



(12) 发明专利申请

(10) 申请公布号 CN 102708874 A

(43) 申请公布日 2012. 10. 03

(21) 申请号 201210052878. 0

(22) 申请日 2012. 03. 02

(30) 优先权数据

13/039, 576 2011. 03. 03 US

(71) 申请人 微软公司

地址 美国华盛顿州

(72) 发明人 H·N·基科利

(74) 专利代理机构 上海专利商标事务所有限公司
31100

代理人 蔡悦

(51) Int. Cl.

G10L 21/02 (2006. 01)

H04R 3/00 (2006. 01)

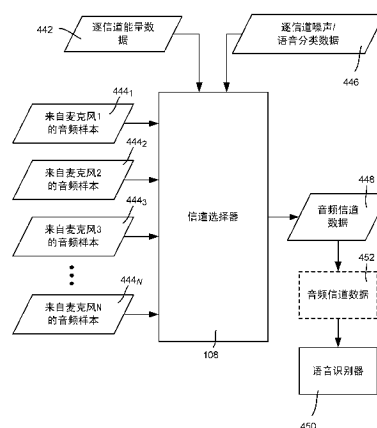
权利要求书 1 页 说明书 6 页 附图 6 页

(54) 发明名称

麦克风阵列的噪声自适应波束形成

(57) 摘要

本发明涉及麦克风阵列的噪声自适应波束形成。本发明涉及基于在不存在实际信号（例如，没有语音）时测量的本底噪声能量水平在麦克风阵列信道之间进行动态选择的噪声自适应波束形成器。当检测到语音（或类似的想要的信号）时，波束形成器选中哪个麦克风信号被用于信号处理，例如，对应于最低噪声信道的那个麦克风信号。可选中多个信道，它们的信号被合并。当不再检测到实际信号时，波束形成器切换回噪声测量阶段，使得波束形成器随着（包括基于每一个麦克风的）噪声水平改变而动态地适应，以考虑麦克风硬件区别、变化的噪声源以及个别的麦克风恶化。



1. 一种在计算环境中的系统,包括:麦克风阵列,所述麦克风阵列包括对应于各自输出信号的信道的多个麦克风 (102_1-102_N),耦合于所述阵列的机制 (104_1-104_N),其被配置成确定每个信道的本底噪声数据,以及信道选择器 (108),其被配置成基于所述每个信道的本底噪声数据来选中哪个或哪些信道被用于信号处理,其中所述信道选择器动态地适应于本底噪声数据的改变。

2. 如权利要求 1 所述的系统,其特征在于,所述信道选择器在任一时段选择单个信道用于信号处理,并且在该时段内丢弃来自每个其它信道的信号。

3. 如权利要求 1 所述的系统,其特征在于,所述信道选择器在任一时段选择一个或多个信道用于信号处理,并且所述系统还包括被配置成当选中了两个或更多信道时合并来自每个被选择的信道的信号的机制。

4. 如权利要求 1 所述的系统,其特征在于,还包括被配置成确定何时获取本底噪声数据的分类器。

5. 一种在计算环境中至少部分地在至少一个处理器上执行的方法,包括:

(a) 在噪声测量阶段确定 (506) 噪声数据,包括对应于麦克风阵列的麦克风的多个信道的每个信道的噪声数据;

(b) 使用 (514) 所述噪声数据来选中哪个或哪些信道被用于噪声测量阶段之后的信号处理;以及

(c) 返回到步骤 (a) 以使信道选择动态地适应于噪声数据随时间的变化。

6. 如权利要求 5 所述的方法,其特征在于,确定噪声数据包括计算对应于每个信道的能量水平的数据。

7. 如权利要求 5 所述的方法,其特征在于,还包括基于信道的一个或多个输入信号对输入信号对应于噪声还是用于信号处理的信号进行分类,以用于确定何时从步骤 (a) 切换到步骤 (b),并且用于确定何时从步骤 (b) 切换到步骤 (c)。

8. 一个或多个具有计算机可执行指令的计算机可读介质,所述计算机可执行指令在被执行时执行以下步骤,包括:

(a) 在噪声测量阶段确定 (508) 噪声数据,包括获取对应于麦克风阵列的麦克风的多个信道的每个信道的本底噪声能量水平;

(b) 检测 (502) 语音,并且切换到选择阶段,选择阶段使用所述噪声数据来选择 (514) 哪个或哪些信道被用于语音识别;

(c) 输出 (518) 对应于所选择的用于语音识别的一个或多个信道的信号;以及

(d) 返回到步骤 (a) 以使信道选择动态地适应于噪声数据随时间的变化。

9. 如权利要求 8 所述的一个或多个计算机可读介质,其特征在于,检测语音包括检测本底噪声能量水平中的改变。

10. 如权利要求 8 所述的一个或多个计算机可读介质,其特征在于,在步骤 (b) 选择多个信道,并且还具有计算机可执行指令,包括将来自所选择的信道的信号合并成合并的信号,用于在步骤 (c) 输出。

麦克风阵列的噪声自适应波束形成

技术领域

[0001] 本发明涉及通过不使用含噪声的信号来丢弃使性能降低的麦克风信号。

背景技术

[0002] 麦克风阵列捕捉来自多个传感器的信号,并处理那些信号以便提高信噪比。在传统的波束形成中,一般方法是合并来自所有传感器(信道)的信号。波束形成的一个典型用途是将经合并的信号提供给语音识别器用于语音识别。

[0003] 然而,在实践中,这一方法实际上会降低整体性能,并且确实,有时候执行起来甚至比单个麦克风更糟。这部分是因为麦克风之间的个体硬件差异,这会导致不同的麦克风拾取不同类型且不同量的噪声。另一个因素是噪声源可能会动态改变。此外,不同的麦克风不同地恶化,这再次导致降低的性能。

