和第二乘积计算信噪比。

[0066] 在一个实施例中,基于前述方案,降噪处理单元对人声片段进行特定降噪处理,包括:

[0067] 按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频;其中,多帧音频中各帧音频均对应于预处理窗长;

[0068] 通过特定噪声检测窗长对参考音频进行加窗,得到多段音频;其中,多段音频中各段音频均对应于特定噪声检测窗长,特定噪声检测窗长大于预处理窗长;

[0069] 若检测到多段音频中存在至少一个包含特定噪声的目标段音频,则对人声片段进行特定降噪处理。

[0070] 在一个实施例中,基于前述方案,多帧音频中相邻帧音频之间存在预设时长的音频重叠。

[0071] 在一个实施例中,基于前述方案,上述装置还包括:

[0072] 特征提取单元,用于在降噪处理单元对人声片段进行特定降噪处理之前,确定多段音频中各段音频对应的音频特征;

[0073] 特征降维单元,用于通过降维模型对各段音频对应的音频特征进行降维处理;

[0074] 噪声检测单元,用于通过分类模型确定降维处理后各段音频的音频特征对应的检测结果;其中,检测结果用于表征音频中存在特定噪声的概率。

[0075] 在一个实施例中,基于前述方案,噪声检测单元通过分类模型确定降维处理后各段音频的音频特征对应的检测结果,包括:

[0076] 通过分类模型计算降维处理后各段音频的音频特征分别对应的特定噪声发生概率;

[0077] 根据特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后各段音频的音频特征对应的检测结果。

[0078] 在一个实施例中,基于前述方案,特征提取单元确定多段音频中各段音频对应的音频特征,包括:

[0079] 对各帧音频进行特征提取,得到各帧音频分别对应的音频特征;

[0080] 对各段音频中分别包含的帧音频进行特征合并,得到各段音频对应的音频特征。

[0081] 在一个实施例中,基于前述方案,上述装置还包括:

[0082] 模型训练单元,用于在特征降维单元通过降维模型对各段音频对应的音频特征进行降维处理之前,按照特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合;其中,特定噪声单元样本集合包括正样本和负样本;通过正样本和负样本训练降维模型;通过降维处理后的正样本和降维处理后的负样本训练分类模型。

[0083] 在一个实施例中,基于前述方案,上述装置还包括:

[0084] 模型测试单元,用于在模型训练单元通过降维处理后的正样本和降维处理后的负样本训练分类模型之后,通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率;根据特定噪声发生概率与测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

[0085] 在一个实施例中,基于前述方案,降噪处理单元对非人声片段进行特定抑制处理,包括:

[0086] 衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止;

[0087] 增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。

[0088] 在一个实施例中,基于前述方案,降噪处理单元根据第一强度对原始音频进行通

用降噪处理,包括:

[0089] 对原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域;

[0090] 确定频域中的各帧音频在每个频带的幅值和相位;

[0091] 通过预训练的掩蔽模型对幅值进行掩蔽处理;

[0092] 将相位和掩蔽处理后的幅值对应的原始音频由频域变换为所述时域,得到通用降噪处理后的原始音频。

[0093] 根据本申请实施例的第三方面,公开了一种电子设备,包括:处理器;以及存储器,存储器上存储有计算机可读指令,计算机可读指令被处理器执行时实现如第一方面公开的音频降噪方法。

[0094] 根据本申请实施例的第四方面,公开了一种计算机程序介质,其上存储有计算机可读指令,当计算机可读指令被计算机的处理器执行时,使计算机执行根据本申请第一方面公开的音频降噪方法。

[0095] 本申请实施例,可以获取原始音频的信噪比;若信噪比小于等于预设阈值,则根据第一强度对原始音频进行通用降噪处理;若信噪比大于预设阈值,则获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。相较现有技术,实施本申请的实施例,一方面可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。另一方面,实施本申请实施例,还能够在信噪比大于预设阈值的情况下(即,人声为主导的情况下),分别对人声片段和非人声片段采用不同的降噪手段,避免统一降噪过程对人声的损耗,即,可以在不损耗人声的前提下实现最佳降噪效果。

[0096] 本申请的其他特性和优点将通过下面的详细描述变得显然,或部分地通过本申请的实践而习得。

[0097] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性的,并不能限制本申请。

## 附图说明

[0098] 通过参考附图阅读下文的详细描述,本申请示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中,以示例性而非限制性的方式示出了本申请的若干实施方式,其中:

[0099] 图1示出的是根据本申请一示例实施方式的音频降噪方法的流程示意图;

[0100] 图2示出的是根据本申请一示例实施方式的语音分割结果示意图;

[0101] 图3示出的是根据本申请一示例实施方式的通用降噪模块示意图;

[0102] 图4示出的是根据本申请一示例实施方式的特定噪声判定模块示意图;

[0103] 图5示出的是根据本申请一示例实施方式的音频降噪方法的流程示意图;

[0104] 图6示出的是根据本申请一示例实施方式的音频降噪方法的流程示意图;

[0105] 图7示出的是根据本申请一可选示例实施方式的音频降噪系统的结构示意图;

[0106] 图8示出的是根据本申请一可选示例实施方式的音频降噪装置的结构框图;

[0107] 图9示出的是根据本申请另一可选示例实施方式的音频降噪装置的结构框图。

[0108] 在附图中,相同或对应的标号表示相同或对应的部分。

### 具体实施方式

[0109] 下面将参考若干示例性实施方式来描述本申请的原理和精神。应当理解,给出这些实施方式仅仅是为了使本领域技术人员能够更好地理解进而实现本申请,而并非以任何方式限制本申请的范围。相反,提供这些实施方式是为了使本申请更加透彻和完整,并且能够将本申请的范围完整地传达给本领域的技术人员。

[0110] 本领域技术人员知道,本申请的实施方式可以实现为一种装置、设备、方法或计算机程序产品。因此,本申请可以具体实现为以下形式,即:完全的硬件、完全的软件(包括固件、驻留软件、微代码等),或者硬件和软件结合的形式。

[0111] 根据本申请的实施方式,提出了一种音频降噪方法、音频降噪装置、电子设备以及计算机可读存储介质。

[0112] 附图中的任何元素数量均用于示例而非限制,以及任何命名都仅用于区分,而不具有任何限制含义。

[0113] 下面参考本申请的若干代表性实施方式,详细阐释本申请的原理和精神。

#### [0114] 发明概述

[0115] 现有的音频降噪算法主要针对语音通信领域,目的是消除背景噪声,提升语音的可懂度。针对语音通信领域的降噪算法主要包含:基于维纳滤波的信号处理算法和基于深度学习的降噪算法;其中,基于维纳滤波的信号处理算法能够处理稳态噪声,基于深度学习的降噪算法能够兼顾对于稳态噪声和非稳态噪声的降噪处理。

[0116] 一般来说,对于人声音频的降噪方式通常可以包括以下方式。

[0117] 方式一:通过对原始音频信号进行处理得到频谱特征,并将频谱特征输入卷积网络进行处理,以获得频谱卷积特征。进而,将频谱卷积特征输入循环网络进行处理,以获得目标频谱或目标频谱掩码,对目标频谱进行处理或使用目标频谱掩码,以获得目标音频信号,从而同时抑制背景环境中的稳态噪声和非稳态噪声。这种方式虽然可以同时抑制稳态噪声和非稳态噪声,但是对于音频录制场景(如,K歌场景)下容易出现的特定噪声(如,齿音、喷麦、触碰耳麦)不能很好地进行抑制,在安静环境下使用方式一进行音频降噪,容易对人声造成一定程度的损耗。

[0118] 方式二:通过对原始信号进行信号处理,得到以环境音为有效信号的第一有效信号和以通话人声为有效信号的第二有效信号,再基于第一有效信号进行场景识别,得到场景识别结果;基于场景识别结果选择相对应的降噪算法对第二有效信号进行降噪处理,得到输出信号。针对方法三,根据场景选择降噪算法意味着需要准备多套降噪算法,从而增加了系统的复杂度,如果场景识别判断错误,则将导致模型选择错误,进而导致降噪效果不佳的问题。

[0119] 基于上述问题,申请人想到可以根据原始音频的信噪比确定原始音频中的主导声音是人声还是背景声,进而根据不同的主导声音执行不同的降噪操作。特别地,还可以在主导声音为人声时,分别对人声片段进行特定抑制处理,对人声片段中的非人声部分进行特定抑制处理。这样可以达到在不用提升系统复杂度的情况下提升降噪效果,也能够解决现有技术对于音频直接使用降噪算法导致的人声损耗问题。

### [0120] 应用场景总览

[0121] 需要注意的是,下述应用场景仅是为了便于理解本申请的精神和原理而示出,本申请的实施方式在此方面不受任何限制。相反,本申请的实施方式可以应用于适用的任何场景。

[0122] 当本申请应用于音频降噪领域时,可以在接收到用户录入的原始音频时,计算出该原始音频的信噪比。若信噪比小于等于预设阈值,则根据第一强度对原始音频进行通用降噪处理;若信噪比大于预设阈值,则获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。其中,信噪比与预设阈值的比对结果用于表征原始音频中不同的主导声音,在不同的主导声音下,可以对应于不同的降噪手段。

[0123] 具体地,当主导声音为人声时,可以对原始音频中的人声片段进行特定降噪处理,对非人声片段进行特定抑制处理;当主导声音为背景声时,可以根据第一强度对原始音频进行通用降噪处理。

