



(12)发明专利

(10)授权公告号 CN 103650040 B

(45)授权公告日 2017.08.25

(21)申请号 201180072331.0

(22)申请日 2011.05.16

(65)同一申请的已公布的文献号
申请公布号 CN 103650040 A

(43)申请公布日 2014.03.19

(85)PCT国际申请进入国家阶段日
2014.01.15

(86)PCT国际申请的申请数据
PCT/US2011/036637 2011.05.16

(87)PCT国际申请的公布数据
W02012/158156 EN 2012.11.22

(73)专利权人 谷歌公司
地址 美国加利福尼亚州

(72)发明人 M·帕尼科尼

(74)专利代理机构 北京市柳沈律师事务所
11105

代理人 邵亚丽

(51)Int.Cl.
G10L 21/02(2013.01)
G10L 25/78(2013.01)
G10L 21/0208(2013.01)

(56)对比文件
EP 2058797 A1,2009.05.13,
EP 1662481 A2,2006.05.31,
CN 1770264 A,2006.05.10,

审查员 张洁

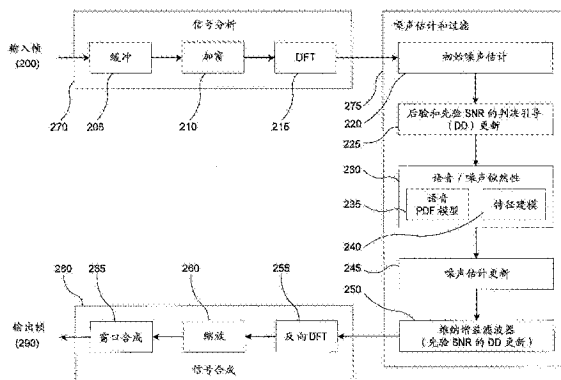
权利要求书2页 说明书18页 附图5页

(54)发明名称

使用多特征建模分析语音/噪声可能性的噪声抑制方法和装置

(57)摘要

噪声抑制系统和方法是基于对噪声频谱的估计而设计的,并采用维纳(Wiener)滤波器来抑制估计得出的噪声。噪声频谱可以依靠一种方法进行估计,即使用语音/噪声似然(如概率)函数,将所接收信号的每次/帧和频率分量分类为语音或噪声。语音/噪声似然函数经过更新和改良,将多个语音/噪声分类特征合并到一个模型中形成基于特征的概率函数,可针对每个输入帧和频率进行计算。



1. 一种通过噪声抑制组件估计和过滤噪声的方法,该方法包括以下步骤:
噪声抑制组件针对其接收到的连续多帧输入信号的每一帧,以对该帧的初始噪声估计为基础,定义语音概率函数;
测量多帧中每一帧的多元信号分类特征;
使用测量出的每帧信号分类特征,计算多帧中每一帧的基于特征的语音概率;
对计算出的每帧基于特征的语音概率应用一个或多个动态加权因子,其中对测量出的每帧信号分类特征应用一个或多个动态加权因子;
根据应用一个或多个动态加权因子后计算出的每帧基于特征的语音概率,修改多帧中每一帧的语音概率函数;
使用修改后的每帧语音概率函数,更新多帧中每一帧的初始噪声估计;以及
使用每帧更新后的初始噪声估计,对多帧中的每一帧过滤噪声,
其中,初始噪声估计是以连续多帧中每一帧的分位数噪声估计为基础。
2. 根据权利要求1所述方法,其特征在于,一个或多个动态加权因子包括每帧信号分类特征的加权和阈值参数。
3. 根据权利要求1所述方法,其特征在于,将一个或多个动态加权因子应用于计算出的基于特征的语音概率,具体步骤包括:
对测量出的每帧信号分类特征应用一个或多个动态加权因子;以及
针对应用一个或多个动态加权因子的帧,更新其基于特征的语音概率。
4. 根据权利要求3所述方法,其特征在于,对测量出的每帧信号分类特征应用一个或多个动态加权因子的步骤,包括将一个或多个动态加权因子与测量出的信号分类特征相结合,形成基于特征的语音概率函数。
5. 根据权利要求4所述方法,该方法进一步包括:
更新多帧中每一帧的基于特征的语音概率函数;以及
根据更新后的基于特征的语音概率函数,更新多帧中每一帧的语音概率函数。
6. 根据权利要求1所述方法,其特征在于,多元信号分类特征用于将输入信号划分为一种语音或噪声分类状态。
7. 根据权利要求5所述方法,其特征在于,使用递归均值更新基于特征的语音概率函数。
8. 根据权利要求4所述方法,其特征在于,基于特征的语音概率函数是通过使用映射函数将每帧的信号分类特征映射到一个概率值而得出的。
9. 根据权利要求8所述方法,其特征在于,映射函数是根据信号分类特征的值定义的,且包括一个或多个阈值和宽度参数。
10. 根据权利要求1所述方法,其特征在于,语音概率函数进一步以帧的似然比因子为基础。
11. 根据权利要求1所述方法,其特征在于,多元信号分类特征至少包括:随时间变化的平均似然比、频谱平坦度测量以及频谱模板差异测量。
12. 根据权利要求1所述方法,其特征在于,一个或多个动态加权因子至少将下述特征之一选为多元信号分类特征:随时间变化的似然比、频谱平坦度测量以及频谱模板差异测量。

13. 根据权利要求11所述方法,其特征在于,频谱模板差异测量以输入信号频谱与模板噪声频谱的对比为基础。

14. 根据权利要求13所述方法,其特征在于,使用更新后的语音概率函数和一组估计出的形状参数基于更新后的噪声估计来估计模板噪声频谱。

15. 根据权利要求14所述方法,其特征在于,估计出的形状参数是一个或多个位移、振幅以及标准化参数。

16. 根据权利要求1所述方法,该方法进一步包括:

为响应对多帧中的每一帧过滤噪声,根据修改后的帧语音概率函数,缩放每个帧的能量。

17. 根据权利要求2所述方法,该方法进一步包括:

为应用于每帧信号分类特征的加权和阈值参数设置初始值;以及在输入信号出现首个间隔之后,更新加权和阈值参数的初始值。

18. 根据权利要求17所述方法,其特征在于,更新加权和阈值参数的初始值的步骤包括:

在出现首次间隔时,计算每帧信号分类特征的直方图;

根据源自直方图的一个或多个数量,确定加权和阈值参数的新值;以及

在输入信号的第二次间隔时,使用加权和阈值参数的新值。

19. 根据权利要求18所述方法,其特征在于,首次和第二次间隔是按照输入信号的帧的序列出现的。

20. 根据权利要求18所述方法,该方法进一步包括:

将源自直方图的一个或多个数量与一个或多个内部参数进行比较,以确定输入信号基于特征的语音概率相应的加权和阈值参数。

使用多特征建模分析语音/噪声可能性的噪声抑制方法和装置

[0001] 发明所属领域

[0002] 本发明大体上与音频信号(如语音通信)传输系统和方法有关。具体而言,本发明的各个方面涉及使用语音概率建模估计和过滤噪声。

[0003] 背景

[0004] 在语音通信中,周边和/或背景噪声过多会干扰通信单方或双方对通信内容的理解,有时甚至导致对话没有意义。周边噪声包括各种外界来源的声音,其中较为常见的噪声源包括计算机、风扇、麦克风和办公设备。

[0005] 发明概述

[0006] 本发明概述以简要形式介绍了一些概念,让读者对本发明的相关方面有个基本了解。本发明概述不是对本发明的广泛性概括,也并非介绍本发明的关键或重要组成部分,亦非描述本发明的范围。本发明概述仅阐述本发明的一些概念,作为下文“详细说明”部分的铺垫。

[0007] 本发明的一个实施例涉及通过噪声抑制组件估计和过滤噪声的方法。该方法包括:针对噪声抑制组件接收到的连续多帧输入信号的每一帧,以对该帧的初始噪声估计为基础,定义语音概率函数;测量多帧中每一帧的多元信号分类特征;使用测量出的每帧信号分类特征,计算多帧中每一帧的基于特征的语音概率;对计算出的多帧中每一帧的基于特征的语音概率应用一个或多个动态加权因子;根据计算出的每帧基于特征的语音概率,修改多帧中每一帧的语音概率函数;以及使用修改后的每帧语音概率函数,更新多帧中每一帧的初始噪声估计。

[0008] 在本发明的另一实施例中,估计和过滤噪声的方法进一步包括:使用每帧更新后的初始噪声估计,对多帧中的每一帧过滤噪声。

[0009] 在本发明的另一实施例中,一个或多个动态加权因子包括每帧信号分类特征的加权和阈值参数。

[0010] 在本发明的另一实施例中,初始噪声估计是以连续多帧中每一帧的分位数噪声估计为基础。

[0011] 在本发明的另一实施例中,估计和过滤噪声的方法进一步包括:对测量出的每帧信号分类特征应用一个或多个动态加权因子;以及针对应用一个或多个动态加权因子的帧,更新其基于特征的语音概率。

[0012] 在本发明的另一实施例中,估计和过滤噪声的方法进一步包括:将一个或多个动态加权因子与测量出的信号分类特征相结合,形成基于特征的语音概率函数。

[0013] 在本发明另一实施例中,估计和过滤噪声的方法进一步包括:更新多帧中每一帧的基于特征的语音概率函数;以及根据更新后的基于特征的语音概率函数,更新多帧中每一帧的语音概率函数。

[0014] 在本发明的另一实施例中,多元信号分类特征用于将输入信号划分为一种语音或噪声分类状态。

- [0015] 在本发明的另一实施例中,使用递归均值更新基于特征的语音概率函数。
- [0016] 在本发明的另一实施例中,基于特征的语音概率函数是通过使用映射函数将每帧的信号分类特征映射到一个概率值而得出的。
- [0017] 在本发明的另一实施例中,映射函数是根据信号分类特征的值定义的,且包括一个或多个阈值和宽度参数。
- [0018] 在本发明的另一实施例中,语音概率函数进一步以帧的似然比因子为基础。
- [0019] 在本发明的另一实施例中,多元信号分类特征至少包括:随时间变化的平均似然比、频谱平坦度测量以及频谱模板差异测量。
- [0020] 在本发明的另一实施例中,一个或多个动态加权因子至少将下述特征之一选为多元信号分类特征:随时间变化的似然比、频谱平坦度测量以及频谱模板差异测量。
- [0021] 在本发明的另一实施例中,频谱模板差异测量以输入信号频谱与模板噪声频谱的对比为基础。
- [0022] 在本发明的另一实施例中,对模板噪声频谱的估计以更新后的噪声估计(使用更新后的语音概率函数和一组估计出的形状参数进行更新)为基础。
- [0023] 在本发明的另一实施例中,估计出的形状参数是一个或多个位移、振幅以及标准化参数。
- [0024] 在本发明的另一实施例中,估计和过滤噪声的方法进一步包括:为响应对多帧中的每一帧过滤噪声,根据修改后的帧语音概率函数,缩放每个帧的能量。
- [0025] 在本发明的另一实施例中,估计和过滤噪声的方法进一步包括:为应用于每帧信号分类特征的加权和阈值参数设置初始值;以及在输入信号出现首个间隔之后,更新加权和阈值参数的初始值。
- [0026] 在本发明的另一实施例中,估计和过滤噪声的方法进一步包括:在出现首次间隔时,计算每帧信号分类特征的直方图;根据源自直方图的一个或多个数量,确定加权和阈值参数的新值;以及在输入信号的第二次间隔时,使用加权和阈值参数的新值。
- [0027] 在本发明的另一实施例中,首次和第二次间隔是按照输入信号的帧的序列出现的。
- [0028] 而在本发明的另一实施例中,估计和过滤噪声的方法进一步包括:将源自直方图的一个或多个数量与一个或多个内部参数进行比较,以确定输入信号基于特征的语音概率相应的加权和阈值参数。
- [0029] 本发明的详细适用范围将在下文“详细说明”部分予以明确阐述。但是,需要认识到的一点是,“详细说明”和为说明本发明的优选实施例而举出的具体示例仅供阐述说明之用,所属领域技术人员通过阅读此“详细说明”,应能显而易见地理解本发明的精神和适用范围中的诸多变更和修改内容。
- [0030] 附图简述
- [0031] 通过阅读下方“详细说明”,并结合随附的权利要求和图示,所属领域技术人员就能更清晰地明白本发明的目标、特征和特性,所有权利要求和附图都是本说明书的一部分。在这些附图中:
- [0032] 图1总体描述了一个代表性实施例,该例可能执行本文介绍的一个或多个方面。
- [0033] 图2是一个方块图,根据本文介绍的一个或多个实施例,阐释了噪声抑制系统的典

