[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 21/02 (2006.01)
G10L 15/20 (2006.01)

[12] 发明专利说明书

专利号 ZL 200510092458.5

[45] 授权公告日 2010 年 1 月 20 日

[11] 授权公告号 CN 100583243C

[22] 申请日 2005.8.17
[21] 申请号 200510092458.5
[30] 优先权
[32] 2004.9.17 [33] US [31] 10/944,235
[73] 专利权人 微软公司
地址 美国华盛顿州
[72] 发明人 A·阿瑟洛 J·G·德罗坡
黄学东 张正友 刘自成
[56] 参考文献
US6754623B2 2004.6.22
US6289309B1 2001.9.11
WO02098169A1 2002.12.5
US20020039425A1 2002.4.4
US5241692A 1993.8.31
US20010037195A1 2001.11.1

Air – and bone – conductive integrated microphones for robustspeech detection and enhancement. YANLI ZHENG ET AL. AUTOMATIC SPEECH RECOGNITION AND UNDERSTANDING, 2003. ASRU '03. 2003 IEEE WORKSHOP ON ST. THOMAS, VI, USA. 2003

Combining standard and throat microphones for robustspeech recognition. GRACIARENA M ET AL. IEEE SIGNAL PROCESSING LETTERS IEEE USA, Vol. 10 No. 3. 2003
审查员 张 岩
[74] 专利代理机构 上海专利商标事务所有限公司
代理人 张政权

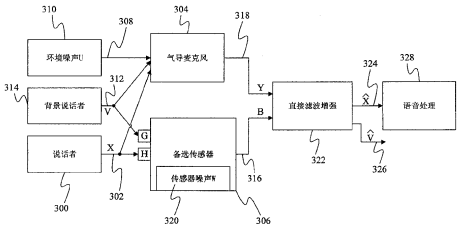
权利要求书 3 页 说明书 13 页 附图 7 页

[54] 发明名称

多传感器语音增强的方法和装置

[57] 摘要

一种方法和装置使用备选传感器信号和气导麦克风信号来确定对备选传感器的信道响应。该信道响应而后用于使用备选传感器信号的至少一部分来估算干净语音值。



1．一种确定对表示经降噪的语音信号的一部分的经降噪的值的估算的方法，所述方法包括：

使用除气导麦克风外的备选传感器生成一备选传感器信号，所述气导麦克风检测语音信号以及噪声信号，而所述备选传感器检测语音信号；

生成一气导麦克风信号；

使用所述备选传感器信号和所述气导麦克风信号来估算自说话者至所述备选传感器的路径的信道响应值；以及

使用所述信道响应来估算所述经降噪的值。

2．如权利要求 1 所述方法，其特征在于，估算信道响应值包括找出一目标函数的极值。

3．如权利要求 1 所述方法，其特征在于，估算信道响应值包括，将所述备选传感器信号建模为干净语音信号与所述信道响应卷积，并将结果与一噪声项相加。

4．如权利要求 1 所述方法，其特征在于，所述信道响应包括对干净语音信号的信道响应。

5．如权利要求 4 所述方法，其特征在于，还包括确定所述备选传感器对背景语音信号的信道响应。

6．如权利要求 5 所述方法，其特征在于，使用所述信道响应来估算所述经降噪的值包括，使用对所述干净语音信号的信道响应和对所述背景语音信号的信道响应来估算所述经降噪的值。

7．如权利要求 1 所述方法，其特征在于，还包括使用所述经降噪的值的估算来估算背景语音信号的值。

8．如权利要求 1 所述方法，其特征在于，估算信道响应值包括，使用所述备选传感器信号和所述气导麦克风信号的帧序列来估算对所述帧序列中的帧的单个信道响应值。

9．如权利要求 8 所述方法，其特征在于，使用所述信道响应来估算经降噪的值包括为所述帧序列中的每一帧估算一单独的经降噪的值。

10．如权利要求 1 所述方法，其特征在于，估算信道响应值包括，通过向在当前帧中的备选传感器信号和气导麦克风信号分配比前一帧中的备选传感器信号和气导麦克风信号更大的权重，来估算当前帧的值。