发明内容

[0004] 提供本发明内容以便以简化形式介绍将在以下的具体实施例中进一步描述的一些代表性概念。本发明内容不旨在标识出所要求保护的主题的关键特征或必要特征,也不旨在以限制所要求保护的主题的范围的任何方式来使用。

[0005] 简单来说,此处所描述的主题的各方面涉及一种自适应波束形成器/选择器藉由其来基于为每个信道所确定的本底噪声(noise floor)数据来选择使用麦克风阵列的哪些信道/麦克风的技术。在一个实现方式中,获取在没有实际信号(例如,没有语音)的时间期间的能量水平,并且一旦存在实际信号,信道选择器就基于本底信号数据选择要在信号处理中使用那个或哪些信道。本底噪声数据被反复测量,藉此,自适应波束形成器动态地适应于本底噪声数据随时间的变化。

[0006] 在一个实现方式中,信道选择器选择在任一时刻用于信号处理(例如,语音识别)的信号信道并丢弃其它信道的信号。在另一实现方式中,信道选择器选择一个或多个信道,其中当选择了两个或更多信道时,来自每个所选择的信道的信号被合并以用于信号处理。

[0007] 在一方面,分类器确定在噪声测量阶段何时获取本底噪声数据,以及在选择阶段何时作出选择。分类器可基于所检测到的能量水平的变化。

[0008] 结合附图阅读以下具体实施方式,本发明的其他优点会变得显而易见。

附图说明

[0009] 作为示例而非限制,在附图中示出了本发明,附图中相同的附图标记指示相同或相似的元素,附图中:

[0010] 图1是表示用于麦克风阵列的噪声自适应波束形成器/选择器的示例组件的方框图。

[0011] 图2是示例的8信道麦克风阵列的麦克风的噪声相对于语音信号的表示。

[0012] 图3是表示估计麦克风阵列的输入信道的本底噪声能量的机制的方框图。

[0013] 图 4 是表示噪声自适应波束形成器 / 选择器可如何使用基于噪声的信道选择来自适应地将信号提供给语音识别器的方框图。

[0014] 图 5 是表示噪声检测阶段和信道选择阶段中的示例步骤的流程图。

[0015] 图 6 是表示其中可实现此处所描述的各种实施例的一个或多个方面的示例性非限制计算系统或操作环境的框图。

具体实施方式

[0016] 此处所描述的技术的各方面总的来说涉及通过不使用含噪声的信号来丢弃使性能降低的麦克风信号。此处所描述的噪声自适应波束形成技术尝试最小化麦克风硬件差异、动态改变的噪声源麦克风恶化和 / 或可能地其它因素所导致的不利效果, 得到例如包括在最初以及在随硬件降级的一段时间内对语音识别有益的信号。

[0017] 应当理解, 此处的任何示例均是非限制性的。举例来说, 虽然语音识别是此处所描述的技术的一项有用的应用, 但是任何声音处理应用 (例如, 方向性放大和 / 或噪声抑制) 可类似地受益。因此, 本发明不限于在此描述的任何具体的实施例、方面、概念、结构、功能或示例。相反, 此处所描述的实施例、方面、概念、结构、功能或示例中的任一个都是非限制性的, 并且本发明一般能够以在声音处理和 / 或语音识别方面提供好处和优点的各种方式来使用。

[0018] 图 1 示出了一个示例噪声自适应波束形成实现的组件。对应于麦克风阵列信道 102_1-102_N 的多个麦克风各自提供用于选择和 / 或波束形成的信号; 可以理解的是, 在一给定的阵列实现方式中, 可存在至少两个 (至多任意实际数量) 这样的麦克风。

[0019] 并且, 麦克风阵列不需要被对称安放, 并且实际上, 在一个实现方式中, 麦克风出于各种原因是非对称安放的。此处所描述的技术的一个应用是用于移动机器人, 其可在等待来自人的语音时自主地四处移动, 并因而可动态地暴露于不同的噪声源。

[0020] 如图 1 中的能量检测器 104_1-104_N 所示, 此处所描述的噪声自适应波束形成技术监视每个麦克风中的噪声能量水平, 包括当没有实际信号, 即只有噪声时。图 2 是一个示例 8 信道麦克风阵列的这种能量水平的表示, 其中, 框 221 表示阵列的 “MIC1” (麦克风 1) 的 “无实际信号” 状态。最初, 不存在真正的输入信号, 因此, 麦克风的输出仅仅是所感测到的噪声。需要注意的是, 图 2 中的框 221 (以及其它的框) 并不意在表示准确的一个采样帧或一组采样帧; (例如, 典型的采样率是 16K 帧 / 秒)。

[0021] 当存在信号时, 如图 2 中框 222 所示, 能量增加, 并且能量检测器 104_1-104_N 提供指示每个信道的增加的估计。噪声 / 语音分类器 106_1-106_N 可被用于确定 (例如, 基于训练的增量能量水平或阈值能量水平) 信号是噪声还是语音, 并且将这一信息馈送给信道选择器 108。需要注意的是, 每个分类器可包括其各自的归一化、过滤、平滑化和 / 或其它这类技术来作出其判断, 例如, 能量可能需要在一定数量的帧上保持增加或者以其它方式匹配将被视为语音的语音模式, 以便消除因被视为语音而可能产生的短时噪声能量毛刺之类。需要注意的是, 将单个噪声或语音分类器用于全部的信道也是可行的, 例如仅仅使用信道中的一个用于分类, 或混合音频信道中的一些或全部用来分类 (同时, 保持它们分开以便选择)。