型组件。

[0034] 图3是一个原理图,根据本文介绍的一个或多个实施例,阐释了缓冲和加窗流程。

[0035] 图4是一个流程图,根据本文介绍的一个或多个实施例,阐释了特征阈值和加权参数的更新流程。

[0036] 图5是一个方块图,根据本文介绍的一个或多个实施例,阐释了一个用于多径路由和处理音频输入信号的示例计算设备。

[0037] 本文中的标题仅为方便阅读而设,不会影响申请专利之发明的范围或意义。

[0038] 在这些附图中,为便于理解和方便阅读,具有相同或相似结构或功能的组成部分或操作,以相同的参考序号或首字母缩略词加以标注。下文“详细说明”部分将对附图进行详细介绍。

[0039] 详细说明

[0040] 本部分将展示本发明的诸多示例。下列描述将提供具体细节,以便读者能够全面了解这些示例。但是,相关领域技术人员应能理解,本发明在实际应用时可能不会实施所述全部细节内容。同样,相关领域技术人员也应理解,本发明也可能包括本文未详细介绍的其他许多明显特征。此外,下文详细说明中可能未显示或介绍一些众所周知的结构或功能,以避免对相关描述造成不必要的干扰和模糊。

[0041] 噪声抑制旨在消除或减少周边背景噪声,提高目标音频的清晰度,从而为听者提供更为舒适的听力环境。在本发明的某些实施例中,噪声抑制是在频域中进行的,而且会针对频域实施噪声估计和噪声过滤。在噪声水平极为不平稳的情况下,如果仅依靠本地信噪比(SNR)实施噪声抑制,通常会导致确定语音和噪声的可能性时出现错误偏差。针对每个输入帧和频率,更新和调节语音/噪声概率测度的过程包括使用多重语音/噪声分类特征(如本文涉及的“信号分类特征”或“噪声估计特征”)估算基于特征的概率,从而对帧中存在的语音/噪声做出更精确和稳定的估计,本文将对此进行更为详尽的描述。在下方描述内容中,“语音/噪声分类特征”、“信号分类特征”和“噪声估计特征”是可互换的,均指每个帧和频率中可用来(如测量)将输入信号分类成语音或噪声状态的特征。

[0042] 本发明与噪声抑制相关的方面是基于对噪声频谱的估计而设计的,并采用维纳(Wiener)滤波器来抑制估计得出的噪声。噪声频谱可以依靠一种方法进行估计,即使用语音/噪声似然(如概率)函数,将所接收信号的每次/帧和频率分量分类为语音或噪声。下文将详细介绍语音/噪声概率函数,及其在噪声频谱估计中的使用。

[0043] 在至少有一些布置中,可对噪声抑制组件进行配置,以执行本文描述的多种语音概率建模。举例来说,对于接收到的语音的每个输入帧,噪声抑制组件可能会执行以下处理:信号分析,包括缓冲、加窗和傅立叶变换;噪声估计和过滤,包括确定初始噪声估计、计算语音/噪声似然函数、根据语音/噪声似然函数更新初始噪声估计,以及使用维纳滤波器抑制估计得出的噪声;以及信号合成,包括反向傅立叶变换、缩放和窗口合成。此外,还可对噪声抑制组件做进一步配置,使其生成估计得出的语音帧,作为上述流程的输出。

[0044] 图1和下方的讨论对本发明的一个代表性实施例进行了简短、概括的描述,该实施例实现了本发明的多个方面。如图1所示,噪声抑制组件40可能位于信号传输路径的近端环境中,同时捕获设备5也位于近端环境中,而发出设备30则位于远端环境中。在一些布置中,噪声抑制组件40可能是大型音频(如声音)通信系统中的一个组件。噪声抑制组件40可能是

该大型系统中的一个独立组件,也可能是该系统独立组件(未显示)中的一个子部件。在图1显示的示例实施例中,噪声抑制组件40被布置用于接收和处理来自捕获设备5的输入内容,并输出至一个或多个其他音频处理组件(未显示)。这些其他音频处理组件可以是声学回声控制(AEC)、自动增益控制(AGC)以及/或者其他声音质量改进组件。在一些实施例中,这些其他处理组件可能在噪声抑制组件40之前接收来自捕获设备5的输入内容。

[0045] 捕获设备5可以是众多音频输入设备中的任意一个,比如用于捕获声音和生成输入信号的一个或多个麦克风。发出设备30可以是众多音频输出设备中的任意一个,包括用于输出一个或多个频道声音的一个或一组扬声器。举例来说,捕获设备5和发出设备30可以是计算机系统的内部硬件设备,也可以是有线和/或无线连接方式接入计算机系统的外围设备。在一些布置中,捕获设备5和发出设备30可以是扩音器、电话听筒等单个设备的组件。此外,捕获设备5和发出设备30中任意一个或两个可能具备模拟数字转换和/或数字模拟转换功能。

[0046] 至少在图1所示的实施例中,噪声抑制组件40包括一个用于协调多个流程和定时考虑的控制器50。噪声抑制组件40也可包括一个信号分析单元10、一个噪声估计单元15、一个维纳滤波器20,以及一个信号合成单元25。这些部件中的任意一个都能与控制器50进行通信,这样控制器50就能加快本文所述的一些流程。信号分析单元10、噪声估计单元15、维纳滤波器20以及信号合成单元25的各方面详情将在后文进一步详述。

[0047] 在本发明的某些实施例中,一个或多个其他组件、模块、单元等都可作为噪声抑制组件40中的一部分,可补充或替代图1中所示部件。1. 噪声抑制组件40中组件单元的标识名称(如信号分析单元、噪声估计单元等)实质仅为示例名称,不是为了限制本发明的范围。

[0048] 图2是一个流程图,阐释了本发明之完整噪声抑制系统和方法的一个示例实施例。图2所示的噪声抑制系统包括三个主要流程:信号分析270、噪声估计和过滤275,以及信号合成280。信号分析流程270可以包括多个预处理过程,这些过程必须针对输入帧200执行,才可在频域内实施噪声抑制。举例来说,信号分析270可能包括缓冲205、加窗210和离散傅立叶变换(DFT) 215的预处理步骤。图2所示的噪声估计和过滤流程275包括的步骤或子流程:初始噪声估计220、后验和先验SNR的判决引导(DD)更新225、语音/噪声可能性测定230,可能性测定是基于似然比(LR)因子进行的,而似然比是使用后验和先验SNR,以及语音概率密度函数(PDF)模型235(如高斯、拉普拉斯算子、伽马、超高斯等),还有根据特征建模240、噪声估计更新245并应用维纳增益滤波器250确定的概率而确定的。此外,在信号合成流程280中,需要将输入帧200转换回时域,因此还包括反向离散傅立叶变换255、缩放260和窗口合成265的步骤。信号合成流程280的结果就是输出帧290,这就是估计得出的语音帧。图2所示的上述噪声抑制系统的每个流程和子流程将在下文做更详细的描述。

[0049] 本文介绍的用于减少和消除语音信号中噪声的噪声抑制方法和系统,按照下方的模型方程进行(按照时域形式显示):

$$[0050] \quad y(t) = x(t) + N(t)$$

[0051] 其中, $x(t)$ 是纯净的语音信号, $y(t)$ 是观测到的有噪信号,而 $N(t)$ 则是噪声。至少在图2所示的多个流程和步骤的下列描述中,该模型假设(未知)语音信号受到加性噪声干扰,且有噪信号 $y(t)$ 与语音信号 $x(t)$ 不相关。在频域中,以上模型方程采用以下形式:

$$[0052] \quad Y_k(m) = X_k(m) + N_k(m)$$

[0053] 其中, k 表示频率, m 代表帧索引(即在短期窗口DFT215中使用的帧号, 详见下文)。

[0054] 信号分析

[0055] 信号分析270可能包括多个预处理步骤, 从而确保噪声抑制在频域中, 而非时域中执行。首先, 输入帧200会经过缓冲步骤205处理, 在这一步骤中, 输入帧200会进行扩展以涵括以前的数据(即上一帧的部分数据, 例如图3所示帧305的数据330; 详情见下文), 从而使缓冲区长度达到2的幂数。

[0056] 至少在某些布置中, 如图2所示的噪声抑制系统是以帧为基础运行的实时系统, 当接收到一个帧(如输入帧200)时, 会对其数据进行缓冲和分析。例如, 输入帧200的帧尺寸是10毫秒(ms)。8kHz的采样率相当于80个样本, 而16kHz的采样率就相当于160个样本。在一个或多个其他的布置中, 本文介绍的如图2所示的噪声抑制系统可能替代和/或额外支持其他的输入帧尺寸, 包括15ms、20ms和30ms。为清楚起见, 以下描述是以输入帧200、帧尺寸为10ms的情况为基础。

[0057] 在经过缓冲205之后, 输入帧200进入到加窗210和DFT215, 以便将输入帧200映射到频域中。因为DFT215经优化适用的数据长度为2的幂数, 所以至少在某些布置中, 该输入帧可用的分析缓冲区长度为128个样本和256个样本。图3是一个原理图, 显示了本文介绍的缓冲205和加窗210步骤示例。图3显示了当采样率为8kHz且仅分析一个信号帧时, 是如何对数据进行缓冲和加窗的。如示例所示, 数据305的新帧的帧尺寸是80个样本, 且该新帧被添加到尺寸为128个样本的缓冲区320中。此外, 加窗函数310在下文显示经扩展的缓冲区。