11．一种识别干净语音信号的方法，所述方法包括：

估算描述一备选传感器信号中的噪声的噪声参数，所述备选传感器检测语音信号；

使用所述噪声参数来估算自说话者至备选传感器的路径的信道响应；以及

使用所述信道响应来估算所述干净语音信号的值。

12．如权利要求 11 所述的方法，其特征在于，估算噪声参数包括，使用所述备选传感器信号来识别用户不发声的时段。

13．如权利要求 12 所述的方法，其特征在于，还包括在一气导麦克风信号中与用户不发声时段相关联的部分上执行语音检测，以识别无语音时段和背景语音时段。

14．如权利要求 13 所述的方法，其特征在于，还包括使用所述备选传感器信号中与无语音时段相关联的部分来估算所述噪声参数。

15．如权利要求 14 所述的方法，其特征在于，还包括使用所述无语音时段来

估算描述所述气导麦克风信号中的噪声的噪声参数。

16. 如权利要求 14 所述的方法，其特征在于，还包括使用所述备选传感器信号中与背景语音时段相关联的部分来估算对背景语音的信道响应。

17. 如权利要求 16 所述的方法，其特征在于，还包括使用对背景语音的信道响应来估算干净语音。

18. 如权利要求 11 所述的方法，其特征在于，还包括确定对背景语音值的估算。

19. 如权利要求 18 所述的方法，其特征在于，确定对背景语音值的估算包括，使用对所述干净语音值的估算来估算所述背景语音值。

20. 如权利要求 11 所述的方法，其特征在于，还包括使用所述信道响应的先验模型来估算所述干净语音值。

多传感器语音增强的方法和装置

技术领域

本发明涉及降噪，尤其涉及从语音信号中去除噪声。

背景技术

在语音识别和语音传输中的一个常见问题是加性噪声对语音信号的破坏。更具体地，由于另一扬声器的语音造成的破坏被证明是难以检测和/或纠正的。

近来，开发了一种试图使用诸如骨导麦克风等备选的传感器和气导麦克风的组合来去除噪声的系统。该系统使用以下三个训练信道来训练：有噪声的备选传感器训练信号、有噪声的气导麦克风训练信号和干净的气导麦克风训练信号。每个信号都被转化至特征域中。有噪声的备选传感器信号和有噪声的气导麦克风信号的特征被组合成表示有噪声的信号的单个向量。干净的气导麦克风信号的特征形成单个干净向量。这些向量而后用来训练有噪声的向量和干净向量间的映射。一旦训练后，该映射被应用于由有噪声的备选传感器测试信号和有噪声的气导麦克风测试信号的组合所形成的有噪声的向量。该映射产生干净信号向量。

当测试信号的噪声条件与训练信号的噪声条件不匹配时，该系统不是最优的，因为该映射是为训练信号的噪声条件而设计的。

发明内容

一种方法和装置使用备选传感器信号和气导麦克风信号来确定对备选传感器的信道响应。该信道响应而后用于使用备选传感器信号的至少一部分来估算干净语音值。

附图说明

图 1 是其中可实现本发明的一个计算环境的框图。

图 2 是其中可实现本发明的另一计算环境的框图。

图 3 是本发明的通用语音处理系统的框图。

图 4 是本发明的一个实施例中增强语音的系统的框图。

图 5 是本发明的一个实施例中增强语音的流程图。

图 6 是本发明的另一实施例中增强语音的流程图。

图 7 是本发明的又一实施例中增强语音的流程图。

具体实施方式

图 1 示出了可在其上实现本发明的合适的计算系统环境 100 的示例。计算环境 100 仅仅是合适的计算环境的一个示例，并不旨在对本发明的使用范围或功能提出任何限制。也不应该把计算环境 100 解释为对在示例性操作环境 100 中示出的任一组件或其组合有任何依赖或要求。

本发明可用众多其它通用或专用计算系统环境或配置来操作。适合在本发明中使用的公知的计算系统、环境和/或配置的示例包括，但不限于，个人计算机、服务器计算机、手持或膝上型设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费者电子产品、网络 PC、小型机、大型机、电话系统、包含上述系统或设备中的任一个的分布式计算机环境等。