[0022] 基于噪声水平, 当检测到语音时, 信道选择器 108 动态地确定哪个 (或哪些) 麦克

风的信号将被用于进一步处理,例如,语音处理,以及哪些信号将被丢弃。在图 1 的示例中,麦克风 MIC1 在没有信号时具有相对大量的噪声,而麦克风 MIC7 在没有信号时具有最少量的噪声(框 227)。因此,当语音确实出现时(对每个信道而言,近似于对应于框 222 的时间),来自麦克风 MIC7 的信号将可能被使用,而来自麦克风 MIC1 的信号则可能被丢弃。

[0023] 在噪声自适应波束形成的一个实现方式中,仅仅对应于最低噪声信号的信道被选择,例如,在图 2 中,只有来自麦克风 MIC7 的信号被选择,因为在没有信号时,它的本底噪声比其它麦克风的本底噪声低。在可替换实现方式中,信道选择器 108 可选择来自多个信道的信号,这些信号随后被合并为一个合并的信号用于输出。例如,两个最低的噪声信道可被选择并合并。可考虑阈值能量水平或相对能量水平数据以便当下一最低噪声信道过于嘈杂或相对过于嘈杂,则只选择最低噪声信道。作为另一可选实现方式,每个信道可被给予与该信道的噪声反相关(以任何适当的数学方式)的权重并且使用加权合并来合并。

[0024] 以此方式,本底噪声追踪的使用自动消除了(或极大地减少了)含噪声的麦克风的不利效果,因为含噪声的麦克风具有较高的噪声水平且因此它们的信号不被使用。这一方法还消除了靠近给定场景中的噪声源(例如,在电视机扬声器附近)的麦克风的效果。类似地,随着麦克风硬件衰竭或以其它方式变得损坏(有些麦克风变坏且通常产生高的噪声水平),噪声自适应波束形成器自动消除了这些麦克风的效果。

[0025] 图 3 是表示示例本底噪声能量估计器机制 330 的方框图,所述机制诸如用于信道中的一个的能量检测器。给定麦克风 X 的传入的音频样本 332 可被过滤(框 334)以从信号中移除任何 DC 分量,并且随后在将结果输入到快速傅立叶变换(FFT)338 之前由已知的汉明窗口函数 336(或其它这类函数)进行处理(例如,平滑化)。基于 FFT 输出,本底噪声能量评估器 340 以普遍公知的方式来计算噪声能量数据 342(例如,代表值)。

[0026] 如图 4 中所示,每个信道的噪声能量数据 442 被馈送到信道选择器 108。取决于表示来自每个麦克风的噪声能量水平估计的数据 442,当检测到对应于音频样本 444₁-444_N 的语音时,如分类数据 446 所示的,信道选择器 108 决定是否使用来自每个麦克风的信号。信道选择器 108 输出所选择的信号作为选定的音频信道数据 448,用于馈送到语音识别器 450。需要注意的是,如框 452 所示,如果信道选择器 108 被配置成选择不止一个信道并且确实这样选择了,则使用各种方法中的任意一种来合并来自多个信道的信号。

[0027] 图 5 概述了与信道选择和使用有关的各个示例操作,从步骤 502 开始,在步骤 502,对当前输入是噪声还是语音作出分类。如果是噪声,则步骤 504 选择一个信道,并且如以上所描述的,步骤 506 确定该信道的本底噪声能量。步骤 508 表示计算这一信道的噪声数据,例如计算一定数量个帧的平均噪声能量水平、执行舍入、归一化和/或诸如此类的计算,以便提供信道选择器所期望的噪声数据。步骤 510 将噪声数据与该信道(例如,该信道的标识符)相关联。

[0028] 步骤 512 对每一个其它信道重复噪声测量阶段处理的步骤 504-510。当每个信道的噪声数据与信道身份相关联,过程返回到如上所述的步骤 502。

[0029] 在某个后续时刻,检测到了语音,由此步骤 502 分支到步骤 514 以切换到选择阶段,该选择阶段选择具有指示最低本底噪声水平的相关联数据的信道(或多个信道),以用于进一步处理。在步骤 514 选择了不止一个信道的情况下,步骤 516 合并来自每个信道的信号。步骤 518 输出所选择的信道或合并的信道的信号以供进一步处理,例如,语音识别,

之后返回步骤 502。

[0030] 需要注意的是,在图 5 中示出了步骤 520 处的可任选的延迟,该延迟可用于在检测到语音之后而在切换回评估噪声之前进行延迟。虽然语音识别器可能正持续地接收包含语音和噪声两者的输入时,但在简短的暂停期间切换麦克风可能导致降低的识别准确性。例如,在简短的暂停期间的扬声器的吸入或其它自然噪声可被麦克风检测为噪声,否则该麦克风将具有最佳噪声结果,而从这一麦克风切换到其它麦克风可能提供来自更嘈杂的另一个麦克风的语音输入。因此,通过延迟,在简短的暂停期间,扬声器有机会恢复发声而不是切换回噪声测量。作为延迟的替代(或补充),信道选择操作可包括平滑化、求平均等等,来消除任何这类快速麦克风变化等。例如,如果麦克风已具有相对于其它麦克风的低噪声并因此其信号被选中一段时间,其本底噪声能量的突然变化可被忽视,以使得不会因为瞬间的短时脉冲干扰之类就切换到另一麦克风。

[0031] 由此可见,所描述的是一种使用本底噪声水平来确定哪个麦克风被用于波束形成的噪声自适应波束形成技术。噪声自适应波束形成技术动态地更新该信息,使得动态地适应于变化的环境(与传统波束形成相对)。

[0032] 示例性计算设备

[0033] 如上所述,有利地,本文所描述的技术可应用于任何设备。因此,应当理解,构想了结合各实施例使用的包含机器人在内的所有种类的手持式、便携式和其它计算设备和计算对象。因此,以下在图 6 中所述的通用远程计算机只是计算设备的一个示例。