[0058] 由于分析缓冲区(如图3所示的缓冲区320)比帧的尺寸要大(如图3所示的帧305), 因此如之前的数据330所示, 连续的缓冲区之间会有重叠, 在所示示例中该重叠部分包括来自帧305的前48个样本。虽然分析缓冲区320彼此间具有相关性, 这样的重叠通常会使得降噪更加顺利, 但也会给合成带来限制。举例来说, 当添加了如帧305这样的缓冲区重叠部分, 必须对信号加窗以避免突变。

[0059] 如上所述, 分析缓冲区之间的任何重叠(如图3所示的缓冲区320)都可能需要加窗处理。至少在一个布置中, 在频域中进行噪声处理前后可加入相同的窗口。具体而言, 根据图2所示, 信号分析流程270的加窗步骤210和信号合成流程280的窗口合成步骤265可使用同一窗口。因此, 在这样的布置中, 窗口函数必须为保幂映射, 即重叠缓冲区部分的窗口的平方和必须为1, 如下所示:

$$[0060] \quad w^2(N) + w^2(M+N) = 1$$

[0061] 其中, N 是缓冲区长度, M 是帧的长度。定义 $y(n, m)$ 为内部缓冲区时间指数 n 和帧 m 中的有噪音频信号, 经过加窗的信号为:

$$[0062] \quad y_w(n, m) = w(n) y(n, m)$$

[0063] 在本发明的一些布置中, 噪声估计和抑制流程在频域中执行。在信号分析流程270的DFT步骤215中, 使用加窗数据的DFT将输入帧200转化到频域中:

$$[0064] \quad Y_k(m) = \sum_{n=0}^{N-1} y_w(n, m) e^{-j2\pi \frac{k}{N} n}$$

[0065] k 表示频率槽指数(子频带)。由于本文介绍的流程使用维纳滤波器进行噪声抑制(详情见下文), 因此在进行噪声估计时仅考虑频率响应的量级 $|Y(m)|$ 。

[0066] 噪声估计和过滤

[0067] 图2所示系统的噪声估计和过滤流程275,使用包含多个信号特征的语音概率模型将接收到的信号的每个输入帧200分类为语音或噪声。语音/噪声分类是针对每个时间/帧和频率定义的,通过一个语音/噪声概率函数实现,下文将加以详细介绍。通过语音/噪声分类,在信号暂停(出现噪声)时就能对噪声频谱的初始估计进行更彻底的更新,这样含有残余噪声的信号听起来就更流畅(如音乐噪声更少),且对非稳定噪声源的噪声频谱的测量就更精确和更稳健。如图2所示的示例系统,噪声估计和过滤流程275包括以下步骤:初始噪声估计220、后验和先验SNR的判决引导(DD)更新225、语音/噪声可能性测定230,可能性测定是基于似然比(LR)因子进行的,似然比是使用后验和先验SNR,以及语音概率密度函数(PDF)模型235(如高斯),还有根据特征建模240、噪声估计更新245并应用维纳增益滤波器250确定的概率而确定的。下文将详细介绍每个包括噪声估计和过滤流程275的步骤。

[0068] 在一个或更多布置中,初始噪声估计220是以分位数噪声估计为基础。噪声估计受分位数参数控制,该参数以 q 表示。根据初始噪声估计步骤220确定的噪声估计,仅能用作促进噪声更新/估计的后续流程的初始条件。

[0069] 用于噪声抑制处理的滤波器通常会用先验SNR和后定SNR(后验SNR)来表述。因此,在执行任何实际抑制之前,需要估计先验和后验的SNR数量。噪声估计和过滤流程275的语音/噪声可能性测定步骤230也需要先验和后验SNR数量,下文将对此加以详述。

[0070] 在一个示例中,后验SNR可能指以观测到的与噪声功率谱相关的输入功率谱为基础的瞬态SNR,定义如下:

$$[0071] \quad \sigma_k(m) = \frac{|Y_k(m)|^2}{\langle |N_k(m)|^2 \rangle}$$

[0072] 其中, $Y_k(m)$ 是输入的有噪频谱, $N_k(m)$ 是噪声频谱,所处的时间/帧是 m ,频率是 k 。在此示例中,先验SNR可能是与噪声功率谱相关的纯净(未知)信号功率谱的期望值,可表示为:

$$[0073] \quad \rho_k(m) = \frac{\langle |X_k(m)|^2 \rangle}{\langle |N_k(m)|^2 \rangle}$$

[0074] 其中, $X_k(m)$ 是未知纯净语音信号的频谱系数。上述每个后验和先验SNR中的噪声功率谱可能得自初始噪声估计步骤220中确定的初始估计噪声频谱,此频谱以分位数估计为基础。在至少一个实施例中,在表达后验和先验SNR时,可能使用量级数量代替上述计算中显示的平方量级:

$$[0075] \quad \sigma_k(m) = \frac{|Y_k(m)|}{\langle |N_k(m)| \rangle}$$

$$[0076] \quad \rho_k(m) = \frac{\langle |X_k(m)| \rangle}{\langle |N_k(m)| \rangle}$$

[0077] 由于纯净信号是未知信号,因此对先验SNR的估计是上一帧(即通过图2所示系统处理的输入帧200之前的输入帧)经估计的先验SNR和瞬态SNR $\sigma_k(m)$ 的平均值:

$$[0078] \quad \rho_k(m) = \gamma_{dd} H(k, m-1) \frac{|Y_k(m-1)|}{\langle |N_k(m-1)| \rangle} + (1 - \gamma_{dd}) \max(\sigma_k(m) - 1, 0)$$

[0079] 其中, $H(k, m-1)$ 是用于上一处理帧的增益滤波器(如噪声估计和过滤流程275所用的维纳增益滤波器250), $|Y_k(m-1)|$ 是观测到的上一帧有噪语音的幅度谱。在上述表达式中, 第一部分为上一时帧的先验SNR, 第二部分为先验SNR的瞬态估计。至少在本例中, 可将上述表达式看成是噪声估计和过滤流程275的先验SNR225步骤的判决引导(DD)更新, 时间平滑参数为 γ_{dd} 。先验SNR是后验SNR的流畅版, 时间有一些滞后。 γ_{dd} 越大, 流畅度越高, 但时间延迟也会增加。在一个或多个布置中, 平滑参数的值是 ~ 0.98 。

[0080] 根据本发明的某些方面, 上文描述和定义的先验和后验SNR是噪声估计和过滤流程275的语音/噪声可能性测定步骤230的组成元素。至少在本示例中, 语音/噪声可能性测定包括两个因素: (1) LR(似然比)因子, 根据先验和后验SNR确定, 以及(2)以特征建模为基础的概率, 将在下文进行更详细的描述。

[0081] 在定义和获得用于语音/噪声可能性的模型中, 语音的状态被定义为 $H^{k,m} = H_1^{k,m}$, 而噪声状态则定义为 $H^{k,m} = H_0^{k,m}$ 。定义每个帧 m 和频率槽 k 的语音和噪声状态。语音/噪声状态的概率可以表示为:

$$[0082] \quad P(H^{k,m} | Y_k(m), \{F\})$$

[0083] 语音/噪声概率取决于观测到的噪声输入频谱系数 $Y_k(m)$, 以及所处理的信号的一些特征数据(如信号分类特征), 在本例中被定义为 $\{F\}$ 。上述语音/噪声可能性的表达式在本文中也称作“语音概率函数”。至少在一个布置中, 特征数据可以是有噪输入频谱、过往频谱数据、模型数据以及脱机数据等的任何函数。例如, 特征数据 $\{F\}$ 可以包括频谱平坦度测量、谐振峰值距、LPC残余、模板匹配等等。

[0084] 在下方表达式中, 语音/噪声状态对 (k, m) 的依赖性受到抑制, $H^{k,m}$ 被写作 H 以简化符号。因此, 根据贝叶斯法则(Bayes rule), 在计算语音/噪声概率时, 可能表达为:

$$[0085] \quad P(H | Y_k(m), \{F\}) \propto P(Y_k(m) | H, \{F\}) q_{k,m}(H | \{F\}) p(\{F\})$$

[0086] 其中, $p(\{F\})$ 是以信号的特征数据为基础的先验概率, 该值在下方一个或多个表达式中被设为一个常数。在本例中, 数量 $q_{k,m}(H | \{F\})$ 是特征数据 $\{F\}$ 下的语音/噪声概率, 详细描述请见下文。在描述本发明的各方面内容时, 上述数量 $q_{k,m}(H | \{F\})$ 也被称为“基于特征的语音概率”。不考虑以 $\{F\}$ 为基础的先验概率, 且为简化符号, 标记 $q_{k,m}(H_1 | \{F\}) = q$ 以及 $q_{k,m}(H_0 | \{F\}) = 1 - q$, 则标准化的语音概率可写作:

$$[0087] \quad P(H_1 | Y_k(m), \{F\}) = \frac{P(Y_k(m) | H_1, \{F\}) q}{P(Y_k(m) | H_1, \{F\}) q + P(Y_k(m) | H_0, \{F\}) (1 - q)}$$

$$[0088] \quad P(H_1 | Y_k(m), \{F\}) = \frac{q \Delta_k}{q \Delta_k + 1 - q}$$

$$[0089] \quad \text{其中似然比(LR) } \Delta_k \text{ 为: } \Delta_k = \frac{P(Y_k(m) | H_1, \{F\})}{P(Y_k(m) | H_0, \{F\})}$$

[0090] 在上述 Δ_k 的表达式中, 至少在本文所述的一个模型布置中, 数量 $P(Y_k(m) | H_{1,0}, \{F\})$ 是通过线性状态模型和针对语音和噪声频谱系数的高斯概率密度函数(PDF)假设来确

定的。更具体而言,有噪输入信号的线性模型可表达为:语音状态下 $Y_k(m) = X_k(m) + N_k(m)$,其中 $H=H_1$;噪音状态下 $Y_k(m) = N_k(m)$,其中 $H=H_0$ 。假设高斯PDF使用复杂系数 $\{X_k, N_k\}$,则数量 $P(Y_k(m) | H, \{F\})$ 表示如下:

$$[0091] \quad P(Y_k(m)|H_0, \{F\}) = P(Y_k(m)|H_0) \infty \frac{1}{\langle |N_k|^2 \rangle} \exp\left(-\frac{|Y_k|^2}{\langle |N_k|^2 \rangle}\right)$$

$$[0092] \quad P(Y_k(m)|H_1, \{F\}) = P(Y_k(m)|H_1) \infty \frac{1}{\langle |N_k|^2 \rangle + \langle |X_k|^2 \rangle} \exp\left(-\frac{|Y_k|^2}{\langle |N_k|^2 \rangle + \langle |X_k|^2 \rangle}\right)$$

[0093] 由于完全可根据线性模型和高斯PDF假设确定概率,因此可将特征依赖从上述表达式中删除。这样,似然比 Δ_k 就变成:

$$[0094] \quad \Delta_k = \frac{P(Y_k(m)|H_1)}{P(Y_k(m)|H_0)} = \frac{\exp\left(\frac{\rho_k(m)\sigma_k(m)}{(1+\rho_k(m))}\right)}{(1+\rho_k(m))}$$