本发明可在诸如由计算机执行的程序模块等的计算机可执行指令通用语境下描述。一般而言，程序模块包括例程、程序、对象、组件、数据结构等，它们执行特定任务或实现特定抽象数据类型。本发明也可以在分布式计算环境下实现，其中任务由通过通信网络连接的远程处理设备执行。在分布式计算环境中，程序模块可以位于包括存储器存储设备在内的本地和远程计算机存储介质中。

参考图 1，用于实现本发明的示例性系统 100 包括计算机 110 形式的通用计算设备。计算机 110 的组件包括，但不限于，处理单元 120、系统存储器 130 和将包括系统存储器在内的各种系统组件耦合至处理单元 120 的系统总线 121。系统总线 121 可以是若干类型的总线结构中的任一种，包括存储器总线或存储器控制器、外围总线和使用多种总线体系结构中的任一种的局部总线。作为示例，而非限制，这样的体系结构包括工业标准体系结构（ISA）总线、微信道体系结构（MCA）总线、增强的 ISA（EISA）总线、视频电子技术标准协会（VESA）局部总线和外围部件互连（PCI）总线（也被称为 Mezzanine 总线）。

计算机 110 通常包括各种计算机可读介质。计算机可读介质可以是能够被计算机 110 访问到的任何可用介质，且包括易失性和非易失性介质、可移动和不可移动介质。作为示例，而非限制，计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以任何方法或技术实现的用于存储诸如计算机可读指令、数据结构、程序模块或其它数据等信息的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括，但不限于，RAM、ROM、EEPROM、闪存或其它存储器技术，CD-ROM、数字多功能盘（DVD）或其它光盘存储，磁带盒、磁带、磁盘存储或其它磁性存储设备、或能用于存储所需信息且可以由计算机 100 访问的任何其它介质。通信介质通常具体化为诸如载波或其它传输机制等已调制数据信号中的计算机可读指令、数据结构、程序模块或其它数据，且包括任何信息传递介质。术语“已调制数据信号”指的是一种信号，其一个或多个特征以在信号中编码信息的方式被设定或更改。作为示例，而非限制，通信介质包括有线介质，诸如有线网络或直接线连接，和无线介质，诸如声学、RF、红外线和其它无线介质。上述中任何的组合也应包括在计算机可读介质范围之内。

系统存储器 130 包括易失性或非易失性存储器形式的计算机存储介质，诸如只读存储器(ROM)131 和随机存取存储器(RAM)132。基本输入/输出系统 133(BIOS)包含有助于诸如启动时在计算机 110 中的元件之间传递信息的基本例程，它通常存储在 ROM 131 中。RAM 132 通常包含处理单元 120 可以立即访问和/或目前正在操作的数据和/或程序模块。作为示例，而非限制，图 2 示出了操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

计算机 110 也可以包括其它可移动/不可移动、易失性/非易失性计算机存储介质。仅作为示例，图 1 示出了从不可移动、非易失性磁介质中读取或向其写入的硬盘驱动器 141，从可移动、非易失性磁盘 152 中读取或向其写入的磁盘驱动器 151，和从诸如 CD ROM 或其它光学介质等可移动、非易失性光盘 156 中读取或向其写入的光盘驱动器 155。可以在示例性操作环境下使用的其它可移动/不可移动、易失性/非易失性计算机存储介质包括，但不限于，盒式磁带、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等。硬盘驱动器 141 通常由不可移动存储器接口，诸如接口 140 连接至系统总线 121，磁盘驱动器 151 和光盘驱动器 155 通常由可移动存储器接口，诸如接口 150 连接至系统总线 121。

以上描述和在图 1 中示出的驱动器及其相关联的计算机存储介质为计算机 110 提供了对计算机可读指令、数据结构、程序模块和其它数据的存储。例如，在图 1 中，硬盘驱动器 141 被示为存储操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。注意到这些组件可以与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同或不同。操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 在这里被标注了不同的标号是为了说明至少它们是不同的副本。