[0124] 可见,本申请在应用于音频降噪领域时,可以根据针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题,以及,还能够在主导声音为人声时,分别对人声片段和非人声片段采用不同的降噪手段,避免统一降噪过程对人声的损耗,即,可以在不损耗人声的前提下实现最佳降噪效果。

### [0125] 示例性方法

[0126] 下面结合上述的应用场景,参考图1和图6来描述根据本申请示例性实施方式的音频降噪方法。

[0127] 请参阅图1,图1示出的是根据本申请一示例实施方式的音频降噪方法的流程示意图,该音频降噪方法可以由服务器或终端设备来实现。如图1所示,该音频降噪方法,可以包括:

[0128] 步骤S110:获取原始音频的信噪比。

[0129] 步骤S120:若信噪比小于等于预设阈值,则根据第一强度对原始音频进行通用降噪处理。

[0130] 步骤S130:若信噪比大于预设阈值,则获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。

[0131] 实施图1所示的音频降噪方法,可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。此外,实施本申请实施例还能够在信噪比大于预设阈值的情况下(即,人声为主导的情况下),分别对人声片段和非人声片段采用不同的降噪手段,避免统一降噪过程对人声的损耗,即,可以在不损耗人声的前提下实现最佳降噪效果。

[0132] 下面对这些步骤进行详细描述。

[0133] 在步骤S110中,获取原始音频的信噪比。

[0134] 其中,需要说明的是,信噪比(SIGNAL-NOISE RATIO, SNR或S/N),是指电子设备或者电子系统中信号与噪声的比例,其中的信号是指是来自当前设备外部的需要通过当前设备进行处理的电子信号,噪声是指经过当前设备后产生的原信号中并不存在的无规则的额

外信号且该种信号不随原信号的变化而变化。一般来说,信噪比越高表明信号中的噪声越少,声音回放的音质越高,否则相反。另外,噪声中一般包括稳态噪声(steady noise)和非稳态噪声,稳态噪声是指频率相对稳定且周期不随时间变化的音频信号,非稳态噪声是与稳态噪声相对的在时间分布上不连续且噪声形态随时间不断变化的音频信号。

[0135] 具体地,原始音频可以为用户的实时录制音频也可以为用户的历史录制音频。基于此,获取原始音频的信噪比之前,上述方法还可以包括:当检测到开始录制操作时,通过麦克风模组采集音频,直到检测到停止录制操作为止,进而将采集到的音频确定为原始音频;或者,当检测到音频选择操作时,将音频选择操作对应的目标音频确定为原始音频,其中,音频选择操作可以用于从多个已录制好的历史音频中选择目标音频。

[0136] 作为一种可选的实施例,获取原始音频的信噪比,包括:计算原始音频中各人声片段的能量均值以及原始音频中各非人声片段的能量均值;对各人声片段的能量均值进行求和,得到第一参数;对各非人声片段的能量均值进行求和,得到第二参数;根据第一参数和第二参数计算信噪比。

[0137] 具体地,计算原始音频中各人声片段的能量均值以及原始音频中各非人声片段的能量均值之前,上述方法还可以包括:根据语音端点检测(VAD)算法对原始音频进行语音分割,得到人声片段集合 $(sb_1, se_1)$ 、 $\dots$ 、 $(sb_n, se_n)$ 和非人声片段集合 $(nb_1, ne_1)$ 、 $\dots$ 、 $(nb_m, ne_m)$ ;其中,人声片段集合中包含 $n$ 个人声片段 $(sb, se)$ , $sb$ 用于表示人声片段的起始点, $se$ 用于表示人声片段的终止点,非人声片段集合中包含 $m$ 个非人声片段 $(nb, ne)$ , $nb$ 用于表示非人声片段的起始点, $ne$ 于表示非人声片段的终止点,下标 $m$ 和下标 $n$ 为正整数。另外,需要说明的是,VAD用于进行语音边界检测,可以检测出原始音频中包含人声的语音片段,VAD具体可以为基于过零率、能量、音高、梅尔倒谱等多维特征构造分类器的传统算法或基于深度学习的端到端算法,本申请优选基于深度学习的端到端算法。

[0138] 举例来说,请参阅图2,图2示出的是根据本申请一示例实施方式的语音分割结果示意图。如图2所示,语音分割结果可以包含对于原始音频进行语音分割后得到的人声片段集合 $(sb_1, se_1)$ 、 $(sb_2, se_2)$ 、 $(sb_3, se_3)$ 和非人声片段集合 $(nb_1, ne_1)$ 、 $(nb_2, ne_2)$ 、 $(nb_3, ne_3)$ 、 $(nb_4, ne_4)$ 。其中, $(sb_1, se_1)$ 、 $(sb_2, se_2)$ 、 $(sb_3, se_3)$ 和 $(nb_1, ne_1)$ 、 $(nb_2, ne_2)$ 、 $(nb_3, ne_3)$ 、 $(nb_4, ne_4)$ 在原始音频中的的排列次序可以为 $(nb_1, ne_1)$ 、 $(sb_1, se_1)$ 、 $(nb_2, ne_2)$ 、 $(sb_2, se_2)$ 、 $(nb_3, ne_3)$ 、 $(sb_3, se_3)$ 、 $(nb_4, ne_4)$ 。

[0139] 基于此,可以计算原始音频中各人声片段的能量均值 $sE_1$ 以及原始音频中各非人声片段的能量均值 $nE_1$ ,进而,对各人声片段的能量均值进行求和,可以得到第一参数 $\sum_1^n sE_i$ ,对各非人声片段的能量均值进行求和,得到第二参数 $\sum_1^m nE_i$ 。

[0140] 可见,实施该可选的实施例,能够基于人声片段和非人声片段的能量计算得到可以作为主导声音判别依据的信噪比,从而有利于根据信噪比确定出不同主导声音对应的噪声抑制方式。

[0141] 作为一种可选的实施例,根据人声能量均值和非人声能量均值计算信噪比,包括:根据第一参数和第二参数计算信噪比,包括:计算第一参数与第一预设权重的第一乘积;计算第二参数与第二预设权重的第二乘积;根据第一乘积和第二乘积计算信噪比。

[0142] 具体地,计算第一参数 $\sum_1^n sE_i$ 与第一预设权重的第一乘积 $m * \sum_1^n sE_i$ ;其中,第

一预设权重可以为非人声片段集合的数量 $m$ ,也可以为预设常数。进而,计算第二参数 $\sum_1^m nE_i$ 与第二预设权重 $n$ 的第二乘积,第二预设权重可以为人声片段集合的数量 $n$ ,也可以为预设常数。

[0143] 基于此,根据第一乘积和第二乘积计算信噪比,包括:将第一乘积 $m * \sum_1^n sE_i$ 和第二乘积 $n * \sum_1^m nE_i$ 代入表达式 $snr = 10\log(\frac{m * \sum_1^n sE_i}{n * \sum_1^m nE_i})$ ,计算得到信噪比 $snr$ 。

[0144] 可见,实施该可选的实施例,能够根据各人声片段的能量均值之和(即,第一参数)以及各非人声片段的能量均值之和(即,第二参数)计算得到信噪比,从而可以便于根据该信噪比区分具体的噪声抑制方式,从而达到不增加系统复杂度的情况下,针对不同主导噪声进行相对应的噪声抑制。

[0145] 在步骤S120中,若信噪比小于等于预设阈值,则根据第一强度对原始音频进行通用降噪处理。

[0146] 具体地,预设阈值可以为预先设定的常数,第一强度用于表示噪声抑制强度,下述的第二强度同理。另外,在信噪比小于等于预设阈值时,可以判定原始音频中的主导声音为背景声。

[0147] 作为一种可选的实施例,根据第一强度对原始音频进行通用降噪处理,包括:对原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域;确定频域中的各帧音频在每个频带的幅值和相位;通过预训练的掩蔽模型对幅值进行掩蔽处理;将相位和掩蔽处理后的幅值对应的原始音频由频域变换为时域,得到通用降噪处理后的原始音频。

[0148] 具体地,通过预训练的掩蔽模型对幅值进行掩蔽处理之前,上述方法还可以包括:通过循环神经网络(如,RNN)训练掩蔽模型。进而通过预训练的掩蔽模型对幅值进行掩蔽处理,包括:通过对应于第一强度的预训练的掩蔽模型对幅值和相位进行掩蔽处理。

[0149] 请参阅图3,图3示出的是根据本申请一示例实施方式的通用降噪模块示意图。如图3所示,通用降噪模块包括:FFT(快速傅里叶变换)模块310、频带特征获取模块320、掩蔽模型330、掩蔽特征获取模块340和IFFT(反向快速傅里叶变换)模块350。其中,快速傅里叶变换(fast Fourier transform),是利用计算机计算离散傅里叶变换(DFT)的高效快速计算方法的统称,简称FFT,可以将信号的时域表达方式转换为频域表达方式。反向快速傅里叶变换(Invert Fast Fourier Transformation)是傅里叶变换的逆变换,可以将信号的频域表达方式转换为时域表达方式。