[0095] 其中, $\rho_k(m)$ 是未知信号的SNR(即先验SNR), $\sigma_k(m)$ 是频率 k 和帧 m 的后定信号SNR(即后验SNR或瞬态SNR)。在一个实现例子中,上述表达式中使用的先验SNR和后验SNR由量级定义进行估计,算式为:

$$[0096] \quad \sigma_k(m) = \frac{|Y_k(m)|}{\langle |N_k(m)| \rangle}$$

$$[0097] \quad \rho_k(m) = \frac{\langle |X_k(m)| \rangle}{\langle |N_k(m)| \rangle}$$

[0098] 根据上述表达式和描述,至少在一个布置中,语音/噪声状态概率可通过似然比(Δ_k)和数量 $q_{k,m}(H_1 | \{F\}) = q$ 获得,其中,似然比根据频变后验和先验SNR确定,数量是基于特征或基于模型的概率,详细描述参见下文。因此,语音/噪声状态概率可表示为:

$$[0099] \quad P(H_1|Y_k(m), \{F\}) = \frac{q\Delta_k}{q\Delta_k + 1 - q}$$

$$[0100] \quad P(H_0|Y_k(m), \{F\}) = 1 - P(H_1|Y_k(m), \{F\})$$

[0101] 因为有时帧到帧之间的频变似然比因子(Δ_k)会有很大波动,所以至少在本文描述的一个噪声抑制系统的布置中,会使用经过时间平滑处理的似然比因子:

$$[0102] \quad \log(\tilde{\Delta}_k(m)) = \gamma_{lnt} \log(\tilde{\Delta}_k(m-1)) + (1 - \gamma_{lnt}) \log(\Delta_k(m))$$

[0103] 此外,经过时间平滑处理的似然比因子的几何平均数(包括所有频率)可用作对基于帧的语音/噪声分类的可靠测量结果:

$$[0104] \quad \log\left(\prod_k \tilde{\Delta}_k(m)\right)^{1/N} = \frac{1}{N} \sum_{k=1}^N \log(\tilde{\Delta}_k(m))$$

[0105] 如上所述,可在语音/噪声可能性确定步骤230使用高斯假设作为语音PDF模型235,从而获得似然比。在一个或多个其他布置中,其他语音PDF模型也可用作测量似然比的

基础,包括拉普拉斯算子、伽马和/或超高斯。举例来说,当高斯假设可合理表示噪声时,该假设并不一定适用于语音,尤其是在较短的时帧中(如~10ms)。在这种情况下,可以使用另一种语音PDF模型,但这很可能会增加复杂性。

[0106] 如图2所示,要在噪声估计和过滤流程275中确定语音/噪声可能性(或概率)230,这不仅需要本地SNR(即先验SNR和瞬态SNR)的引导,还要结合从特征建模240中获得的语音模型/认知内容。将语音模型/认知内容并入到语音/噪声概率确定中,能让本文所述的噪声抑制流程更好地处理和/或区分极不稳定的噪声水平,如果仅依靠本地SNR,可能会造成可能性偏差。至少在一个布置中,系统使用了一个流程,对包含本地SNR和语音特征/模型数据的每个帧和频率更新和适应基于特征的概率 $q_{k,m}(H|F)$ 。在下方所述的此更新和适应流程的各方面内容中,符号 $q_{k,m}(H|F)=q_{k,m}$ 。因为此处所述流程仅以帧为基础对数量 $q_{k,m}(H|F)$ 建模和更新,所以变量k就受到了抑制。

[0107] 根据本发明的一个或多方面内容,对基于特征的概率的更新可采用以下模型:

$$[0108] \quad q_m = \gamma q_{m-1} + (1 - \gamma) M(z, w)$$

[0109] 其中, γ 是一个平滑常数, $M(z)$ 是给定时间和频率的映射函数(如在0和1之间)。此映射函数中的变量 z 是 $z=F-T$,其中 F 是被测特征, T 是阈值。参数 w 则代表映射函数的形状/宽度特征。映射函数根据测量出的特征以及阈值和宽度参数,将时频槽划分为语音(M 接近1)或噪声(M 接近0)。

[0110] 在一个布置中,噪声估计和过滤流程275在实施特征建模240以确定语音/噪声可能性230时,会考虑语音信号的以下特征:(1) LRT均值,可以基于本地SNR得出,(2) 频谱平坦度,可基于语音谐波模型得出,以及(3) 频谱模板差异测量。下文将对这三个特征做更详细的描述。需要认识到的一点是,除下文所述的三个示例特征外,还可使用很多其他语音信号特征来作为补充或替代特征。

[0111] 1. LRT均值特征

[0112] 如上所述,经过时间平滑处理的似然比(LR)因子的几何平均数是语音/噪声状态的可靠指标:

$$[0113] \quad F_1 = \log \left(\prod_k \tilde{\Delta}_k(m) \right)^{1/N} = \frac{1}{N} \sum_{k=1}^N \log(\tilde{\Delta}_k(m))$$

[0114] 其中经过时间处理的LR因子根据前文所述表达式得出。使用LRT均值特征时,映射函数 $M(z)$ 的一个示例可能是“S”型曲线函数,例如:

$$[0115] \quad M(z) = 0.5 * (\tanh(w_1 z) + 0.5)$$

$$[0116] \quad z = T_1 - F_1$$

[0117] 其中, F_1 是特征, w_1 是一个过渡/宽度参数,用于控制从0到1的映射的平滑性。阈值参数 T_1 需要根据参数设置来确定,本文将对此做更详细的描述。

[0118] 2. 频谱平坦度特征

[0119] 为获得频谱平坦度特征,假设语音比噪声有更多的谐波行为。然而,语音频谱往往会在基频(基音)和谐波中出现峰值,而噪声频谱则相对平坦。因此,至少在某些布置中,本地频谱平坦度测量的总和可用作指示/区分语音和噪声的良好判断依据。

[0120] 在计算频谱平坦度时, N 代表频率槽的数量, B 代表频率带的数量。 k 是频率槽指数, j 是频率带指数。每个频率带将包括大量的频率槽。举例来说,128槽的频率频谱可分成4个

频率带(低带、中低带、中高带和高带),每个频率带包括32个槽。在另一个示例中,仅使用一个包括所有频率的频率带。频谱平坦度可以通过计算输入幅度谱的几何平均数与算术平均数的比值得出:

$$[0121] \quad F_2 = \frac{(\prod_k |Y_k|)^{1/N}}{\frac{1}{N} \sum_k |Y_k|}$$

[0122] 其中N表示频率带中的频率数。对于噪声,计算出的数量 F_2 偏大且为常数,而对于语音,计算出的数量则偏小且为变量。同样,用于对基于特征的先验概率进行更新的映射函数 $M(z)$ 的一个示例可表示为“S”型曲线函数:

$$[0123] \quad M(z) = 0.5 * (\tanh(w_2 z_2) + 0.5)$$

$$[0124] \quad z = T_2 - F_2$$

[0125] 3. 频谱模板差异特征

[0126] 除了上述用于频谱平坦度特征的噪声相关假设之外,有关噪声频谱的另一个假设是,噪声频谱比语音频谱更稳定。因此,可假设噪声频谱的整体形状在任何给定节段都倾向于保持相同。根据这种假设,可继续在本例的语音/噪声概率确定中融入第三个特征。该附加特征可测量输入频谱与噪声频谱形状的偏差。

[0127] 此第三特征可通过对比输入频谱与作为学习模板的噪声频谱来确定。至少在某些布置中,模板频谱通过更新频谱(最初被设为零)中极有可能是噪声或语音停顿的区段来确定。该比较结果是对噪声的保守估计,其中仅对语音概率确定低于阈值(如 $P(H|F)$, $\{F\} < \lambda$)的区段处更新了噪声。在其他布置中,模板频谱也可能被导入到算法中,或从对应不同噪声的形状表中筛选出来。考虑到输入频谱 $Y_k(m)$ 和模板频谱(可表示为 $\alpha_k(m)$),如想获得频谱模板差异特征,可首先将频谱差异测量定义为:

$$[0128] \quad J = \sum_k |Y_k(m) - (\alpha \alpha_k(m) + u)|^2$$

[0129] 其中, (α, u) 是形状参数,包括线性位移和振幅参数,是通过将J最小化获得的。参数 (α, u) 通过线性方程获得,因此可对每个帧轻松抽取此参数。在某些示例中,这些参数可表明输入频谱(在音量增加的情况下)的任何简单位移/标度变化。之后该特征将成为标准化的测度,

$$[0130] \quad F_3 = \frac{J}{Norm}$$

[0131] 其中标准化是所有频率以及之前时帧在某些时间窗口的平均输入频谱:

$$[0132] \quad Norm = \frac{1}{W} \sum_{n=0}^W \sum_k |Y_k(n)|^2$$

[0133] 如上所述,频谱模板差异特征可测量出模版或习得噪声频谱与输入频谱的差异/偏差。至少在某些布置中,这种频谱模板差异特征可用于修正基于特征的语音/噪声概率 $q_{k,m}(H|F)$ 。如果 F_3 较小,则可将输入帧频谱视作“接近”模板频谱,且很可能将该输入帧视作噪声。另一方面,如果频谱模板差异特征值较大,则表示输入帧(如输入帧200)频谱与噪声模板频谱有很大差异,就可判断该帧为语音。在一个或多个变化情况中,模板频谱可以导入至语音/噪声概率算法,或作为在线资源用于数字化测量和使用。

[0134] 与LRT均值特征和频谱平坦度特征相似,可使用上述相同的“S”型曲线函数,将频谱模板差异特征值映射为概率权数。需要重点强调的是,频谱模板差异特征测量比频谱平坦度特征测量更普遍。如果一个模板具备恒定(即接近完美)的平坦频谱,则频谱模板差异特征可简化为对频谱平坦度的测量。

[0135] 至少在一个布置中,可以在频谱模板差异测量中加入加权期限 W_k ,以突出频谱中的特定频率带:

$$[0136] \quad J = \sum_k W_k |Y_k(m) - (\alpha \alpha_k(m) + \mu)|^2$$

[0137] 在本示例中,所有频率的加权期限可保持为 $W_k=1$ 。

[0138] 上文所述的多个特征(即LRT均值、频谱平坦度和频谱模板差异)可在语音/噪声概率的更新模板中同时出现,如下所示:

[0139] $q_m(H|F_1, F_2, F_3) = q_m = \gamma_p q_{m-1} + (1 - \gamma_p) [\tau_1 M(F_1 - T_1) + \tau_2 M(F_2 - T_2) + \tau_3 M(F_3 - T_3)]$ 不同的特征源自不同的信号(即不同的特征传达不同的信息,例如第一特征传达的是电能测量或本地SNR,第二特征传达的是噪声频谱平坦度,第三特征传达的是噪声稳定性和一般形状),这些特征互相补充,以提供一个更稳定、更具适应性的语音/噪声概率更新。上文所示的语音/噪声概率的更新模型中,包括各种加权期限(τ_i)、阈值参数 $\{T_i\}$,以及用于映射函数的宽度参数。举例来说,如果给定输入的频谱平坦度特征(F_2)不可靠,例如噪声频谱不是十分平坦,则第二个加权期限 τ_2 可能被设置为零,即 $\tau_2=0$,从而避免对更新模型的测量出现不可靠结果。对于这些加权期限和阈值参数的设置将在下文做更详细的描述。