用户可以通过输入设备，诸如键盘 162、麦克风 163 和定点设备 161（通常指鼠标、跟踪球或触摸垫）向计算机 110 输入命令和信息。其它输入设备（未示出）可以包括操纵杆、游戏垫、圆盘式卫星天线、扫描仪等。这些和其它输入设备通常由耦合至系统总线的用户输入接口 160 连接至处理单元 120，但也可以由其它接口或总线结构，诸如并行端口、游戏端口或通用串行总线（USB）连接。监视器 191 或其它类型的显示设备也经由一接口，诸如视频接口 190，连接至系统总线 121。除监视器以外，计算机也可以包括其它外围输出设备，诸如扬声器 197 和打印机 196，它们可以通过输出外围接口 195 连接。

计算机 110 可使用至一个或多个远程计算机，诸如远程计算机 180 的逻辑连接在网络化环境下操作。远程计算机 180 可以是个人计算机、手持式设备、服务器、路由器、网络 PC、对等设备或其它常见网络节点，且通常包括上文相对于计算机 110 所描述的许多或所有元件。图 1 中所示逻辑连接包括局域网（LAN）171 和广域网（WAN）173，但也可以包括其它网络。这样的网络环境在办公室、企业范围计算机网络、内联网和因特网中是常见的。

当在 LAN 网络环境中使用时，计算机 110 通过网络接口或适配器 170 连接至 LAN 171。当在 WAN 网络环境中使用时，计算机 110 通常包括调制解调器 172 或用于通过诸如因特网等 WAN 173 建立通信的其它装置。调制解调器 172 可以是内部的或外部的，可以通过用户输入接口 160 或其它合适的机制连接至系统总线 121。在网络化环境中，相对于计算机 110 所描述的程序模块或其部分可以存储在远程存储器存储设备中。作为示例，而非限制，图 1 示出了远程应用程序 185 驻留在存储器设备 181 上。可以理解，所示的网络连接是示例性的，且可以使用在计算机之间建立通信链路的其它手段。

图 2 是移动设备 200 的框图，它是一个示例性计算环境。移动设备 200 包括微处理器 202、存储器 204、输入/输出（I/O）组件 206 和用于同远程计算机或其它移动设备通信的通信接口 208。在一个实施例中，上述组件为相互通信而通过合适的总线 210 被耦合在一起。

存储器 204 被实现为诸如带有电池备用模块(未示出)的随机存取存储器(RAM)等的非易失性电子存储器，以使当移动设备 200 的总电源被关闭时，存储在存储器 204 中的信息也不会丢失。存储器 204 的一部分较佳地被分配为用于程序执行的可寻址存储器，而存储器 204 的另一部分较佳地用于存储，诸如模拟在硬盘驱动器上的存储。

存储器 204 包括操作系统 212、应用程序 214 和对象存储 216。在操作期间，操作系统 212 较佳地由处理器 202 从存储器 204 处执行。在一个较佳的实施例中，操作系统 212 是可从微软公司购买的 WINDOWS® CE 操作系统。操作系统 212 较佳地是为移动设备所设计的，且实现可由应用程序 214 通过一组所展现的应用程序编程接口和方法来使用的数据库特征。对象存储 216 中的对象由应用程序 214 和操作系统 212 至少部分地响应于对所展现的应用程序编程接口和方法的调用来维护。

通信接口 208 表示允许移动设备 200 发送和接收信息的众多设备和技术。仅举几个示例，这些设备包括有线和无线调制解调器、卫星接收器和广播调谐器。移动设备 200 也能够被直接连接至计算机以与其交换数据。在这些情况下，通信接口 208 能够是红外线收发器或者串行或并行通信连接，上述所有都能够传输流信息。

输入/输出组件 206 包括各种输入设备，诸如触敏屏幕、按钮、滚轮和麦克风，还包括各种输出设备，包括音频发生器、振动设备和显示器。以上列出的设备仅作为示例，且不需在移动设备 200 上全部存在。另外，其它输入/输出设备可以在本发明的范围内被附加至移动设备 200 或与其一同出现。