[0034] 各实施例可部分地经由操作系统来实现,以供设备或对象的服务开发者使用和/或被包括在用于执行此处所述的各实施例的一个或多个功能方面的应用软件内。软件可以在由诸如客户机工作站、服务器或其它设备等一个或多个计算机执行的诸如程序模块等计算机可执行指令的通用上下文中描述。本领域的技术人员可以理解,计算机系统具有可用于传递数据的各种配置和协议,并且由此没有特定配置或协议应当被认为是限制性的。

[0035] 图 6 由此示出了其中可实现本文所述的各实施例的一个或多个方面的合适的计算机系统环境 600 的一个示例,尽管如上所述,计算机系统环境 600 仅为合适的计算环境的一个示例,并非对使用范围或功能提出任何限制。此外,也不应当将计算机系统环境 600 解释为对在示例性计算机系统环境 600 中所示的组件中的任何一个或其组合有任何依赖。

[0036] 参考图 6,用于实现一个或多个实施例的示例性远程设备包括计算机 610 形式的通用计算设备。计算机 610 的组件可包括,但不限于,处理单元 620、系统存储器 630、以及将包括系统存储器的各种系统组件耦合到处理单元 620 的系统总线 622。

[0037] 计算机 610 通常包括各种计算机可读介质,并且可以是可由计算机 610 访问的任何可用介质。系统存储器 630 可包括诸如只读存储器 (ROM) 和/或随机存取存储器 (RAM) 之类的易失性和/或非易失性存储器形式的计算机存储介质。作为示例而非限制,系统存储器 630 还可包括操作系统、应用程序、其他程序模块、以及程序数据。

[0038] 用户可通过输入设备 640 向计算机 610 输入命令和信息。监视器或其他类型的显示设备也经由诸如输出接口 650 之类的接口连接到系统总线 622。除监视器以外,计算机还可包括诸如扬声器和打印机之类的其他外围输出设备,它们可通过输出接口 650 连接。

[0039] 计算机 610 可使用到一个或多个其他远程计算机(诸如远程计算机 670)的逻辑连接在联网或分布式环境中操作。远程计算机 670 可以是个人计算机、服务器、路由器、网

络 PC、对等设备或其他常见网络节点、或者任何其他远程媒体消费或传输设备,并且可包括以上关于计算机 610 所述的任何或全部元件。图 6 所示的逻辑连接包括诸如局域网 (LAN) 或广域网 (WAN) 之类的网络 672,但也可包括其他网络 / 总线。这些联网环境在家庭、办公室、企业范围的计算机网络、内联网和因特网中是常见的。

[0040] 如上所述,尽管结合各种计算设备和网络体系结构描述了各示例性实施例,但基本概念可被应用于其中期望改进资源使用的效率的任何网络系统和任何计算设备或系统。

[0041] 而且,存在实现相同或相似功能的多种方法,例如适当的 API、工具箱、驱动程序代码、操作系统、控件、独立或可下载软件对象等,它们使得应用和服务能够使用此处提供的技术。由此,此处的各实施例从 API (或其他软件对象) 的观点以及从实现如此处描述的一个或多个实施例的软件或硬件对象构想。由此,此处所述的各实施例可具有完全采用硬件、部分采用硬件并且部分采用软件、以及采用软件的方面。

[0042] 本文中所使用的词语“示例性”意味着用作示例、实例、或说明。为避免疑惑,本文所公开的主题不限于这些示例。另外,在此所述的被描述为“示例性”的任意方面或设计并不一定要被解释为相比其它方面或设计更优选或有利。此外,在使用术语“包括”、“具有”、“包含”和其他类似词语的程度上,为避免疑惑,这些术语旨在当用于权利要求中时以类似于术语“包括”作为开放的过渡词的方式是包含性的而不排除任何附加或其他元素。

[0043] 如所述的,此处所述的各种技术可结合硬件或软件或,在适当时,以两者的组合来实现。如此处所使用的,术语“组件”、“模块”、“系统”等同样旨在指计算机相关实体,或者是硬件、硬件和软件的组合、软件或者是执行中的软件。例如,组件可以是,但不限于是,在处理器上运行的进程、处理器、对象、可执行码、执行的线程、程序和 / 或计算机。作为说明,在计算机上运行的应用和计算机都可以是组件。一个或多个组件可以驻留在进程和 / 或执行线程中,并且组件可以位于一个计算机内和 / 或分布在两个或更多计算机之间。

[0044] 如前所述的系统已经参考若干组件之间的交互来描述。可以理解,这些系统和组件可包括组件或指定的子组件、某些指定的组件或子组件和 / 或附加的组件,并且根据上述内容的各种置换和组合。子组件还可作为通信地耦合到其他组件的组件来实现,而不是被包括在父组件内 (层次性)。另外,应注意到一个或多个组件可被组合成提供聚集功能的单个组件,或被分成若干单独的子组件,且诸如管理层等任何一个或多个中间层可被设置成通信耦合到这样的子组件以便提供集成功能。此处所述的任何组件也可与一个或多个此处未专门描述的但本领域技术人员一般已知的其他组件进行交互。

[0045] 鉴于本文所述的示例性系统,可根据参考各附图的流程图还可理解根据所述的主题来实现方法。尽管为了说明简洁起见,作为一系列框示出和描述的方法,但是应当理解,各实施例不仅仅限于框的次序,因为一些框可以与此处所描绘和描述的框不同的次序发生和 / 或与其他框并发地发生。尽管经由流程图示出了非顺序或分支的流程,但可以理解,可实现达到相同或类似结果的各种其他分支、流程路径和框的次序。此外,某些示出的框在实现下文所述的方法中是可选的。

[0046] 结语

[0047] 尽管本发明易于作出各种修改和替换构造,但其某些说明性实施例在附图中示出并在上面被详细地描述。然而应当了解,这不旨在将本发明限于所公开的具体形式,而是相反地,旨在覆盖落入本发明的精神和范围之内内的所有修改、替换构造和等效方案。