[0150] 具体地,可以将原始音频进行分帧处理,并将分帧处理得到的各帧音频输入FFT模块310,以使得FFT模块310将各帧音频由时域变换为频域,得到各帧音频在每个频带的幅值和相位,进而将幅值输入频带特征获取模块320并将相位输入IFFT模块350。频带特征获取模块320可以根据将幅值转化为频带特征输入掩蔽模型330,以使得掩蔽模型330对幅值进行掩蔽处理并将掩蔽处理结果输入掩蔽特征获取模块340。进而,掩蔽特征获取模块340可以将掩蔽处理结果转化为掩蔽幅值,并将掩蔽幅值输入IFFT模块350。IFFT模块350可以将相位和掩蔽幅值对应的原始音频由频域变换为时域,从而得到通用降噪处理后的原始音频。

[0151] 可见,实施该可选的实施例,能够通过强度更大的通用降噪处理方式消除原始音

频中的底噪,在主导声音为背景声的情况下,提升对于原始音频中噪声的抑制力度,有针对性地改善噪声抑制效果。

[0152] 在步骤S130中,若信噪比大于预设阈值,则获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。

[0153] 具体地,在信噪比大于预设阈值时,可以判定原始音频中的主导声音为人声。

[0154] 作为一种可选的实施例,若信噪比大于预设阈值,上述方法还包括:根据第二强度对人声片段和/或非人声片段进行通用降噪处理;其中,第二强度小于第一强度。

[0155] 具体地,第二强度对应的底噪消除力度小于第一强度对应的底噪消除力度,第二强度和第一强度对应的通用降噪处理中的掩蔽模型参数不同,即,根据第二强度对人声片段和/或非人声片段进行通用降噪处理的方式与根据第一强度对原始音频进行通用降噪处理的方式相同但参数不同。

[0156] 需要说明的时,根据第二强度对人声片段和/或非人声片段进行通用降噪处理,可以在对人声片段进行特定降噪处理以及对非人声片段进行特定抑制处理之前执行,也可以在对人声片段进行特定降噪处理以及对非人声片段进行特定抑制处理之后执行,本申请实施例不作限定。

[0157] 可见,实施该可选的实施例,能够通过强度更轻的通用降噪处理方式消除原始音频中的底噪,在主导声音为人声的情况下,有针对性地改善对于原始音频中噪声的抑制效果,减少对于原始音频中人声的损害。

[0158] 作为一种可选的实施例,对人声片段进行特定降噪处理,包括:按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频;其中,多帧音频中各帧音频均对应于预处理窗长;通过特定噪声检测窗长对参考音频进行加窗,得到多段音频;其中,多段音频中各段音频均对应于特定噪声检测窗长,特定噪声检测窗长大于预处理窗长;若检测到多段音频中存在至少一个包含特定噪声的目标段音频,则对人声片段进行特定降噪处理。

[0159] 具体地,特定降噪处理包括对于一种或多种特定噪声(如,喷麦噪声)进行噪声抑制处理,其中,喷麦指的是,说话人离耳麦过近导致音频中有“噗噗”的声音。预处理窗长为预处理窗口(如,汉明窗)的长度,可以通过时长进行表示(如,5ms);特定噪声检测窗长为特定噪声检测窗口的长度,也可以通过时长进行表示(如,10ms)。另外,多帧音频中相邻帧音频之间存在预设时长(如,1ms)的音频重叠,以保证特征平滑性。

[0160] 另外,按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频,包括:按照预处理窗长以及预设窗移(如,6ms)对原始音频进行加窗,得到多帧音频构成的参考音频。

[0161] 另外,对人声片段进行特定降噪处理之前,上述方法还可以包括:若至少一个包含特定噪声的目标段音频中存在相邻的连续目标段音频,则对相邻的连续目标段音频进行边界合并,根据合并后的目标段音频对人声片段进行特定降噪处理;其中,合并后的目标段音频的数量小于合并前的目标段音频的数量。

[0162] 可见,实施该可选的实施例,能够根据对于原始音频中包含特定噪声的音频段的检测,从而有利于根据包含特定噪声的音频段对原始音频中的人声片段进行特定噪声抑制,从而提升对于原始音频的噪声抑制效果。

[0163] 作为一种可选的实施例,对人声片段进行特定降噪处理之前,上述方法还包括:确

定多段音频中各段音频对应的音频特征;通过降维模型对各段音频对应的音频特征进行降维处理;通过分类模型确定降维处理后各段音频的音频特征对应的检测结果;其中,检测结果用于表征音频中存在特定噪声的概率(如,80%)。

[0164] 具体地,通过降维模型对各段音频对应的音频特征进行降维处理,包括:通过降维模型将各段音频对应的音频特征由原始维数降低至预设维数(如, $M$ 维, $M$ 为正整数);其中,降维模型可以采用PCA(Principal Component Analysis)算法,PCA用于高维数据的降维,可以提取数据的主要特征分量。

[0165] 请参阅图4,图4示出的是根据本申请一示例实施方式的特定噪声判定模块示意图。如图4所示,特定噪声判定模块中可以包括降维模型410和分类模型420。以原始音频中的音频段400为例,可以将确定出的音频段400对应的音频特征输入降维模型410,以使得降维模型410对音频特征进行降维处理并输出后的音频特征,作为分类模型420的输入,进而,分类模型420可以计算降维处理后的音频特征对应的特定噪声发生概率并输出,特定噪声发生概率可以作为音频段400中存在特定噪声的判定依据;其中,音频段400可以为通过特定噪声检测窗长对参考音频进行加窗得到的多段音频中的任一音频段。

[0166] 可见,实施该可选的实施例,能够通过对于音频特征的降维,提升分类模型的输出准确率,进而有利于改善对于特征噪声的抑制效果。

[0167] 作为一种可选的实施例,通过分类模型确定降维处理后各段音频的音频特征对应的检测结果,包括:通过分类模型计算降维处理后各段音频的音频特征分别对应的特定噪声发生概率;根据特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后各段音频的音频特征对应的检测结果。

[0168] 具体地,特定噪声判定条件用于作为每段音频中是否包含特定噪声的判定依据,例如,特定噪声判定条件可以包含判定阈值(如,60%),特定噪声判定条件也可以包含多个不存在交集的判定阈值范围(如,0%~30%、31%~60%、61%~100%)。

[0169] 基于此,若特定噪声判定条件也可以包含多个不存在交集的判定阈值范围,根据特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后各段音频的音频特征对应的检测结果,包括:将特定噪声发生概率与判定阈值进行比对,确定特定噪声发生概率所属的判定阈值范围,并确定该判定阈值范围对应的特定噪声存在概率(如,0.8),进而根据该特定噪声存在概率生成检测结果,检测结果可以通过文本进行表示,如,该特定噪声存在概率对应的音频段存在特定噪声的概率为0.8。

[0170] 可见,实施该可选的实施例,能够通过对于特定噪声发生概率的计算以及依据特定噪声发生概率对音频段的特定噪声判定,提升对于特定噪声的检测精度。

[0171] 作为一种可选的实施例,确定多段音频中各段音频对应的音频特征,包括:对各帧音频进行特征提取,得到各帧音频分别对应的音频特征;对各段音频中分别包含的帧音频进行特征合并,得到各段音频对应的音频特征。

[0172] 具体地,对各帧音频进行特征提取,得到各帧音频分别对应的音频特征,包括:提取各帧音频的 $N$ 维声学特征, $N$ 为正整数(如,20),从而得到各帧音频分别对应的 $N$ 维声学特征;其中, $N$ 维声学特征的特征类型可以为梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient,MFCC)、对数域的梅尔频率特征(FilterBank)或原始梅尔频率特征等,本申请实施例不作限定。基于此,各段音频对应的音频特征对应的声学特征维数可以为组成该段

的帧音频的声学特征维数之和(如,200)。

[0173] 可见,实施该可选的实施例,能够通过对各帧音频的音频特征合并,从而得到基于特定噪声检测窗口的每段音频对应的音频特征,从而有利于提升针对每段音频包含特定噪声的计算精度,进而有利于更准确地对人声片段进行特定噪声抑制,改善对于原始音频的噪声抑制效果。

[0174] 作为一种可选的实施例,通过降维模型对各段音频对应的音频特征进行降维处理之前,上述方法还包括:按照特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合;其中,特定噪声单元样本集合包括正样本和负样本;通过正样本和负样本训练降维模型;通过降维处理后的正样本和降维处理后的负样本训练分类模型。

[0175] 具体地,样本音频的数量可以为一个或多个,每个样本音频对应于一个特定噪声单元样本集合。

[0176] 另外,按照特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合,包括:按照特定噪声检测窗长对样本音频进行分割,得到多个样本音频段,将包含特定噪声的样本音频段标记为X(如,1),并将不包含特定噪声的样本音频段标记为Y(如,0),从而得到特定噪声单元样本集合,特定噪声单元样本集合中包括一个或多个包含特定噪声的样本音频段以及一个或多个不包含特定噪声的样本音频段;其中,X和Y可以表示为不同的字符。

[0177] 另外,通过正样本和负样本训练降维模型,包括:对正样本和负样本进行特征提取,并根据提取到的样本特征训练降维模型。

[0178] 可见,实施该可选的实施例,能够对降维模型和分类模型进行联合训练,从而有利于提升对于特定噪声的检测精度。

[0179] 作为一种可选的实施例,通过降维处理后的正样本和降维处理后的负样本训练分类模型之后,上述方法还包括:通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率;根据特定噪声发生概率与测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