图 3 提供了本发明实施例的基本框图。图 3 中，说话者 300 生成语音信号 302（X），该信号由气导麦克风 304 和备选传感器 306 检测。备选传感器的示例包括测量用户喉部振动的喉部麦克风、位于或接近用户面部骨骼或颅骨（诸如颌骨）的或在用户耳朵中，传感与用户生成的语音相对应的颅骨或颌骨的振动的骨导传感器。气导麦克风 304 是常用于将音频空气波转换成电信号的麦克风类型。

气导麦克风 304 还接收由一个或多个噪声源 310 生成的环境噪声 308（U）和

由背景说话者 314 生成背景语音 312（V）。取决于备选传感器的类型和背景语音的级别，背景语音 312 也可以由备选传感器 306 检测。然而，在本发明的实施例中，备选传感器 306 通常对环境噪声和背景语音不如气导麦克风 304 敏感。这样，由备选传感器 306 生成的备选传感器信号 316（B）一般比由气导麦克风 304 所生成的气导麦克风信号 318（Y）包含更少的噪声。尽管备选传感器 306 对环境噪声较不敏感，但它的确生成某些传感器噪声 320（W）。

从说话者 300 至备选传感器信号 316 的路径能够被建模为拥有信道响应 H 的信道。从背景说话者 314 至备选传感器信号 316 的路径能够被建模为拥有信道响应 G 的信道。

备选传感器信号 316（B）和气导麦克风信号 318（Y）被提供给干净信号估算器 322，它估算干净信号 324，并且在某些实施例中估算背景语音信号 326。干净信号估算 324 被提供给语音处理 328。干净信号估算 324 可以是经滤波的时域信号或傅里叶变换向量。如果干净信号估算 324 是时域信号，则语音处理 328 可以采用收听器、语音编码系统或语音识别系统的形式。如果干净信号估算 324 是傅里叶变换向量，则语音处理 328 通常可以是语音识别系统，或包含傅里叶反变换用于将傅里叶变换向量转换为波形。

在直接滤波增强 322 中，备选传感器信号 316 和麦克风信号 318 被转换到用于估算干净语音的频域。如图 4 所示，备选传感器信号 316 和气导麦克风信号 318 分别被提供给模一数转换器 404 和 414，用于生成一数字值序列，这些数字值分别由帧构造器 406 和 416 组合成值的帧。在一个实施例中，模一数转换器 404 和 414 以 16 kHz 和每个样值 16 比特对模拟信号进行采样，从而创建了每秒 32 千字节的语音数据，且帧构造器 406 和 416 每 10 毫秒分别创建一个包含 20 毫秒数据的新帧。

由帧构造器 406 和 416 提供的每一各自的数据帧分别使用快速傅里叶变换（FFT）408 和 418 转换到频域。

备选传感器信号和气导麦克风信号的频域值被提供给干净信号估算器 420，它使用该频域值来估算干净语音信号 324，并在某些实施例中估算背景语音信号 326。

在某些实施例中，干净语音信号 324 和背景语音信号 326 使用快速傅里叶反变换 422 和 424 转换回时域。这样创建了干净语音信号 324 和背景语音信号 326 的时域形式。

本发明提供了用于估算干净语音信号 324 的直接滤波技术。在直接滤波中，备选传感器 306 的信道响应的最大似然估计是通过最小化与该信道响应相关的函数来确定的。这些估算而后被用来通过最小化与干净语音信号相关的函数来确定干净语音信号的最大似然估计。

在本发明的一个实施例中，与由被选传感器所检测的背景语音相对应的信道响应 G 被认为是零，且背景语音和环境噪声被结合在一起形成单个噪声项。这能够获得在干净语音信号和气导麦克风信号及备选传感器信号之间的模型：

$$y(t) = x(t) + z(t) \qquad \text{公式 1}$$

$$b(t) = h(t) * x(t) + w(t) \qquad \text{公式 2}$$

其中，$y(t)$是气导麦克风信号，$b(t)$是备选传感器信号，$x(t)$是干净语音信号，$z(t)$是包括背景语音和环境噪声的组合噪声信号，$w(t)$是备选传感器噪声，$h(t)$是对与备选传感器相关联的干净语音信号的信道响应。从而，在公式 2 中，备选传感器信号被建模为干净语音的经滤波形式，其中滤波器拥有脉冲响应$h(t)$。