[0048] 除此处所述的各实施例以外,应当理解,可使用其他类似实施例,或者可对所述实施例作出修改和添加以便执行对应实施例的相同或等效功能而不背离这些实施例。此外,多个处理芯片或多个设备可共享此处所述的一个或多个功能的性能,并且类似地,存储可跨多个设备实现。因此,本发明不应限于任何单个实施例,而是应当根据所附权利要求书的广度、精神和范围来解释。

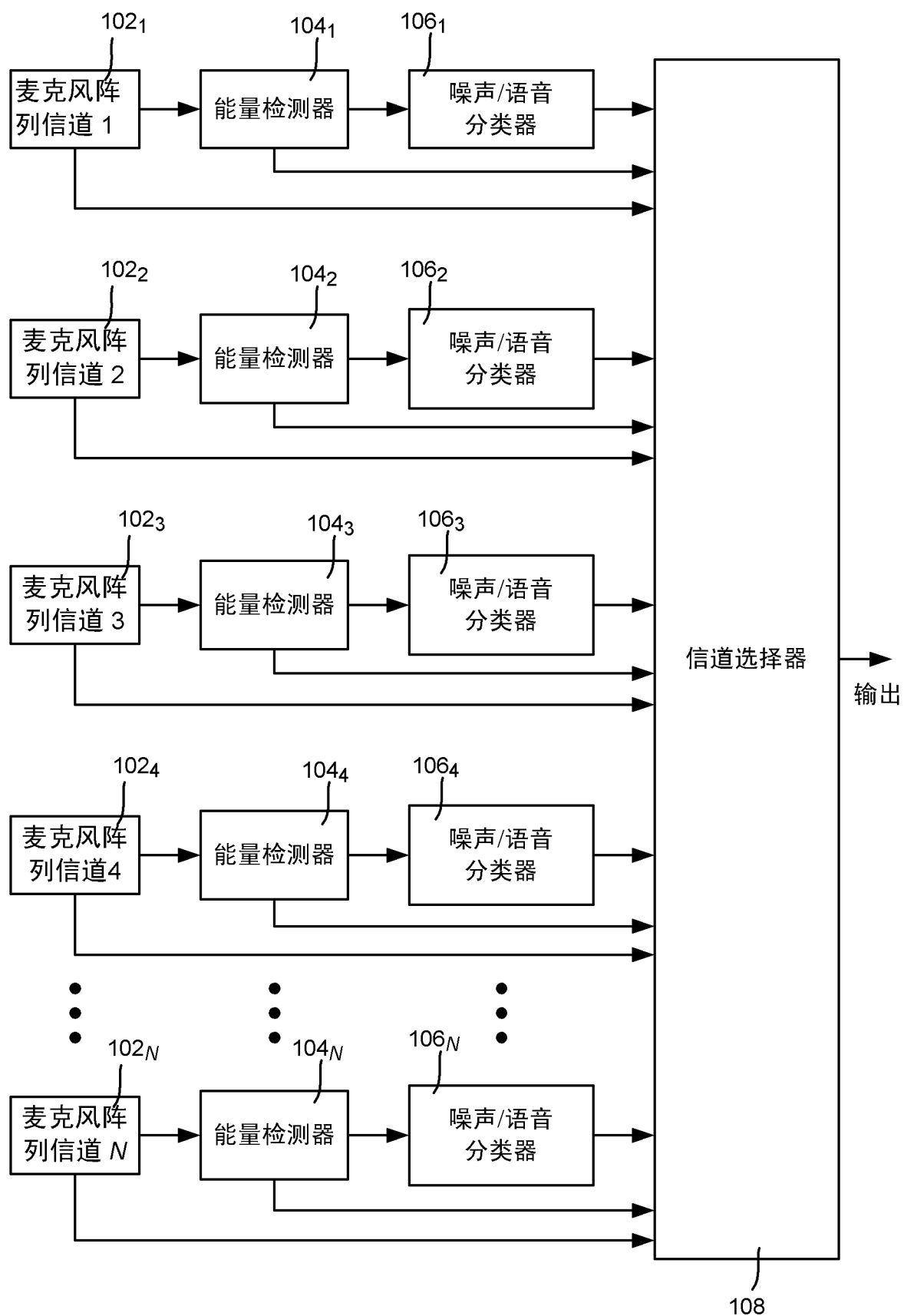


图 1