[0180] 具体地,测试样本包含一个或多个特定噪声单元。

[0181] 另外,根据特定噪声发生概率与测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整,包括:根据特定噪声发生概率与测试样本的样本值之间的差计算损失函数,根据损失函数对训练后的降维模型和训练后的分类模型进行参数调整。

[0182] 可见,实施该可选的实施例,能够对训练后的降维模型和训练后的分类模型进行进一步测试以及参数修正,降低降维模型和分类模型的出错概率。

[0183] 作为一种可选的实施例,对非人声片段进行特定抑制处理,包括:衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止;增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。

[0184] 具体地,目标阈值可以为预设常数,如,1。基于此,衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止,包括:按照预设衰减比例对各非人声片段的前边界音量和中间音频音量进行衰减淡出处理,直到前边界音量和中间音频音量衰减至目标阈值为止。进而,增强处理各非人声片段

的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止,包括:对各非人声片段的后边界音量进行音量淡入处理,直到后边界音量达到相邻语音片段的前边界音量为止。

[0185] 可见,实施该可选的实施例,能够通过对于非人声片段的衰减处理以及增强处理,提升非人声片段的音量并提升非人声片段与人声片段衔接处的平滑性,改善原始音频的声音效果。

[0186] 请参阅图5,图5示出的是根据本申请一示例实施方式的音频降噪方法的流程示意图。如图5所示,音频降噪方法可以包括:步骤S500~步骤S560。

[0187] 步骤S500:获取原始音频的信噪比。

[0188] 步骤S510:检测信噪比是否大于预设阈值。如果是,则执行步骤S530。如果不是,则执行步骤S520。

[0189] 步骤S520:设置严模式。进而,执行步骤S560。具体地,在信噪比是否小于等于预设阈值时,可以设置严模式,以使得通用降噪处理基于第一强度对原始音频进行噪声抑制。

[0190] 步骤S530:对原始音频中的人声片段进行特定噪声检测和抑制。具体地,可以按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频,通过特定噪声检测窗长对参考音频进行加窗,得到多段音频,若检测到多段音频中存在至少一个包含特定噪声的目标段音频,则对人声片段进行特定降噪处理。

[0191] 步骤S540:对原始音频中的非人声片段进行抑制处理。具体地,可以衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止,进而增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。

[0192] 步骤S550:设置松模式。进而,执行步骤S560。具体地,在信噪比是否大于预设阈值时,可以设置松模式,以使得通用降噪处理基于第二强度对原始音频进行噪声抑制。

[0193] 步骤S560:执行通用降噪处理。

[0194] 可见,实施图5所示的方法,可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。此外,实施本申请实施例还能够能够在信噪比大于预设阈值的情况下(即,人声为主导的情况下),分别对人声片段和非人声片段采用不同的降噪手段,避免统一降噪过程对人声的损耗,即,可以在不损耗人声的前提下实现最佳降噪效果。

[0195] 请参阅图6,图6示出的是根据本申请一示例实施方式的音频降噪方法的流程示意图。如图6所示,该音频降噪方法包括:步骤S600~步骤S624。

[0196] 步骤S600:计算原始音频中各人声片段的能量均值以及原始音频中各非人声片段的能量均值,对各人声片段的能量均值进行求和,得到第一参数,对各非人声片段的能量均值进行求和,得到第二参数,根据第一参数和第二参数计算信噪比。

[0197] 步骤S602:计算第一参数与第一预设权重的第一乘积;计算第二参数与第二预设权重的第二乘积;根据第一乘积和第二乘积计算信噪比;若信噪比小于等于预设阈值,则执行步骤S604;若信噪比大于预设阈值,则执行步骤S606。

[0198] 步骤S604:对原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换

为频域,确定频域中的各帧音频在每个频带的幅值和相位,通过对应于第一强度的预训练的掩蔽模型对幅值和相位进行掩蔽处理,进而将相位和掩蔽处理后的幅值对应的原始音频由频域变换为所述时域,得到通用降噪处理后的原始音频。

[0199] 步骤S606:获取原始音频的人声片段和非人声片段。进而,执行步骤S608。

[0200] 步骤S608:对各帧音频进行特征提取,得到各帧音频分别对应的音频特征;对各段音频中分别包含的帧音频进行特征合并,得到各段音频对应的音频特征。

[0201] 步骤S610:按照特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合,其中,特定噪声单元样本集合包括正样本和负样本,进而通过正样本和负样本训练降维模型,通过降维处理后的正样本和降维处理后的负样本训练分类模型。

[0202] 步骤S612:通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率,根据特定噪声发生概率与测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

[0203] 步骤S614:通过训练后的降维模型对各段音频对应的音频特征进行降维处理。

[0204] 步骤S616:通过分类模型计算降维处理后各段音频的音频特征分别对应的特定噪声发生概率,根据特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后各段音频的音频特征对应的检测结果;其中,检测结果用于表征音频中存在特定噪声的概率。

[0205] 步骤S618:按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频;其中,多帧音频中各帧音频均对应于预处理窗长。

[0206] 步骤S620:通过特定噪声检测窗长对参考音频进行加窗,得到多段音频,若检测到多段音频中存在至少一个包含特定噪声的目标段音频,则对人声片段进行特定降噪处理;其中,多段音频中各段音频均对应于特定噪声检测窗长,特定噪声检测窗长大于预处理窗长。

[0207] 步骤S622:衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止,进而增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。

[0208] 步骤S624:对特定降噪处理以及衰减处理后的原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域,确定频域中的各帧音频在每个频带的幅值和相位,通过对应于第二强度的预训练的掩蔽模型对幅值和相位进行掩蔽处理,进而将掩蔽处理后的幅值和相位由频域变换为时域,得到噪声抑制后的原始音频。

[0209] 需要说明的是,步骤S600~步骤S624与图1所示的各步骤及其实施例相对应,针对步骤S600~步骤S624的具体实施方式,请参阅图1所示的各步骤及其实施例,此处不再赘述,

[0210] 可见,实施图6所示的方法,可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。此外,实施本申请实施例还能够在信噪比大于预设阈值的情况下(即,人声为主导的情况下),分别对人声片段和非人声片段采用不同的降噪手段,避免统一降噪过程对人声的损耗,即,可以在不损耗人声的前提下实现最佳降噪效果。

[0211] 此外,尽管在附图中以特定顺序描述了本申请中方法的各个步骤,但是,这并非要

求或者暗示必须按照该特定顺序来执行这些步骤,或是必须执行全部所示的步骤才能实现期望的结果。附加的或备选的,可以省略某些步骤,将多个步骤合并为一个步骤执行,以及/或者将一个步骤分解为多个步骤执行等。

#### [0212] 示例性系统

[0213] 在介绍了本申请示例性实施方式的方法之后,接下来,对本申请示例性系统进行说明。

[0214] 请参阅图7,图7示出的是根据本申请一可选示例实施方式的音频降噪系统的结构示意图。如图7所示,该音频降噪系统可以包括:第一降噪系统710和第二降噪系统720。其中,第一降噪系统710可以包括:VAD模块711、信噪比计算模块712、针对人声片段的特定降噪处理模块1 713、针对人声片段的特定降噪处理模块2 714、……、针对人声片段的特定降噪处理模块n 715和针对非人声片段的特定抑制处理模块716;其中,n为正整数。第二降噪系统720可以包括通用降噪模块721。

[0215] 需要说明的是,针对人声片段的特定降噪处理模块1 713、针对人声片段的特定降噪处理模块2 714、……、针对人声片段的特定降噪处理模块n 715分别用于处理不同的特定噪声。

[0216] 具体地,可以将原始音频输入第一降噪系统710,以使得第一降噪系统710中的VAD模块711对原始音频进行语音分割,得到人声片段集合和非人声片段集合。进而,信噪比计算模块712可以根据人声片段集合中各人声片段和非人声片段集合各非人声片段计算原始音频的信噪比。若信噪比大于预设阈值,则通过针对人声片段的特定降噪处理模块1 713、针对人声片段的特定降噪处理模块2 714、……、针对人声片段的特定降噪处理模块n 715中至少一种特定降噪处理模块对人声片段进行特定降噪处理;其中,针对人声片段的特定降噪处理模块1 713、针对人声片段的特定降噪处理模块2 714、……、针对人声片段的特定降噪处理模块n 715对于特定噪声的检测方式可以为按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频,通过特定噪声检测窗长对参考音频进行加窗,得到多段音频,检测到多段音频中是否存在至少一个包含特定噪声的目标段音频。进而,再通过针对非人声片段的特定抑制处理模块716衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止,进而增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。进而,通用降噪模块721根据第二强度对特定降噪处理以及衰减处理后的原始音频进行通用降噪处理,得到通用降噪处理后的原始音频。若信噪比小于等于预设阈值,则通用降噪模块721可以根据第一强度对原始音频进行通用降噪处理,得到通用降噪处理后的原始音频。

[0217] 可见,实施图7所示的系统,可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。此外,实施本申请实施例还能够在信噪比大于预设阈值的情况下(即,人声为主导的情况下),分别对人声片段和非人声片段采用不同的降噪手段,避免统一降噪过程对人声的损耗,即,可以在不损耗人声的前提下实现最佳降噪效果。

#### [0218] 示例性介质

[0219] 在介绍了本申请示例性系统之后,接下来,对本申请示例性实施方式的介质进行

说明。

[0220] 在一些可能的实施方式中,本申请的各个方面还可以实现为一种介质,其上存储有程序代码,当程序代码被设备的处理器执行时用于实现本说明书上述“示例性方法”部分中描述的根据本申请各种示例性实施方式的音频降噪方法中的步骤。