[0140] 图2显示了系统在噪声估计和过滤流程275中确定语音/噪声可能性230的过程,可能性确定后,将执行噪声估计更新245(如软决策递归噪声更新)。举例来说,噪声估计更新245可表示如下:

$$[0141] \quad |\hat{N}_k(m)| = \gamma_n |\hat{N}_k(m-1)| + (1 - \gamma_n) [P(H_1 | Y_k(m)) |\hat{N}_k(m-1)| + P(H_o | Y_k(m)) |Y_k(m)|]$$

[0142] 其中 $|\hat{N}_k(m)|$ 是帧/时间为 m 、频率槽为 k 时对噪声频谱量级的估计。参数 γ_n 控制噪声更新的平滑度,第二个期限则使用输入频谱和上次噪声估计对噪声进行更新,然后根据如上所述的语音/噪声概率进行加权,这可表示为:

$$[0143] \quad P(H_1 | Y_k(m)) = \frac{q_m \Delta_k(m)}{q_m \Delta_k(m) + 1 - q_m}$$

[0144] 其中LR因子 $\Delta_k(m)$ 是:

$$[0145] \quad \Delta_k(m) = \frac{\exp\left(\frac{\rho(m)\sigma_k(m)}{(1 + \rho_k(m))}\right)}{(1 + \rho_k(m))}$$

[0146] 数量 q_m 是基于模型或基于特征的语音/噪声概率,得自上述具有多个特征的更新模型。上述噪声估计模型会对噪声可能性较大(即语音可能性较小)的每个帧和频率槽的噪声进行更新。对于噪声可能性不大的帧和频率槽,则将对信号中上一个帧的估计作为噪声估计。

[0147] 至少在一个布置中,噪声估计更新流程受到语音/噪声可能性和平滑度参数 γ_n 的

控制,平滑度参数可被设为像0.85这样的值。在不同的示例中,对于语音概率超过阈值参数 λ 的区域,平滑度参数可能会增加到 $\gamma_n \approx 0.99$,以防止语音开始处的噪声水平增加过高。在一个或多个布置中,阈值参数被设定为 $\lambda=0.2/0.25$,后文将对此加以详细描述。

[0148] 完成噪声估计更新245后,噪声估计和过滤流程275会采用维纳增益滤波器250,以减少或消除来自输入帧200的估计噪声量。标准维纳滤波器表达如下:

$$[0149] \quad H_w(k, m) = \frac{\langle |X_k(m)|^2 \rangle}{\langle |Y_k(m)|^2 \rangle} = 1 - \frac{\langle |N_k(m)|^2 \rangle}{\langle |Y_k(m)|^2 \rangle} \approx 1 - \frac{|\hat{N}_k(m)|^2}{|Y_k(m)|^2}$$

[0150] 其中, $\hat{N}_k(m)$ 是估计得出的噪声频谱系数, $Y_k(m)$ 是观测到的有噪频谱系数, $X_k(m)$ 是纯净语音频谱(帧为 m ,频率为 k)。之后,平方量级可被量级代替,维纳滤波器就变成:

$$[0151] \quad H_k(k, m) = 1 - \frac{|\hat{N}_k(m)|}{|Y_k(m)|}$$

[0152] 在一种或多种常规方法中,会对滤波器直接应用时间平均法,以减少任何的帧间波动。根据本发明的某些方面,维纳滤波器用先验SNR表示,而判决引导(DD)更新则用于对先验SNR进行时间平均计算。维纳滤波器可用先验SNR表示为:

$$[0153] \quad H_k(k, m) = \frac{\rho_k(m)}{1 + \rho_k(m)}$$

[0154] 其中, $\rho_k(m)$ 代表上文定义的先验SNR,将噪声频谱替换为估计得出的噪声频谱:

$$[0155] \quad \rho_k(m) = \frac{\langle |X_k(m)| \rangle}{|N_k(m)|}$$

[0156] 如上所述,按照DD更新估计先验SNR。该增益滤波器通过取底和过相减参数,可得出:

$$[0157] \quad H_{w,dd}(k, m) = \text{最大值} \left(H_{\text{最小值}}, \frac{\rho_k(m)}{\beta + \rho_k(m)} \right)$$

[0158] 在本布置和其他布置中,因为DD更新明确对先验SNR进行时间平均计算,所以不会对该增益滤波器再进行外部时间平均计算。参数 β 是根据噪声抑制系统中实施的噪声抑制器(即图1所示的噪声抑制组件15)的主动配置(如模式)定义的。

[0159] 维纳滤波器应用到输入量级频谱中,以获得经抑制的信号(如对基础语音信号的估计)。在噪声估计和过滤流程275中采用维纳滤波器250会得出:

$$[0160] \quad \hat{X}_k(m) = H_{w,dd}(k, m) Y_k(m)$$

[0161] 信号合成

[0162] 信号合成280包括各种后验噪声抑制处理,以生成包括纯净语音的输出帧290。在应用维纳滤波器后,使用反向DFT255将帧转换回时域。在一个或多个布置中,转换回时域可表达为:

$$[0163] \quad \hat{x}(n, m) = \text{Re} \left[\frac{1}{N} \sum_k^{N-1} \hat{X}_k(m) e^{j2\pi \frac{n}{N} k} \right]$$

[0164] 其中, $\hat{X}_k(m)$ 是经维纳滤波器抑制后估计得出的语音, $\hat{x}(n, m)$ 是相应的时域信号, 其中时间索引为 n , 帧索引为 m 。

[0165] 在反向DFT255之后, 作为信号合成流程280的一部分, 对经噪声抑制的信号实施能量缩放260。能量缩放可用于帮助重建语音帧, 且重建方式可增加经抑制后的语音的能量。例如, 实施缩放时应确保只有语音帧会放大到一定程度, 而噪声帧保持不变。由于噪声抑制可能降低语音信号水平, 因此在缩放260过程中对语音区段适当放大是有益处的。在一个布置中, 根据语音帧在噪声估计和过滤流程275中的能量损失, 对该帧实施缩放260。增益情况可通过该语音帧在噪声抑制处理前后的能量比来确定:

$$[0166] \quad K = \sqrt{\frac{\text{能量}_{\text{处理后}}}{\text{能量}_{\text{处理前}}}}$$

[0167] 在当前示例中, 可根据下方模型提取标度:

$$[0168] \quad \text{标度} = A(K)\tilde{P}(H_1 | m) + B(K)(1 - \tilde{P}(H_1 | m))$$

[0169] 其中, $\tilde{P}(H_1 | m)$ 是帧 m 的语音概率, 通过取所有频率的语音概率函数 $P(H_1 | Y_k(m), \{F\})$ 的平均值而得:

$$[0170] \quad \tilde{P}(H_1 | m) = \sum_k P(H_1 | Y_k(m), \{F\})$$

[0171] 在上述标度方程中, 如果概率 $\tilde{P}(H_1 | m)$ 接近1 (即该帧可能是语音), 则第一项将较大; 如果该帧可能是噪声, 则第二项将较大。

[0172] 在上述标度方程中, 参数 $A(K)$, $B(K)$ 控制输入帧 (如输入帧200) 的缩放。举例来说, 在一个布置中, $A(K)$, $B(K)$ 可能按如下公式控制缩放: 如果 $K > 0.5$, $A(K) = 1.0 + 1.3 * (K - 0.5)$, 最大值按 $1/K$ 取得。如果 $K < 0.5$, 则 $A(K) = 1.0$ 。参数 $B(K) = 1.0$, 所以该帧不会针对噪声区域进行缩放。这些区域的标度可由维纳滤波器中的取底项确定。

[0173] 信号合成280也包括窗口合成操作265, 该操作提供估计得出的语音的最终输出帧290。在一个示例中, 窗口合成265为:

$$[0174] \quad \text{标度}(m-1)w(n)\hat{x}(n+M, m-1) + \text{标度}(m)w(n)\hat{x}(n, m) \dots \dots \dots 0 \leq n < M$$

[0175] 其中, 标度参数由每个帧的上述标度方程式得出。

[0176] 参数估计

[0177] 基于特征的语音/噪声概率函数的更新模型 (公式如下) 包括应用到特征测量的多个特征加权 (τ_i) 和阈值 $\{T_i\}$ 参数:

$$[0178] \quad q_m(H | F_1, F_2, F_3) = q_m = \gamma_p q_{m-1} + (1 - \gamma_p) [\tau_1 M (F_1 - T_1) + \tau_2 M (F_2 - T_2) + \tau_3 M (F_3 - T_3)]$$

[0179] 这些加权 (τ_i) 和阈值 $\{T_i\}$ 参数用于防止不可靠的特征测量进入更新模型。映射函数也包括宽度参数 $\{w_i\}$ 以控制映射函数的形状:

$$[0180] \quad M = M(F_i - T_i; w_i)$$

[0181] 例如, 如果给定输入的LRT均值特征 (F_i) 不可靠, 例如如果初始噪声评估中出现错误, 则首个加权参数 τ_1 可被设置为零, 即 $\tau_1 = 0$, 从而避免将不可靠的LRT均值测量带入到更新模型中。

[0182] 至少在一个实施例中,特征加权和阈值参数的初始设置为,仅使用LRT均值特征(F_1),因此 $\tau_1=\tau_3=0$,且特征的初始阈值为 $T_1=0.5$ 。表1列出了根据本发明多个实施例得出的示例参数设置。表1标明了每个参数,并为每个参数提供了简短描述和示例默认值。需要认识到的一点是,除了表1中列出的这些参数,还可以使用多个其他参数设置和/或默认值作为补充或替代参数。在表1中,每个特征对应的映射函数的宽度参数都设置为相同的值,即 $w=4$ 。

[0183] 表1

示例参数设置			
参数	描述	默认值	
q	分位数: 初始噪声估计	0.25	
γ_{dd}	DD 更新的平滑度	0.98	
γ_{lrt}	似然比因子的平滑度	0.9	
T_1	似然比特征的阈值	初始值: 0.5; 实时修改	
[0184] T_2	频谱平坦度的阈值	实时确定	
T_3	频谱模板的阈值	实时确定	
$\tau_i (i=1,2,3)$	特征的加权重	(1,0,0); 实时更新	
w	先验的“S”型映射的宽度	4.0	
γ_p	先验概率的更新参数	0.9	
γ_n	噪声估计的更新参数	0.9	
λ	噪声状态的阈值	0.2	
H_{min}	增益滤波器中的取底	根据模式 # 确定	

项			
		增益滤波器中的过相	
	β	减	根据模式 # 确定
[0185]	融合	启动时刻	200 帧
	区块大小	最大帧尺寸	320
	分析大小	最大分析区块	512

[0186] 在一个或多个实施例中,用于特征测量的特征阈值和加权参数(例如,语音/噪声概率的更新模型中列出的 T_1 、 T_2 、 T_3 和 τ_1 、 τ_2 、 τ_3 ,这些参数也包含在上文的表1中)将在设置间隔后动态更新。在一个示例中,特征阈值和加权参数可能针对每个窗口 W 进行更新,其中 $W=500$ 帧。在其他示例中,可能使用替代更新间隔,包括多个帧数或时间设置间隔。在本发明的这些和其他实施例中,如图4所示,可能针对特征测量执行特征阈值和加权参数的更新流程。4.