在频域中，公式 1 和公式 2 可以被表达成：

$$Y_t(k) = X_t(k) + Z_t(k) \qquad \text{公式 3}$$

$$B_t(k) = H_t(k)X_t(k) + W_t(k) \qquad \text{公式 4}$$

其中，记法$Y_t(k)$表示以时间 t 为中心的一个信号帧的第 k 个频率分量。这个记法适用于$X_t(k)$，$Z_t(k)$，$H_t(k)$，$W_t(k)$和$B_t(k)$。在以下描述中，对频率分量 k 的引用为清楚起见而被省略。但是，本领域的技术人员应该认识到，下文执行的计算是在每个频率分量的基础上执行的。

在该实施例中，噪声Z_t和W_t的实部和虚部被建模为独立的零均值高斯型，使得：

$$Z_t = N(O, \sigma_z^2) \qquad \text{公式 5}$$

$$W_t = N(O, \sigma_w^2) \qquad \text{公式 6}$$

其中，σ_z^2是噪声Z_t的方差，σ_w^2是噪声W_t的方差。

H_t也被建模为高斯型，使得：

$$H_t = N(H_0, \sigma_H^2) \qquad \text{公式 7}$$

其中，H_0是信道响应的均值，σ_H^2是信道响应的方差。

给定这些模型参数，干净语音值X_t和信道响应值H_t的概率由条件概率描述：

$$p(X_t,H_t|Y_t,B_t,H_0\sigma_z^2,\sigma_w^2,\sigma_H^2) \quad \text{公式 8}$$

它与下述成比例：

$$p(Y_t,B_t|X_t,H_t,\sigma_z^2\sigma_w^2)p(H_t|H_0,\sigma_H^2)p(X_t) \quad \text{公式 9}$$

它等价于：

$$p(Y_t|X_t,\sigma_z^2)p(B_t|X_t,H_t,\sigma_w^2)p(H_t|H_0,\sigma_H^2)p(X_t) \quad \text{公式 10}$$

在一个实施例中，信道响应的先验概率 $p(H_t|H_0,\sigma_H^2)$，和干净语音信号的先验概率 $p(X_t)$ 被忽略，且剩下的概率被作为高斯分布处理。使用这些简化，公式 10 变为：

$$\frac{1}{(2\pi)^2\sigma_z^2\sigma_w^2}\exp[-\frac{1}{2\sigma_z^2}|Y_t-X_t|^2-\frac{1}{2\sigma_w^2}|B_t-H_tX_t|^2] \quad \text{公式 11}$$

从而，话语的最大似然估计 H_t,X_t 是通过把公式 11 在该话语中的所有时间帧 T 上的指数项最小化来确定的。这样，该最大似然估计通过最小化以下公式来给出：

$$F=\sum_{t=1}^{T}(\frac{1}{2\sigma_z^2}|Y_t-X_t|^2+\frac{1}{2\sigma_w^2}|B_t-H_tX_t|^2) \quad \text{公式 12}$$

因为公式 12 是相对于两个变量 H_t,X_t 来最小化的，因此相对于每个变量的偏导可以被用来确定使该函数最小化的变量的值。特别地，$\frac{\partial F}{\partial X_t}=0$ 时可以得到：

$$X_t=\frac{1}{\sigma_w^2+\sigma_z^2|H_t|^2}(\sigma_w^2Y_t+\sigma_z^2H_t^*B_t) \quad \text{公式 13}$$