[0221] 具体地,所述设备的处理器执行所述程序代码时用于实现如下步骤:获取原始音频的信噪比;若信噪比小于等于预设阈值,则根据第一强度对原始音频进行通用降噪处理;若信噪比大于预设阈值,则获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。

[0222] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:计算原始音频中各人声片段的能量均值以及原始音频中各非人声片段的能量均值;对各人声片段的能量均值进行求和,得到第一参数;对各非人声片段的能量均值进行求和,得到第二参数;根据第一参数和第二参数计算信噪比。

[0223] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:计算第一参数与第一预设权重的第一乘积;计算第二参数与第二预设权重的第二乘积;根据第一乘积和第二乘积计算信噪比。

[0224] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:根据第二强度对人声片段和/或非人声片段进行通用降噪处理;其中,第二强度小于第一强度。

[0225] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:对原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域;确定频域中的各帧音频在每个频带的幅值和相位;通过预训练的掩蔽模型对幅值进行掩蔽处理;将相位和掩蔽处理后的幅值对应的原始音频由频域变换为所述时域,得到通用降噪处理后的原始音频。

[0226] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频;其中,多帧音频中各帧音频均对应于预处理窗长;通过特定噪声检测窗长对参考音频进行加窗,得到多段音频;其中,多段音频中各段音频均对应于特定噪声检测窗长,特定噪声检测窗长大于预处理窗长;若检测到多段音频中存在至少一个包含特定噪声的目标段音频,则对人声片段进行特定降噪处理。

[0227] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:确定多段音频中各段音频对应的音频特征;通过降维模型对各段音频对应的音频特征进行降维处理;通过分类模型确定降维处理后各段音频的音频特征对应的检测结果;其中,检测结果用于表征音频中存在特定噪声的概率。

[0228] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:通过分类模型计算降维处理后各段音频的音频特征分别对应的特定噪声发生概率;根据特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后各段音频的音频特征对应的检测结果。

[0229] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:对各帧音频进行特征提取,得到各帧音频分别对应的音频特征;对各段音频中分

别包含的帧音频进行特征合并,得到各段音频对应的音频特征。

[0230] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:按照特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合;其中,特定噪声单元样本集合包括正样本和负样本;通过正样本和负样本训练降维模型;通过降维处理后的正样本和降维处理后的负样本训练分类模型。

[0231] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率;根据特定噪声发生概率与测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

[0232] 在本申请的一些实施方式中,所述设备的处理器执行所述程序代码时还用于实现如下步骤:衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止;增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。

[0233] 需要说明的是:上述的介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以是但不限于:电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0234] 可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了可读程序代码。这种传播的数据信号可以采用多种形式,包括但不限于:电磁信号、光信号或上述的任意合适的组合。可读信号介质还可以是可读存储介质以外的任何可读介质,该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0235] 可读信号介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、有线、光缆、RF等,或者上述的任意合适的组合。另外,可以以一种或多种程序设计语言的任意组合来编写用于执行本申请操作的程序代码,所述程序设计语言包括面向对象的程序设计语言—诸如Java、C++等,还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算设备上执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中,远程计算设备可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算设备,或者,可以连接到外部计算设备(例如利用因特网服务提供商来通过因特网连接)。

[0236] 示例性装置

[0237] 在介绍了本申请示例性实施方式的介质之后,接下来,参考图8对本申请示例性实施方式的音频降噪装置进行说明。

[0238] 请参阅图8,图8示出的是根据本申请一示例实施方式的音频降噪装置的结构框图。如图8所示,本申请一示例实施方式的音频降噪装置800包括:参数获取单元801和降噪处理单元802,其中:

- [0239] 参数获取单元801,用于获取原始音频的信噪比;
- [0240] 降噪处理单元802,用于在信噪比小于等于预设阈值时,根据第一强度对原始音频进行通用降噪处理;
- [0241] 降噪处理单元802,还用于在信噪比大于预设阈值时,获取原始音频的人声片段和非人声片段,并对人声片段进行特定降噪处理,对非人声片段进行特定抑制处理。
- [0242] 可见,实施图8所示的装置,可以通过原始音频的信噪比区分不同的噪声情况(例如,信噪比小于等于预设阈值的情况、信噪比大于预设阈值的情况),针对不同的噪声情况采用不同的降噪手段,以改善现有技术中存在的降噪效果不佳的问题。此外,实施本申请实施例还能够在信噪比大于预设阈值的情况下(即,人声为主导的情况下),分别对人声片段和非人声片段采用不同的降噪手段,避免统一降噪过程对人声的损耗,即,可以在不损耗人声的前提下实现最佳降噪效果。
- [0243] 在一个实施例中,基于前述方案,若信噪比大于预设阈值,降噪处理单元802,还用于根据第二强度对人声片段和/或非人声片段进行通用降噪处理;其中,第二强度小于第一强度。
- [0244] 可见,实施该可选的实施例,能够通过强度更轻的通用降噪处理方式消除原始音频中的底噪,在主导声音为人声的情况下,有针对性地改善对于原始音频中噪声的抑制效果,减少对于原始音频中人声的损害。
- [0245] 在一个实施例中,基于前述方案,参数获取单元801获取原始音频的信噪比,包括:
- [0246] 计算原始音频中各人声片段的能量均值以及原始音频中各非人声片段的能量均值;
- [0247] 对各人声片段的能量均值进行求和,得到第一参数;
- [0248] 对各非人声片段的能量均值进行求和,得到第二参数;
- [0249] 根据第一参数和第二参数计算信噪比。
- [0250] 可见,实施该可选的实施例,能够基于人声片段和非人声片段的能量计算得到可以作为主导声音判别依据的信噪比,从而有利于根据信噪比确定出不同主导声音对应的噪声抑制方式。
- [0251] 在一个实施例中,基于前述方案,参数获取单元801根据第一参数和第二参数计算信噪比,包括:
- [0252] 计算第一参数与第一预设权重的第一乘积;
- [0253] 计算第二参数与第二预设权重的第二乘积;
- [0254] 根据第一乘积和第二乘积计算信噪比。
- [0255] 可见,实施该可选的实施例,能够根据各人声片段的能量均值之和(即,第一参数)以及各非人声片段的能量均值之和(即,第二参数)计算得到信噪比,从而可以便于根据该信噪比区分具体的噪声抑制方式,从而达到不增加系统复杂度的情况下,针对不同主导噪声进行相对应的噪声抑制。
- [0256] 在一个实施例中,基于前述方案,降噪处理单元802对人声片段进行特定降噪处理,包括:
- [0257] 按照预处理窗长对原始音频进行加窗,得到多帧音频构成的参考音频;其中,多帧音频中各帧音频均对应于预处理窗长;

[0258] 通过特定噪声检测窗长对参考音频进行加窗,得到多段音频;其中,多段音频中各段音频均对应于特定噪声检测窗长,特定噪声检测窗长大于预处理窗长;

[0259] 若检测到多段音频中存在至少一个包含特定噪声的目标段音频,则对人声片段进行特定降噪处理。

[0260] 其中,多帧音频中相邻帧音频之间存在预设时长的音频重叠。

[0261] 可见,实施该可选的实施例,能够根据对于原始音频中包含特定噪声的音频段的检测,从而有利于根据包含特定噪声的音频段对原始音频中的人声片段进行特定噪声抑制,从而提升对于原始音频的噪声抑制效果。

[0262] 在一个实施例中,基于前述方案,上述装置还包括:

[0263] 特征提取单元(未图示),用于在降噪处理单元802对人声片段进行特定降噪处理之前,确定多段音频中各段音频对应的音频特征;

[0264] 特征降维单元(未图示),用于通过降维模型对各段音频对应的音频特征进行降维处理;

[0265] 噪声检测单元(未图示),用于通过分类模型确定降维处理后各段音频的音频特征对应的检测结果;其中,检测结果用于表征音频中存在特定噪声的概率。

[0266] 可见,实施该可选的实施例,能够通过对于音频特征的降维,提升分类模型的输出准确率,进而有利于改善对于特征噪声的抑制效果。

[0267] 在一个实施例中,基于前述方案,噪声检测单元通过分类模型确定降维处理后各段音频的音频特征对应的检测结果,包括:

[0268] 通过分类模型计算降维处理后各段音频的音频特征分别对应的特定噪声发生概率;

[0269] 根据特定噪声发生概率与特定噪声判定条件的比对结果生成降维处理后各段音频的音频特征对应的检测结果。

[0270] 可见,实施该可选的实施例,能够通过对于特定噪声发生概率的计算以及依据特定噪声发生概率对音频段的特定噪声判定,提升对于特定噪声的检测精度。

[0271] 在一个实施例中,基于前述方案,特征提取单元确定多段音频中各段音频对应的音频特征,包括:

[0272] 对各帧音频进行特征提取,得到各帧音频分别对应的音频特征;

[0273] 对各段音频中分别包含的帧音频进行特征合并,得到各段音频对应的音频特征。

[0274] 可见,实施该可选的实施例,能够通过对各帧音频的音频特征合并,从而得到基于特定噪声检测窗口的每段音频对应的音频特征,从而有利于提升针对每段音频包含特定噪声的计算精度,进而有利于更准确地对人声片段进行特定噪声抑制,改善对于原始音频的噪声抑制效果。