[0187] 图4阐述了针对特征测量(如LRT均值特征(F_1)、频谱平坦度特征(F_2)和频谱模板差异特征(F_3))更新特征阈值和加权参数的示例流程。该流程从步骤400开始,用于语音序列首个 W 帧(即500帧)的特征阈值和加权参数(如 T_1 、 T_2 、 T_3 和 τ_1 、 τ_2 、 τ_3)被设置为初始值。举例来说,阈值和加权参数的初始值可能为 $\{T_1=0.5\}$ 和 $\{\tau_1=1.0, \tau_2=0, \tau_3=0\}$ 。

[0188] 在步骤405中,可能会计算相关(如当前或目前)参数估计窗口的 W 帧特征,并绘制直方图。对于语音序列的初始窗口,步骤405包括该序列的首个 W 帧,此节段的阈值和加权参数被固定为在步骤400中设定的初始值。在语音序列的后续窗口(即除初始窗口外的其他序列窗口)中,阈值和加权参数固定为从之前的 W 帧所得的值。

[0189] 流程继续进行到步骤410,在处理完 W 帧之后,从步骤405计算出的直方图的数量中,抽取用于特征的新阈值和加权参数。在一个示例中,特征的阈值和加权参数来自一些直方图数量,包括直方图峰位、直方图高度、每个特征各自直方图的一定范围内的平均值,以及每个特征各自直方图的一定范围的波动。除了上述数量外,还可从步骤405中计算的直方图中抽取许多其他数量,作为补充或替代数值,用于在步骤410中提取新的特征阈值和加权参数。

[0190] 至少在一个布置中,从步骤410的直方图中抽取的数量与一些内部参数作比较,以确定相应的先验模型阈值和加权参数。此等内部参数的示例可能包括以下集合:(1)标度参数,应用于测量得出的直方图中的支配峰值或两个峰值的总和,以获得特征阈值;(2)用于将两个过于接近的直方图进行合并的参数;(3)在峰值平均高度过小的情况下,用于拒绝特征的参数;(4)在平均峰位过小的情况下,用于拒绝特征的参数;(5)在直方图范围内的LRT特征波动过低的情况下,用于拒绝一些特征的参数;以及(6)每个特征的阈值的最大和最小限度。除了上述示例参数,还可使用许多其他参数作为内部参数,与步骤410中抽取的数量作比较。

[0191] 在步骤415中,将从步骤410提取的阈值和加权参数固定或设置为语音序列下一个

W帧的特征阈值和加权参数。如果在步骤420到达语音序列的末端,则该流程结束。但是,如果在步骤420未到达语音序列的末端,则该流程将返回步骤405,使用序列的下一个W帧重复步骤405到步骤420,并在步骤415固定阈值和加权参数。

[0192] 在本发明的一些实施例中,如图4所示,在步骤400中设置的初始特征阈值和加权参数,将会用在整个语音序列中,而无需更新这些参数的值。在其他实施例中,在处理完序列的首个W帧窗口后,可能会更新一次阈值和加权参数(即继阈值和加权参数的初始值后,进行一次更新)。

[0193] 在本发明的另一些实施例中,图4所示的特征阈值和加权参数更新流程,可能使用序列的重叠窗口,其中 W_1 包括帧1-500, W_2 包括帧250-750, W_3 包括帧500-1000等等。另一替代方法是使用非重叠窗口,其中 W_1 包括帧1-500, W_2 包括帧500-1000, W_3 包括帧1000-1500,以此类推。此外,虽然一些布置使用固定的窗口,即每个W包括500帧;其他布置则可能使用可变或变化的窗口。举例来说, W_1 可能包括500帧, W_2 包括250帧,而 W_3 包括750帧。此外,在一个或多个布置中,这些可变或变化的窗口可能重叠或非重叠,如 W_1 包括帧1-500(500帧), W_2 包括帧500-750(250帧、非重叠),以及 W_3 包括帧500-1250(750帧、重叠)。需要认识到的一点是,阈值和加权参数可以根据其他多种窗口配置进行更新,这些配置包括给定序列的多种其他特征。

[0194] 根据图4所示的更新流程,在一些情况下,从步骤410中抽取的特征阈值和加权参数可以会阻止一个或多个特征(如LRT均值特征(F_1)、频谱平坦度特征(F_2)以及/或者频谱模板差异特征(F_3))用于计算语音/噪声概率的更新模型。在这种情况下,未包括在更新模型中的每个特征的加权参数将被设置为0。

[0195] 在计算语音/噪声概率的更新模型时使用了三种特征的情况下,参数更新流程的特征阈值和加权参数抽取步骤(如图4所示步骤410)可能产生以下结果:(1)使用所有三个特征 $\{\tau_1=1/3, \tau_2=1/3, \tau_3=1/3\}$; (2)使用两个特征,如特征1和3 $\{\tau_1=1/2, \tau_2=0, \tau_3=1/2\}$;或(3)仅使用一个特征,如特征1 $\{\tau_1=1.0, \tau_2=0, \tau_3=0\}$ 。

[0196] 图5是一个方块图,阐释了一个用于多路径路由的示例计算设备500,是根据本发明的一个或多个实施例绘制而成的。在最基本的配置501中,计算设备500通常包括一个或多个处理器510和系统内存520。内存总线530可用于实现处理器510和系统内存520之间的通信。

[0197] 根据所需的配置,处理器510可以是任意类型,包括但不限于:微处理器(μP)、微控制器(μC)、数字信号处理器(DSP)或它们的任意组合。处理器510可包括一个或多个缓存级别,如一级缓存511和二级缓存512、处理器内核513及寄存器514。处理器内核513包括一个算术逻辑单元(ALU)、一个浮点单元(FPU)、一个数字信号处理核心(DSP核心),或它们的任意组合。内存控制器515还可与处理器510一同使用,在某些实施例中,内存控制器515是处理器510的内部组成部分。

[0198] 根据所需的配置,系统内存520可以是任意类型,包括但不限于:易失性存储器(如RAM)、非易失性存储器(如ROM和闪存等)或它们的任意组合。系统内存520通常包括一个操作系统521、一个或多个应用程序522及程序数据524。至少在某些实施例中,应用程序522包括一个多径处理算法523,该算法配置为将有噪输入信号传到噪声抑制组件中。多径处理算法进一步用于将经噪声抑制处理的输出从噪声抑制组件传递到信号处理途径中的其他组

件。程序数据524也包括多径路由数据525,可用于将有噪输入信号沿多个信号途径传递至噪声抑制组件等处,这样可确保该组件在其他音频处理过程掌控或改变该有噪信号之前接收到该信号。

[0199] 计算设备500拥有附加特征和/或功能以及额外的接口,能够促进基本配置501与任何必要设备和接口之间的通信。例如,总线/接口控制器540能够用于促进基本配置501与一个或多个数据存储设备550之间的通信,此类通信是通过存储接口总线541实现的。数据存储设备550可以是可移动存储设备551、非移动存储设备552,或者二者的组合。可移动存储设备和非移动存储设备的示例包括:软盘驱动器和硬盘驱动器(HDD)等磁盘设备、光盘(CD)机或数字通用光盘(DVD)机等光盘驱动器、固态硬盘(SSD)和磁带机等。计算机存储介质示例包括:采用任何方法或技术实施的易失性和非易失性介质、可移动和非移动介质,用于存储计算机可读指令、数据结构、程序模块和/或其他数据等信息。

[0200] 系统内存520、可移动存储设备551和非移动存储设备552均属于计算机存储介质。计算机存储介质包括但不限于RAM、ROM、EEPROM、闪存或其他存储技术、CD-ROM、数字通用光盘(DVD)或其他光存储器、盒式磁带、磁带、磁盘存储器或其他磁存储器,或者能够用于存储所需信息并能通过计算设备500进行存取的任何其他介质。任何此类计算机存储介质都可能是计算设备500的一部分。

[0201] 计算设备500也包括接口总线542,该接口总线用于促进从各种接口设备(例如输出接口、外围接口和通信接口等)到基本设置501的通信,此类通信是通过总线/接口控制器540实现的。示例输出设备560包括一个图形处理单元561和一个音频处理单元562,配置其中任一单元或同时配置两个单元,可与显示器或扬声器等各种外部设备进行通信,此类通信是通过一个或多个A/V端口563实现的。示例外围接口570包括一个串行接口控制器571或者一个并行接口控制器572,这两种接口控制器均可经过配置,实现与输入设备(例如,键盘、鼠标、笔、语音输入设备或接触式输入设备等)等外部设备或其他外围设备(例如打印机或扫描仪等)的通信,此类通信是通过一个或多个I/O端口573实现的。示例通信设备580包括一个网络控制器581,对该控制器进行布置,可以促进与一个或多个其他计算设备590之间的网络通信(未显示),此类通信是通过一个或多个通信端口582实现的。这样的通信连接是通信介质的一个示例。常见的通信介质包括计算机可读指令、数据结构、程序模块或者调制数据信号形式的其他数据,比如载波或其他传输机制,还包括任何信息传递介质。“调制数据信号”可以是拥有一个或多个特征集的信号,或者可在信号中对信息进行编码,使其发生变化。举例来说,通信介质包括但不限于:有线网络或直接导线连接等有线介质,以及声频、无线电频率(RF)、红外线(IR)和其他无线介质等无线介质。本文中使用的“计算机可读介质”一词包括存储介质和通信介质。

[0202] 计算设备500可以实现为小型便携式(或移动)电子设备的一部分,比如手机、个人数据助理(PDA)、个人媒体播放器设备、无线网络监视设备、个人耳机设备、特定应用设备或者具有上述任何功能的混合型设备。计算设备500也可作为个人电脑来实现,包括手提电脑和非手提电脑配置。

[0203] 系统方面的硬件和软件实现几乎没有什么区别;使用硬件还是软件通常(但并非总是如此,在某些情况下,选择硬件还是软件可能非常重要)是一种设计选择,代表了成本与效率的权衡。本文描述的流程和/或系统和/或其他技术可以借助多种媒介物(例如硬件、

软件和/或固件)发挥作用,而且流程和/或系统和/或其他技术部署的环境不同,首选的媒介物也将有所不同。例如,如果实现人员确定速度和准确性是最重要的,则他可能会倾向选择硬件和/或固件媒介物;如果确定灵活性是最重要的,则可能会倾向选择软件实现。在一个或多个其他情境中,实现人员可能也会选择硬件、软件和/或固件的结合使用。

[0204] 以上详细说明通过使用方块图、流程图和/或示例,列出了设备和/或流程的多种实施例。由于这些方块图、流程图和/或示例中包含一个或多个功能和/或操作,相关领域的人员将获得这样的理解:这些方块图、流程图或示例中的每个功能和/或操作可借助广泛的硬件、软件、固件或者三者的任意组合来单独和/或同时实现。