其中，H_t^* 表示 H_t 的复共轭，而 $|H_t|$ 表示 H_t 的复值的幅度。

将 X_t 的该值代入公式 12，令偏导 $\frac{\partial F}{\partial H_t}=0$，且然后假定 H 在所有时间帧 T 上是常数，得到 H 的解：

$$H=\frac{\sum_{t=1}^{T}(\sigma_z^2|B_t|^2-\sigma_w^2|Y_t|^2)\pm\sqrt{(\sum_{t=1}^{T}(\sigma_z^2|B_t|^2-\sigma_w^2|Y_t|^2))^2+4\sigma_z^2\sigma_w^2|\sum_{t=1}^{T}B_t^*Y_t|^2}}{2\sigma_z^2\sum_{t=1}^{T}B_t^*Y_t} \quad \text{公式 14}$$

在公式 14 中，对 H 的估算需要对最后 T 帧的多个求和，其形式为：

$$S(T)=\sum_{t=1}^{T}s_t \quad \text{公式 15}$$

其中，s_t 为 $(\sigma_z^2|B_t|^2-\sigma_w^2|Y_t|^2)$ 或 $B_t^*Y_t$。

由上述公式，第一帧(t = 1)与最后一帧(t = T)同样重要。然而，在其它实施例中，较佳的是在对 H 的估算中让最近的帧比较早的帧起更大的作用。为达到该目的的一种技术是“指数衰退（exponential aging）”，其中公式 15 中的求和被替代

为：

$$S(T)=\sum_{t=1}^{T}c^{T-t}s_t \qquad \text{公式 16}$$

其中，$c\leq 1$。如果 c = 1，那么公式 16 等价于公式 15。如果 c < 1，那么最后一帧的权重为 1，最后一帧的前一帧由 c 加权（即，它起的作用比最后一帧小），且第一帧由 c^{T-1} 加权（即，它起的作用远小于最后一帧）。举一个例子。令 c = 0.99 且 T = 100，那么第一帧的权重仅为 $0.99^{99}=0.37$。

在一个实施例中，公式 16 被递归地估算为：

$$S(T)=cS'(T-1)+s_T \qquad \text{公式 17}$$

因为公式 17 自动地给旧的数据分配更小的权重，因此不需要使用固定窗长度，且最后 T 帧的数据不需存储在存储器中。相反，只有前一帧的 S(T-1)的值需要被存储。

使用公式 17，公式 14 变为：

$$H_T=\frac{J(T)\pm\sqrt{(J(T))^2+4\sigma_z^2\sigma_w^2\,|K(T)|^2}}{2\sigma_z^2K(T)} \qquad \text{公式 18}$$

其中：

$$J(T)=cJ(T-1)+(\sigma_z^2\,|B_T|^2-\sigma_w^2\,|Y_T|^2) \qquad \text{公式 19}$$

$$K(T)=cK(T-1)+B_T^*Y_T \qquad \text{公式 20}$$

公式 19 和 20 中的 c 的值为用于计算 J(T)和 K(T)当前值的过去的帧的数目提供了有效长度。特别地，有效长度由以下公式给出：

$$L(T)=\sum_{t=1}^{T}c^{T-t}=\sum_{i=0}^{T-1}c^i=\frac{1-c^T}{1-c} \qquad \text{公式 21}$$

渐近的有效长度为：

$$L=\lim_{T\to\infty}L(T)=\frac{1}{1-c} \qquad \text{公式 22}$$

或等价地，

$$c=\frac{L-1}{L} \qquad \text{公式 23}$$

这样，使用公式 23，c 能够被设置以便在公式 18 中得到不同的有效长度。例如，为得到 200 帧的有效长度，c 被设为：

$$c=\frac{199}{200}=0.995 \qquad \text{公式 24}$$

一旦使用公式 14 估算了 H，它可以被用于代替公式 13 中所有的 H_t，以便确

定在每个时间帧 t 时 X_t 的单独值。可选地，公式 18 可以用于估算在每个时间帧 t 时的 H_t。在每个时间帧时的 H_t 的值而后被用在公式 13 中来确定 X_t。

图 5 提供了本发明的一方法的流程图，它使用公式 13 和 14 来估算话语的干净语音值。

在步骤 500 处，气导麦克风信号和备选传感器信号的帧的频率分量在整段话语上捕捉。

在步骤 502 处，气导麦克风噪声的方差 σ_z^2 和被选传感器噪声的方差 σ_w^2 分别从气导麦克风信号和备选传感器信号的帧确定，这些帧在早先说话者不发声的时段的话语中捕捉。