[0275] 在一个实施例中,基于前述方案,上述装置还包括:

[0276] 模型训练单元(未图示),用于在特征降维单元通过降维模型对各段音频对应的音频特征进行降维处理之前,按照特定噪声检测窗长对样本音频进行标记,得到特定噪声单元样本集合;其中,特定噪声单元样本集合包括正样本和负样本;通过正样本和负样本训练降维模型;通过降维处理后的正样本和降维处理后的负样本训练分类模型。

[0277] 可见,实施该可选的实施例,能够对降维模型和分类模型进行联合训练,从而有利

于提升对于特定噪声的检测精度。

[0278] 在一个实施例中,基于前述方案,上述装置还包括:

[0279] 模型测试单元(未图示),用于在模型训练单元通过降维处理后的正样本和降维处理后的负样本训练分类模型之后,通过训练后的降维模型对测试样本进行降维处理,并通过训练后的分类模型计算降维处理结果对应的特定噪声发生概率;根据特定噪声发生概率与测试样本的样本值之间的差对训练后的降维模型和训练后的分类模型进行参数调整。

[0280] 可见,实施该可选的实施例,能够对训练后的降维模型和训练后的分类模型进行进一步测试以及参数修正,降低降维模型和分类模型的出错概率。

[0281] 在一个实施例中,基于前述方案,降噪处理单元802对非人声片段进行特定抑制处理,包括:

[0282] 衰减处理原始音频中各非人声片段的前边界音量和中间音频音量,直到前边界音量和中间音频音量均满足目标阈值为止;

[0283] 增强处理各非人声片段的后边界音量,直到后边界音量达到相邻语音片段的前边界音量为止。

[0284] 可见,实施该可选的实施例,能够通过对于非人声片段的衰减处理以及增强处理,提升非人声片段的音量并提升非人声片段与人声片段衔接处的平滑性,改善原始音频的声音效果。

[0285] 在一个实施例中,基于前述方案,降噪处理单元802根据第一强度对原始音频进行通用降噪处理,包括:

[0286] 对原始音频进行分帧处理,并将分帧处理得到的各帧音频由时域变换为频域;

[0287] 确定频域中的各帧音频在每个频带的幅值和相位;

[0288] 通过预训练的掩蔽模型对幅值进行掩蔽处理;

[0289] 将相位和掩蔽处理后的幅值对应的原始音频由频域变换为所述时域,得到通用降噪处理后的原始音频。

[0290] 可见,实施该可选的实施例,能够通过强度更大的通用降噪处理方式消除原始音频中的底噪,在主导声音为背景声的情况下,提升对于原始音频中噪声的抑制力度,有针对性地改善噪声抑制效果。

[0291] 应当注意,尽管在上文详细描述中提及了音频降噪装置的若干模块或者单元,但是这种划分并非强制性的。实际上,根据本申请的实施方式,上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之,上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[0292] 示例性电子设备

[0293] 在介绍了本申请示例性实施方式的方法、介质和装置之后,接下来,介绍根据本申请的另一示例性实施方式的电子设备。

[0294] 所属技术领域的技术人员能够理解,本申请的各个方面可以实现为系统、方法或程序产品。因此,本申请的各个方面可以具体实现为以下形式,即:完全的硬件实施方式、完全的软件实施方式(包括固件、微代码等),或硬件和软件方面结合的实施方式,这里可以统称为“电路”、“模块”或“系统”。

[0295] 下面参照图9来描述根据本申请的又一可选示例实施方式的音频降噪装置900。图

9显示的音频降噪装置900仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0296] 如图9所示,音频降噪装置900以电子设备的形式表现。音频降噪装置900的组件可以包括但不限于:上述至少一个处理单元910、上述至少一个存储单元920、连接不同系统组件(包括存储单元920和处理单元910)的总线930。

[0297] 其中,所述存储单元存储有程序代码,所述程序代码可以被所述处理单元910执行,使得所述处理单元910执行本说明书上述示例性方法的描述部分中描述的根据本申请各种示例性实施方式的步骤。例如,所述处理单元910可以执行如图1和图6中所示的各个步骤。

[0298] 存储单元920可以包括易失性存储单元形式的可读介质,例如随机存取存储单元(RAM)9201和/或高速缓存存储单元9202,还可以进一步包括只读存储单元(ROM)9203。

[0299] 存储单元920还可以包括具有一组(至少一个)程序模块9205的程序/实用工具9204,这样的程序模块9205包括但不限于:操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。

[0300] 总线930可以为表示几类总线结构中的一种或多种,包括地址总线、控制总线和/或数据总线。

[0301] 音频降噪装置900也可以与一个或多个外部设备1000(例如键盘、指向设备、蓝牙设备等)通信,还可与一个或者多个使得用户能与音频降噪装置900交互的设备通信,和/或与使得该音频降噪装置900能与一个或多个其它计算设备进行通信的任何设备(例如路由器、调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口950进行。并且,音频降噪装置900还可以通过网络适配器960与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图9所示,网络适配器960通过总线930与音频降噪装置900的其它模块通信。应当明白,尽管图中未示出,可以结合音频降噪装置900使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0302] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本申请实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、终端装置、或者网络设备等)执行根据本申请实施方式的方法。

[0303] 虽然已经参考若干具体实施方式描述了本申请的精神和原理,但是应该理解,本申请并不限于所发明的具体实施方式,对各方面的划分也不意味着这些方面中的特征不能组合以进行受益,这种划分仅是为了表述的方便。本申请旨在涵盖所附权利要求的精神和范围内所包括的各种修改和等同布置。