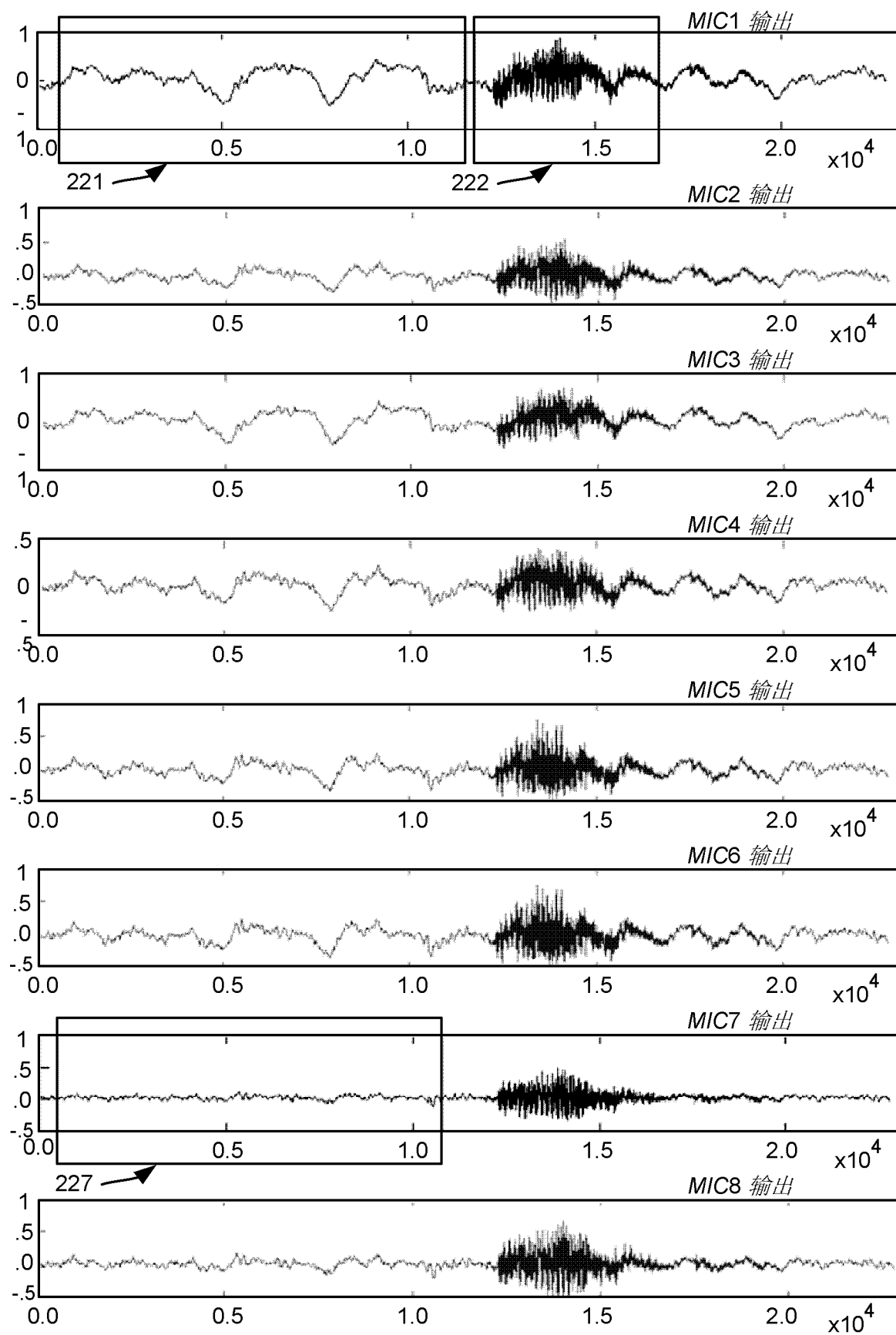


图 2

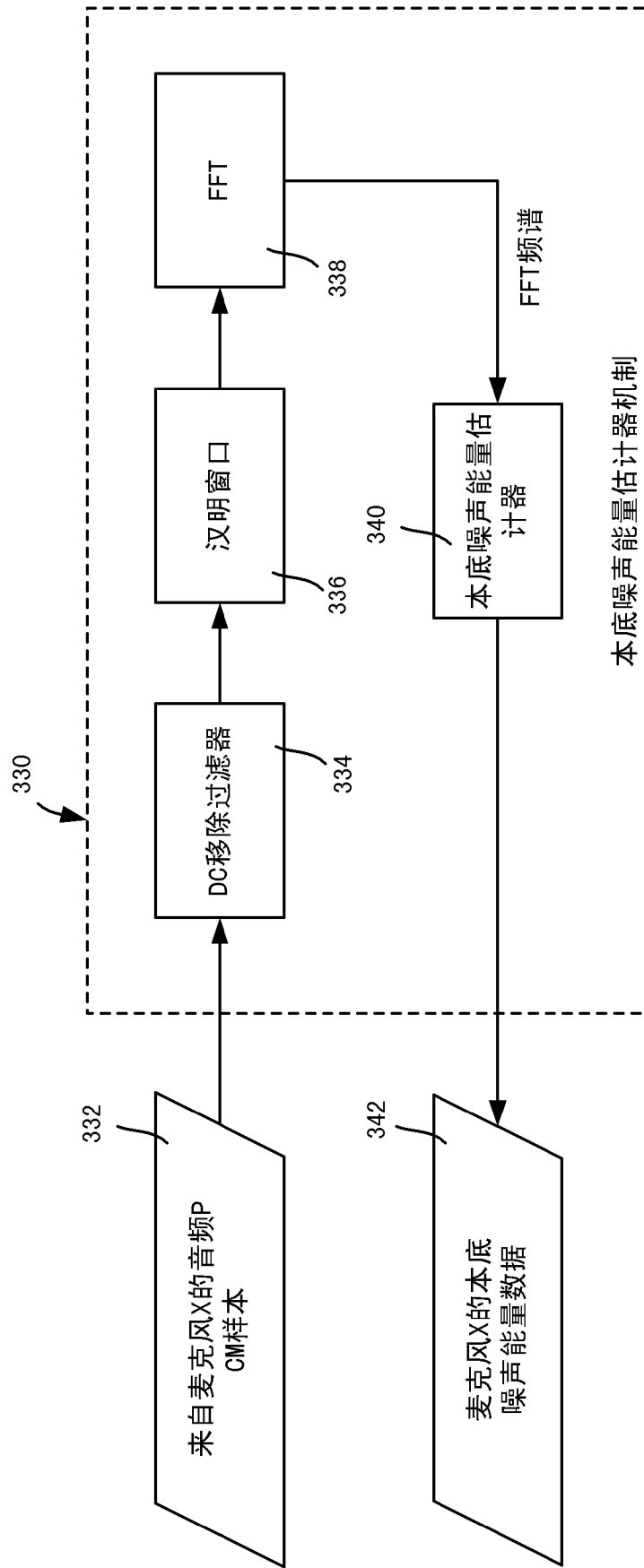


图 3

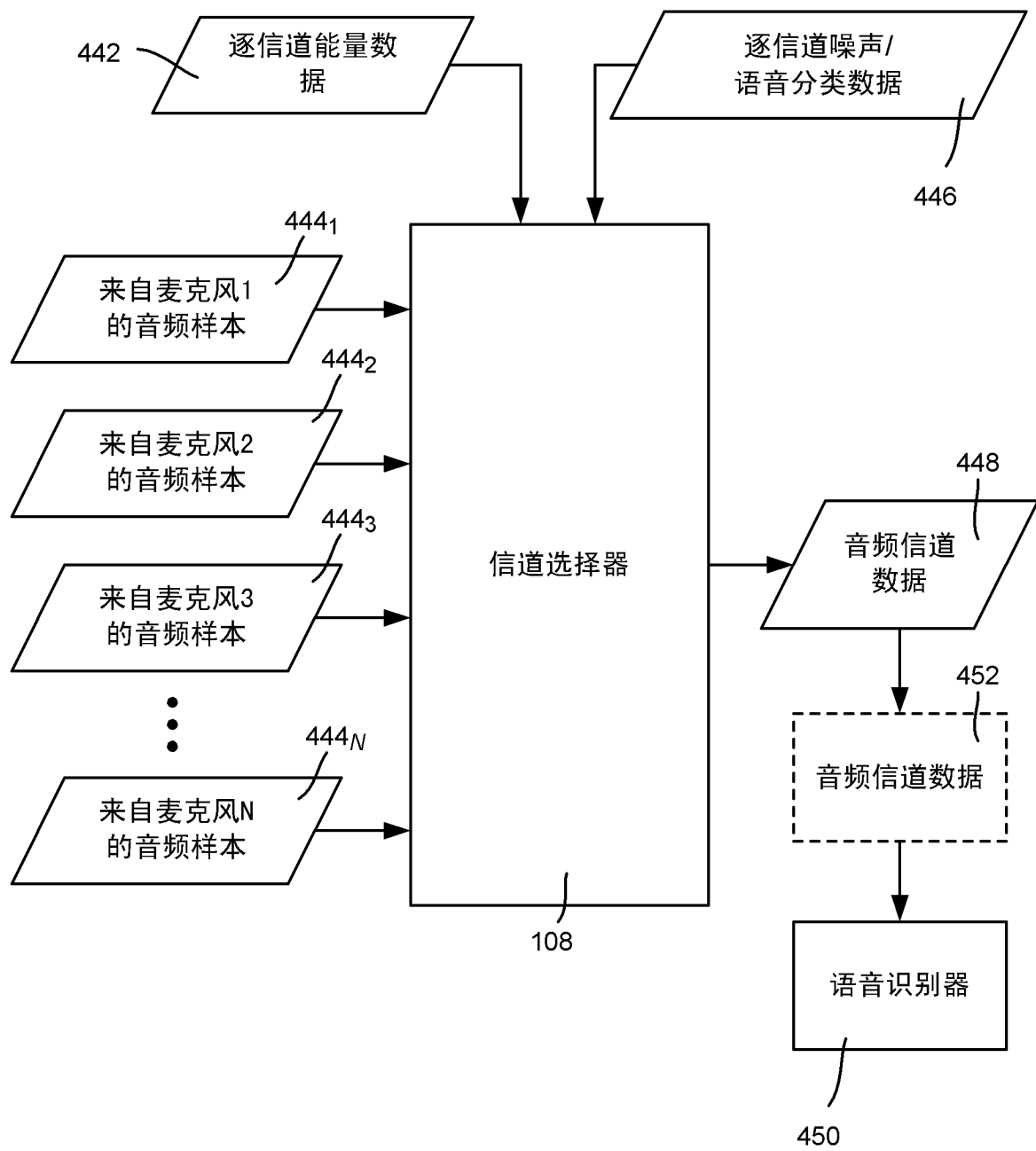