[0205] 在一个或多个实施例中,本文描述的发明的几个部分可以通过专用集成电路(ASIC)、现场可编程门阵列(FPGA)、数字信号处理器(DSP)或者其他集成格式得以实现。然而,所属领域技术人员会发现,本文描述的实施例(全部或部分)的某些方面能够在集成电路中等效实现,作为在一个或多个计算机上运行的一个或多个计算机程序(例如在一个或多个计算机系统上运行的一个或多个程序)、作为在一个或多个处理器上运行的一个或多个程序(例如在一个或多个微处理器上运行的一个或多个程序)、作为固件或者作为上述形式的任意组合。所属领域技术人员还将进一步认识到,根据本发明,所属领域技术人员可以轻松地为软件和/或固件设计电路和/或编写代码。

[0206] 此外,所属领域技术人员将会理解到,本文描述的发明的运行机制能够作为各种形式的程序产品进行分配,而且无论用于实际执行这种分配的信号传输介质为何种类型,本文描述的发明的说明性实施例均适用。信号传输介质的示例包括但不限于以下内容:可记录类型的介质,比如软盘、硬盘驱动器、光盘(CD)、数字视频光盘(DVD)、数字磁带和计算机存储器等;以及传输类型的介质,比如数字和/或模拟通信介质(例如光纤电缆、波导管、有线通信链路和无线通信链路等)。

[0207] 所属领域技术人员也将认识到,在这一领域内,以本文所述方式描述设备和/或流程,然后使用工程实践将如此描述的设备 and/或流程集成到数据处理系统中,这种做法是很常见的。也就是说,本文描述的设备 and/或流程中,至少有一部分可以通过合理数量的实验集成到数据处理系统中。所属领域技术人员将认识到,典型的数据处理系统通常包括一个或多个系统单元外壳;一个视频显示设备;一个易失性或非易失性存储器;微处理器和数字信号处理器等处理器;操作系统、驱动器、图形用户界面和应用程序等计算实体;一个或多个交互设备,比如触摸板或触摸屏;和/或控制系统,包括反馈回路和控制电动机(例如,针对传感位置和/或速率的反馈;用于移动和/或调整组件和/或数量的控制电动机)。典型的数据处理系统可能利用市面上可买到的适用组件来实现,比如数据计算/通信和/或网络计算/通信系统中通常使用的那些组件。

[0208] 关于本文中大量使用的复数和/或单数名词,所属领域技术人员可根据上下文和/或适用情况,采用单数或复数形式。为清楚起见,本文中明确列出了各种单数/复数的变换情况。

[0209] 鉴于本文已经披露了多个不同的方面和实施例,所属领域技术人员应能够触类旁通,想到其他方面和实施例。本文披露的各个方面和实施例是为了解释说明之用,而并非为了提出限制,以下权利要求指明了专利的真正范围和内在意义。