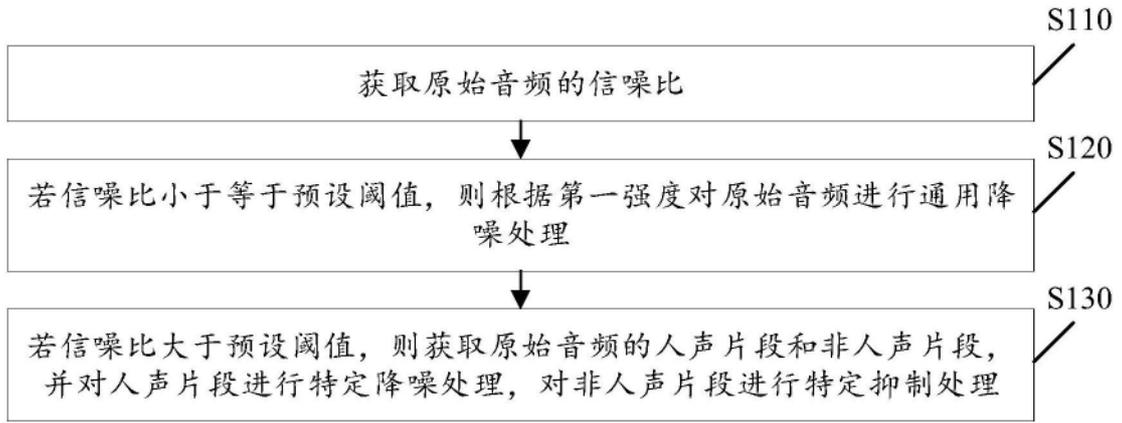


图1

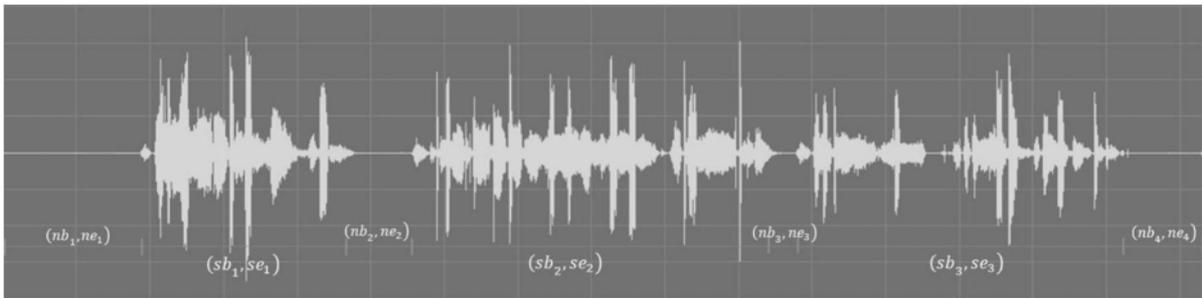


图2

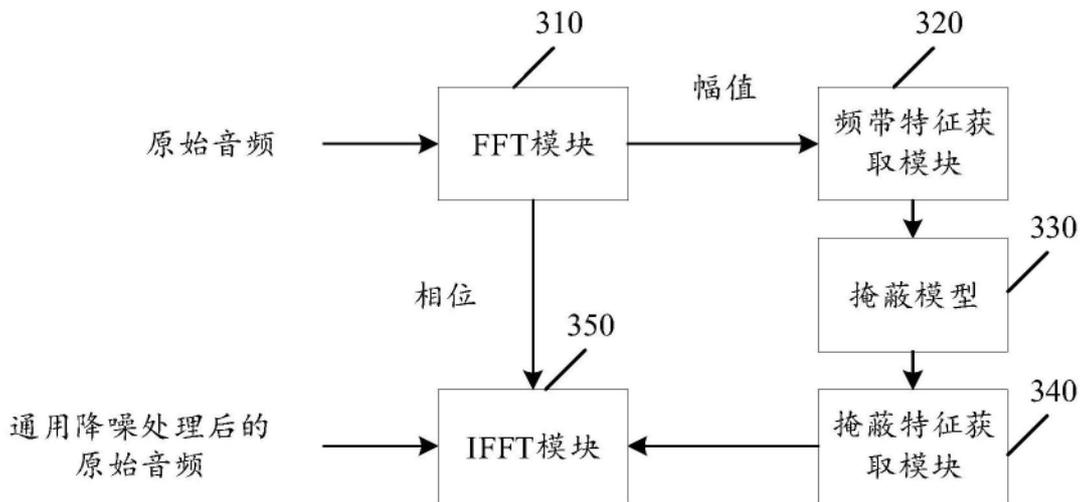


图3

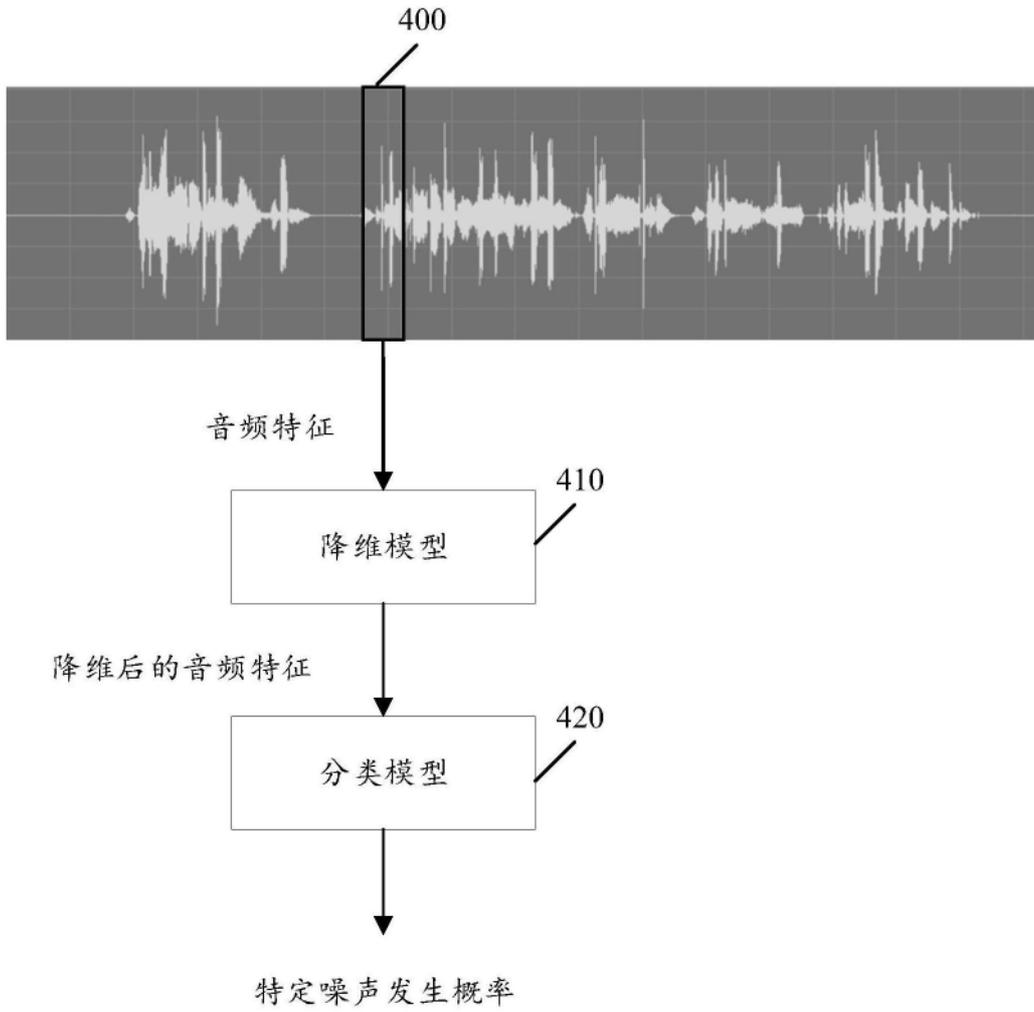


图4

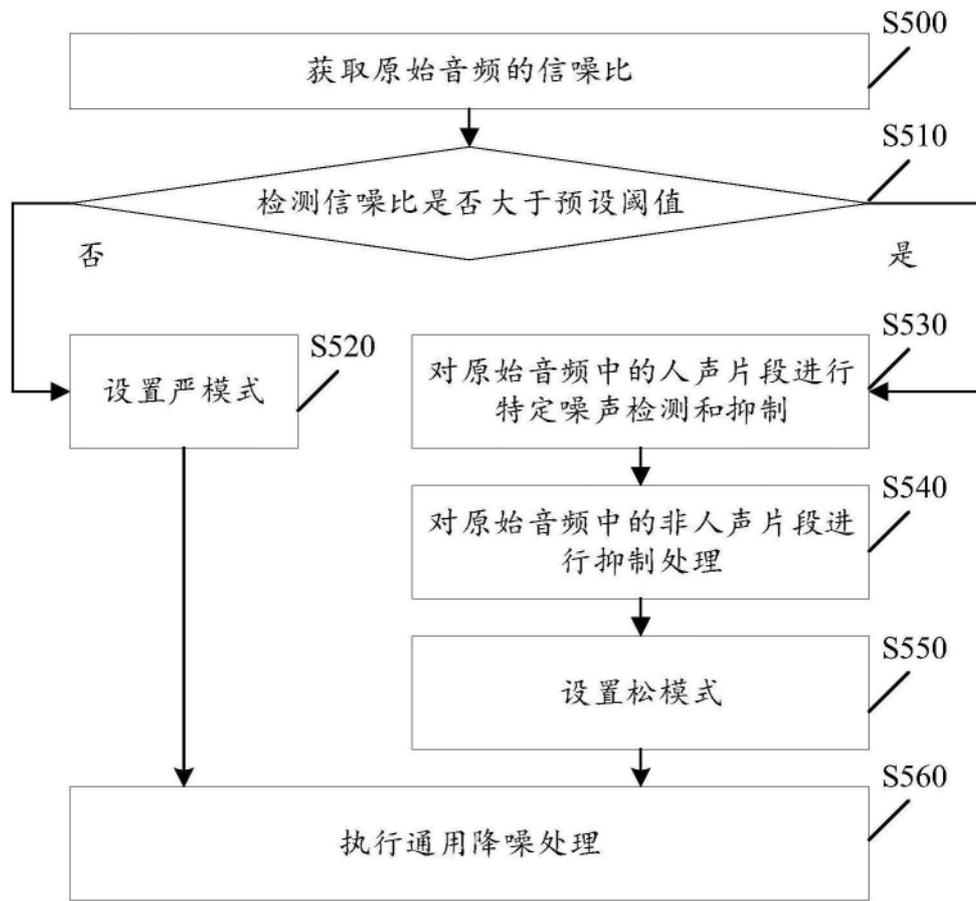