图 4

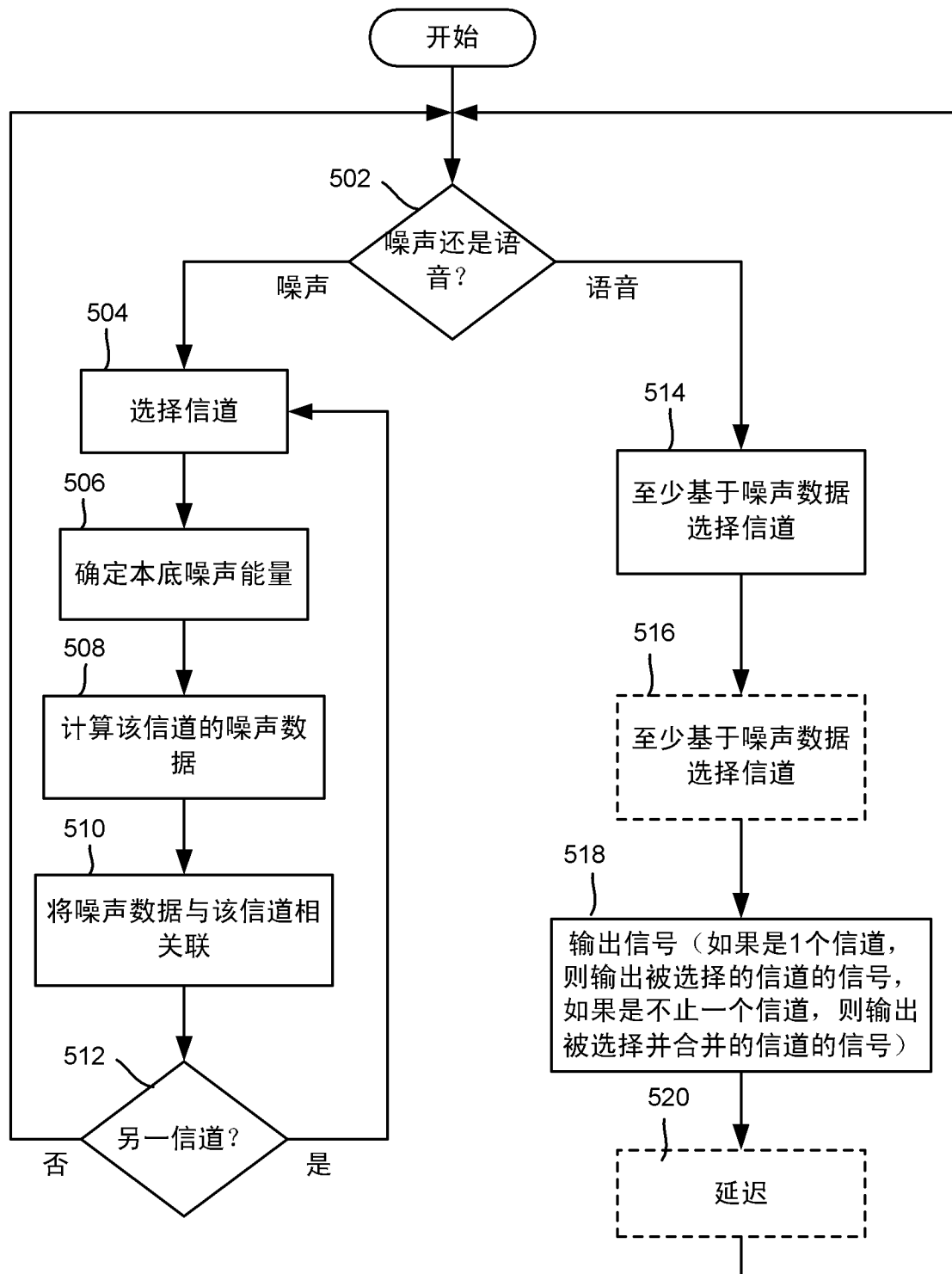


图 5

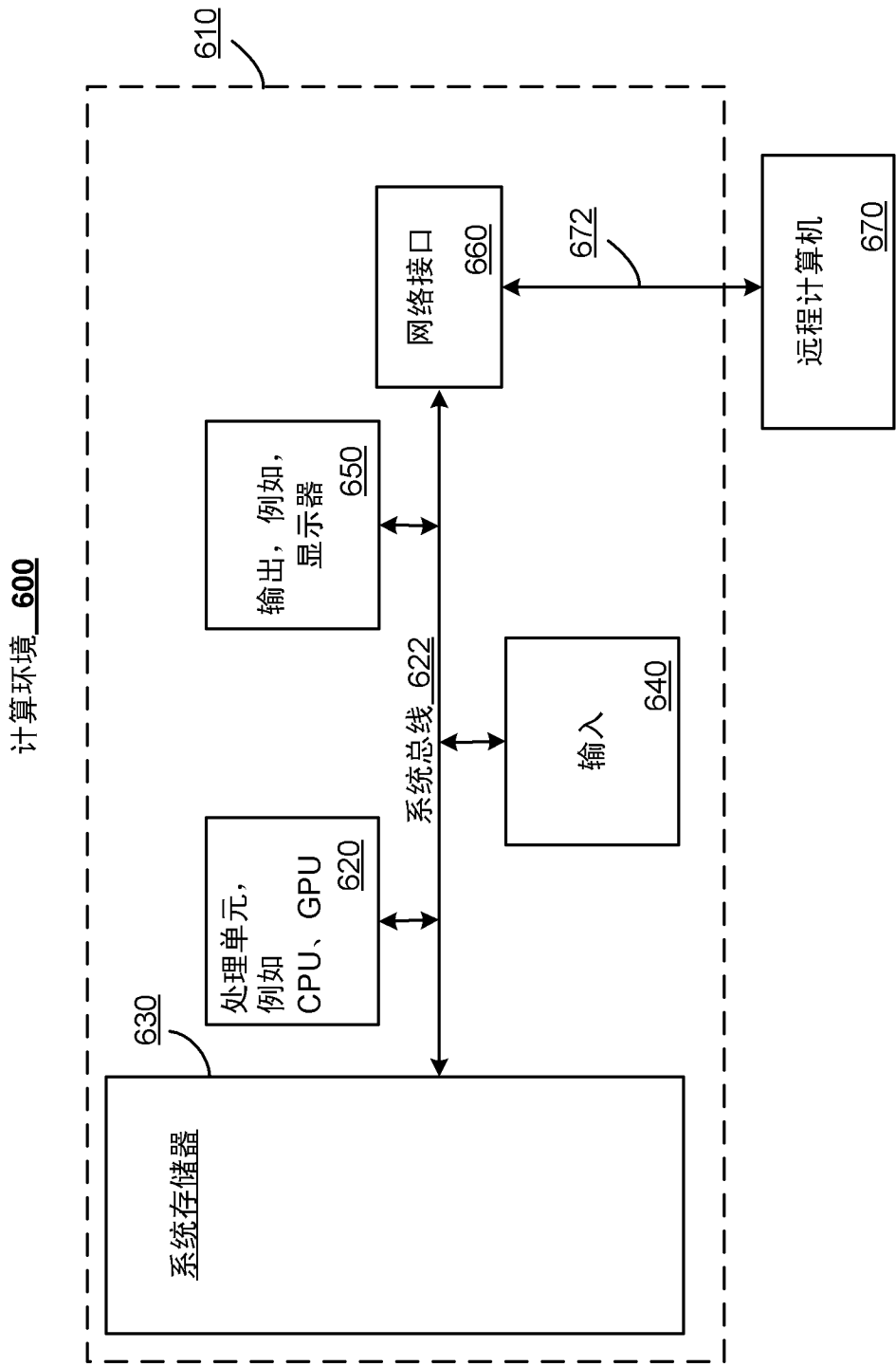


图 6