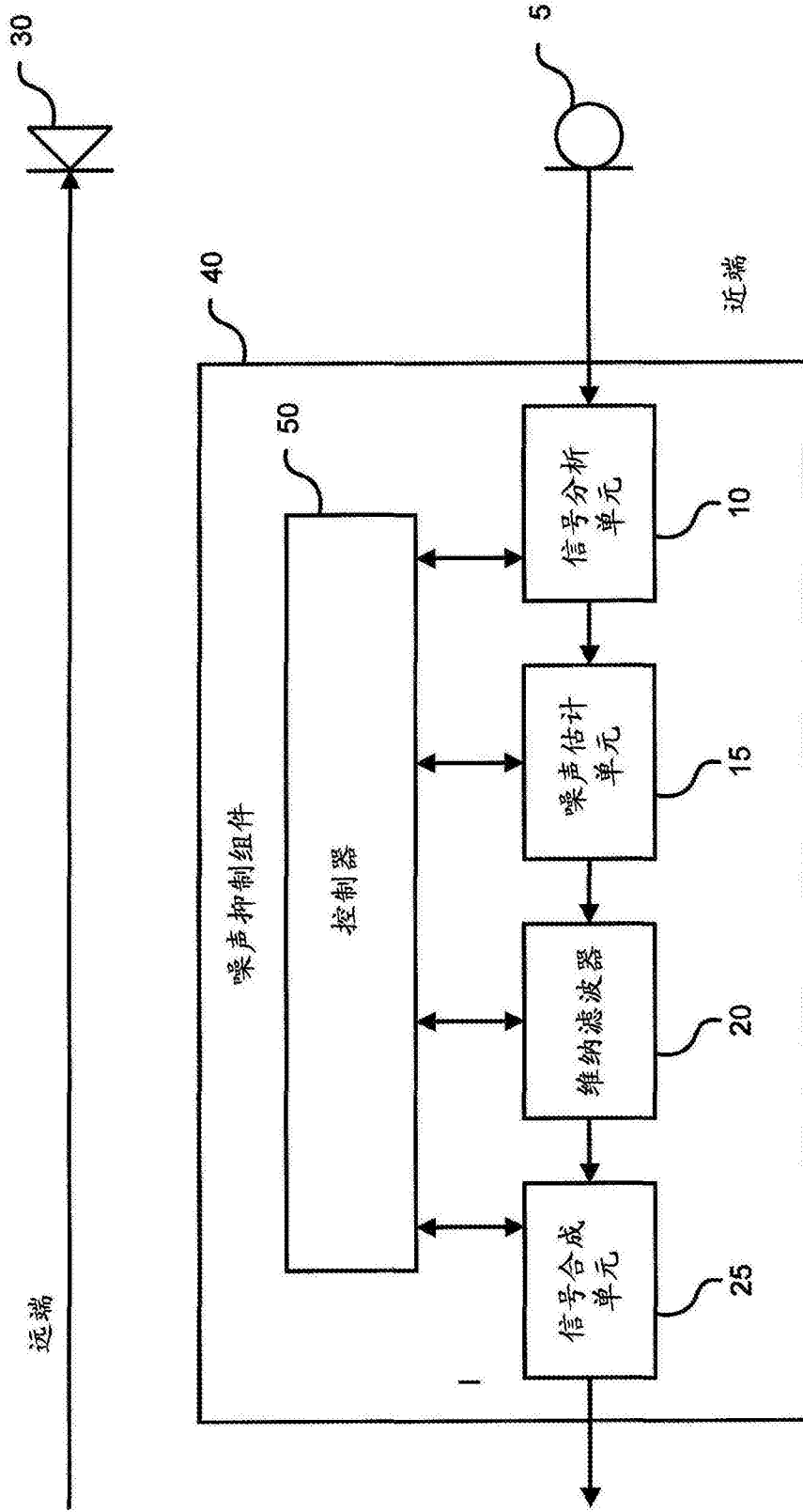


图1

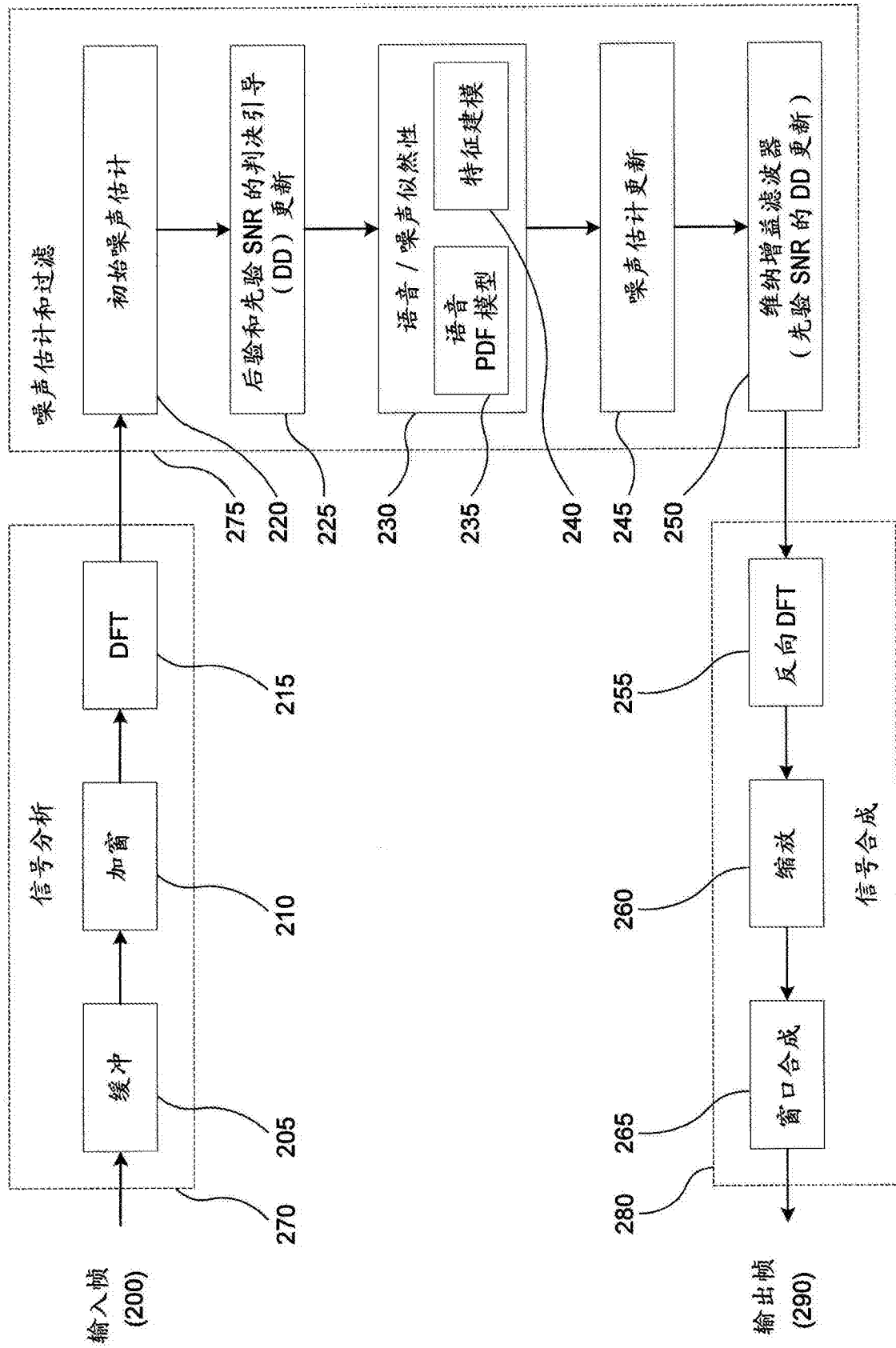


图2

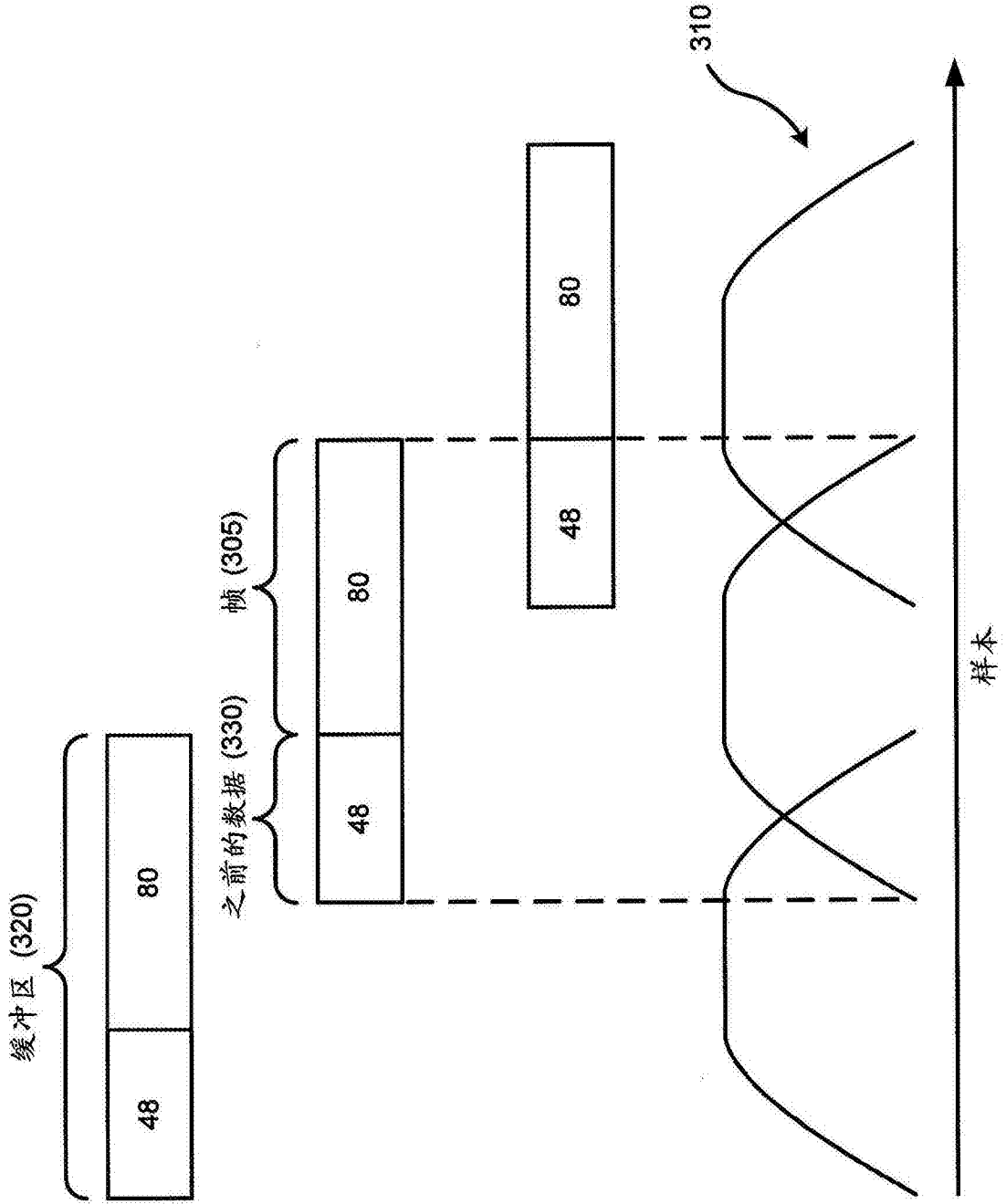


图3

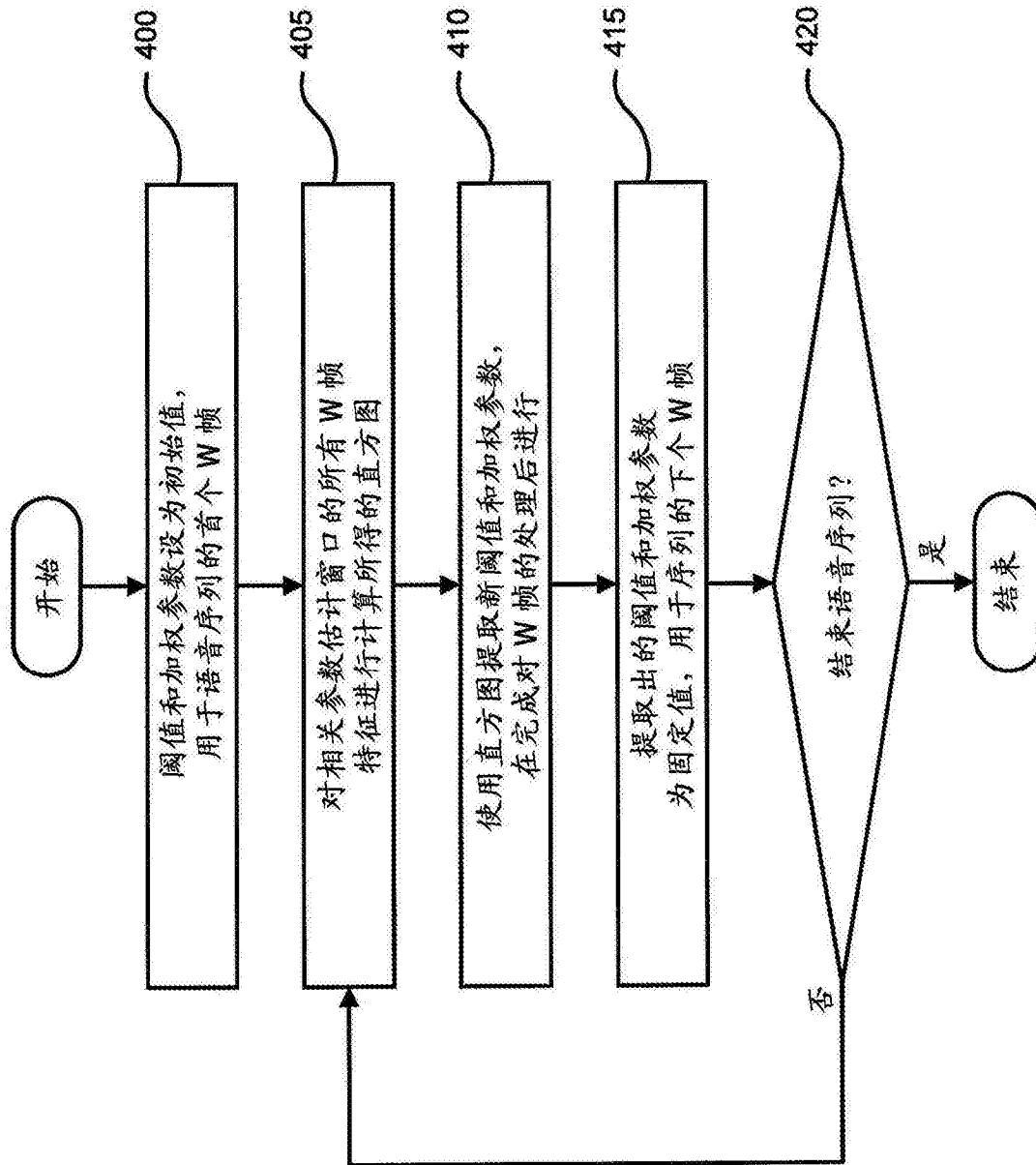


图4

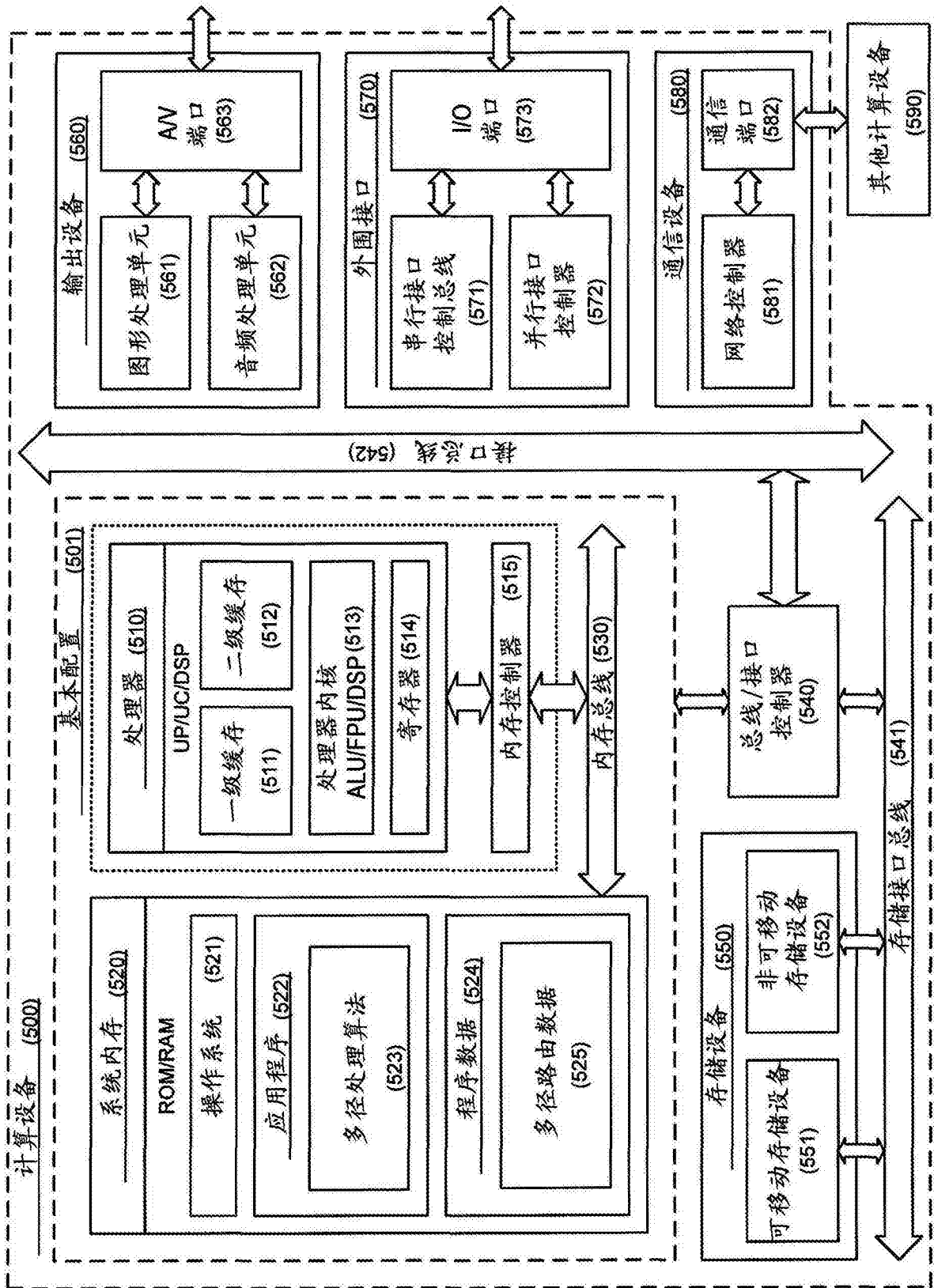


图5