图5

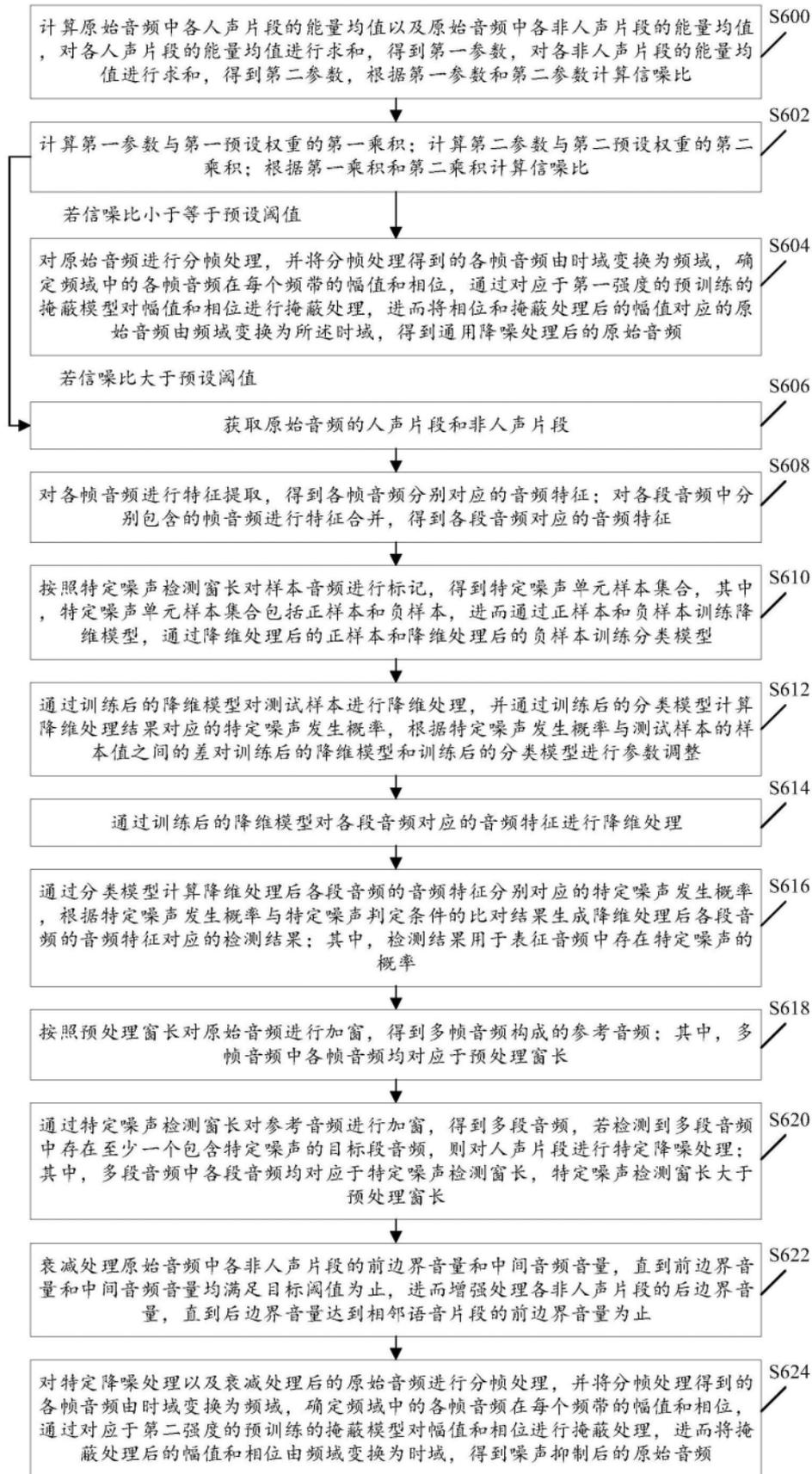


图6

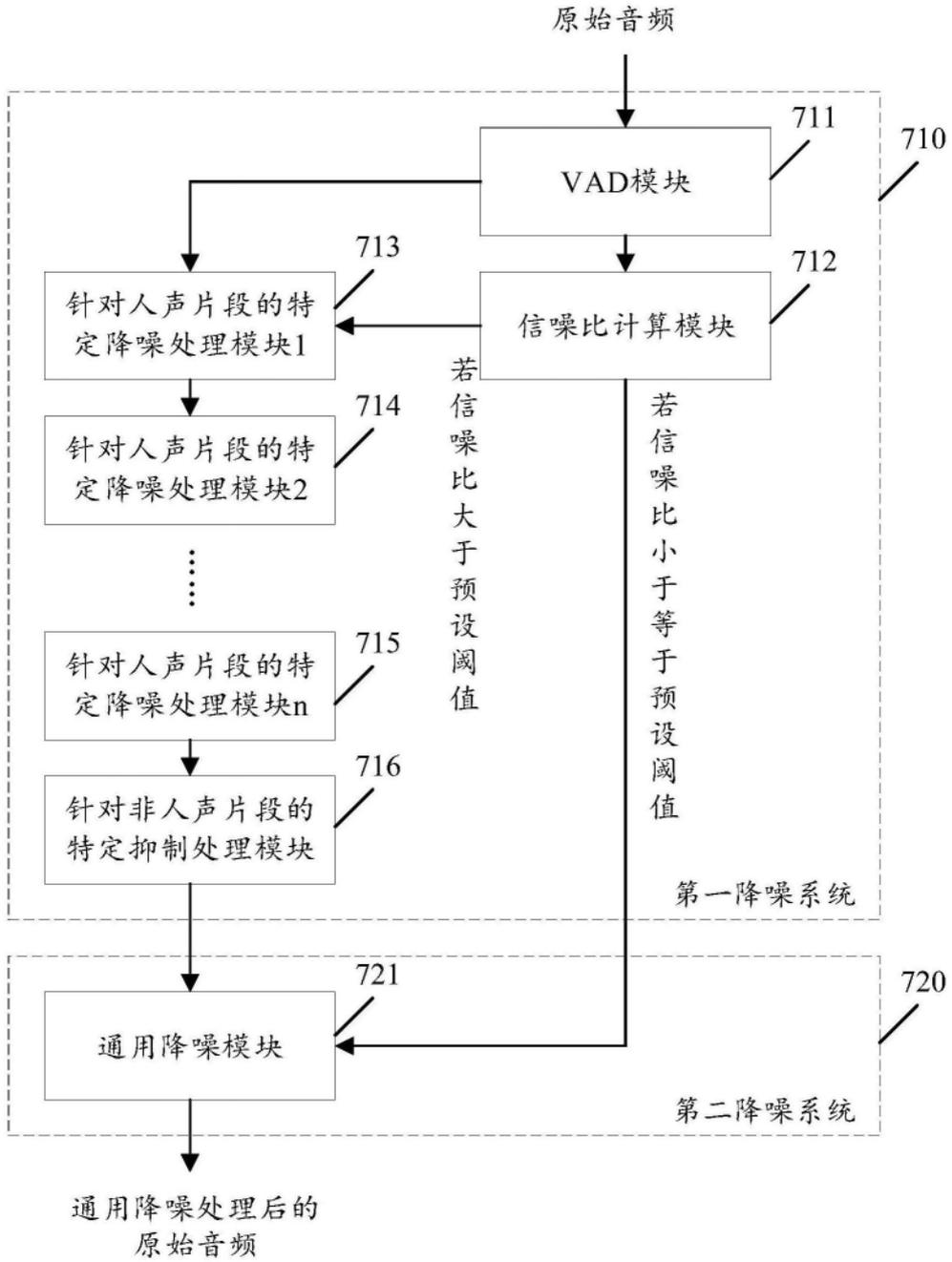


图7

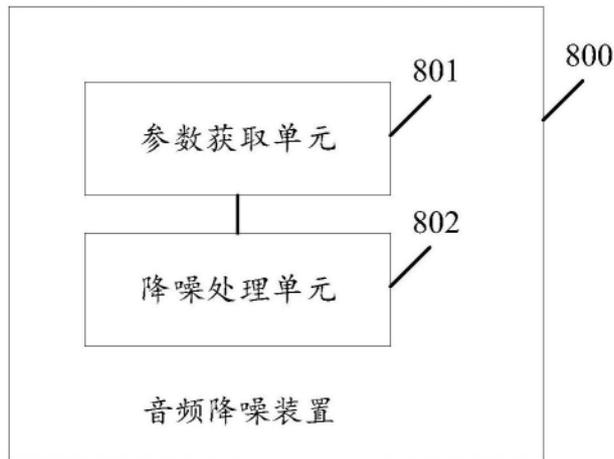


图8

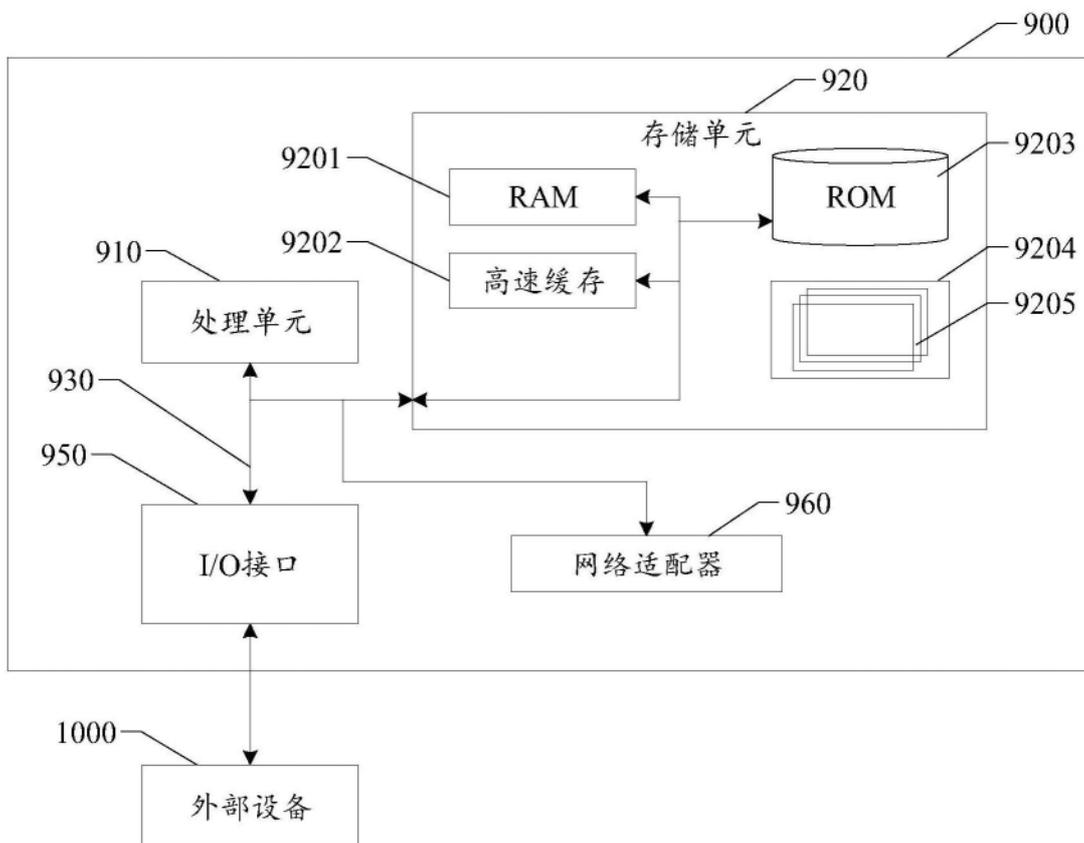


图9