因为备选传感器噪声的能量比由备选传感器信号捕捉到的语音信号的能量小得多，因此本方法通过识别备选传感器信号的低能量段来确定说话者何时不发声。在其它实施例中，已知的语音检测技术可以应用于气导语音信号，以识别说话者何时发声。在说话者被认为不在发声时，X_t 被假定为零，且来自气导麦克风或备选传感器的任何信号被认为是噪声。这些噪声值的样本从非语音的帧中收集，且用于估算在气导麦克风信号和备选传感器信号中的噪声的方差。

在步骤 504 处，通过使用上述公式 14，使用在话语的所有帧上的备选传感器信号和气导麦克风信号的值来确定 H 的值。在步骤 506 处，使用上述公式 13，该 H 的值与每一时间帧上的个别气导麦克风信号和备选传感器信号的值一起用来确定每一时间帧上的增强的或经降噪的语音值。

在其它实施例中，使用公式 18 为每一帧确定 H_t，而不是使用公式 14 使用话语中的所有帧来确定单个 H 值。然后使用上述公式 13，使用 H_t 的值来计算该帧的 X_t。

在本发明的第二实施例中，备选传感器对背景语音的信道响应被认为是非零的。在该实施例中，气导麦克风信号和备选传感器信号被建模为：

$$Y_t(k) = X_t(k) + V_t(k) + U_t(k) \qquad \text{公式 25}$$

$$B_t(k) = H_t(k)X_t(k) + G_t(k)V_t(k) + W_t(k) \qquad \text{公式 26}$$

其中，噪声 $Z_t(k)$ 被分成背景语音 $V_t(k)$ 和环境噪声 $U_t(k)$，且对背景语音的备选传感器信道响应是非零值 $G_t(k)$。

在该实施例中，对干净语音 X_t 的先验知识仍旧被忽略。作以下假定，干净语

音 X_t 的最大似然性能够通过最小化下述目标函数来找到：

$$F=\frac{1}{\sigma_w^2}|B_t-H_tX_t-G_tV_t|^2+\frac{1}{\sigma_u^2}|Y_t-X_t-V_t|^2+\frac{1}{\sigma_v^2}|V_t|^2 \qquad \text{公式 27}$$

这就得到以下干净语音的公式：

$$X_t=\frac{(\sigma_w^2+\sigma_u^2H_t^*G_t)Y_t+[(\sigma_u^2+\sigma_v^2)H_t^*-\sigma_v^2G_t^*](B_t-G_tY_t)}{\sigma_v^2|H_t-G_t|^2+\sigma_w^2+\sigma_u^2|H_t|^2} \qquad \text{公式 28}$$

为了解出公式 28，方差 σ_w^2,σ_u^2 和 σ_v^2 以及信道响应值 H_t 和 G_t 必须已知。图 6 提供了用于识别这些值和用于确定每一帧的增强的语音值的流程图。

在步骤 600 处，话语的帧在用户不发声和没有背景语音的时候被识别。这些帧而后用于分别确定备选传感器和气导麦克风的方差 σ_w^2 和 σ_u^2。

为识别用户不发声时的那些帧，可检查备选传感器信号。因为备选传感器信号为背景语音产生的信号值远小于为噪声产生的信号值，则如果备选传感器信号的能量较低，可以假定说话者不在发声。在基于备选信号识别的帧中，能够向气导麦克风信号应用语音检测算法。该语音检测系统可以检测当用户不发声时在气导麦克风信号中是否存在背景语音。这样的语音检测算法在本领域中是公知的，且包括诸如音调跟踪系统等系统。

当确定了与气导麦克风和备选传感器相关联的噪声的方差后，图 6 所示的方法继续前进至步骤 602 处，在这里识别出用户不发声但是存在背景语音的帧。这些帧使用上述相同的技术来识别，但只是选择当用户不发声时包含背景语音的那些帧。对用户不发声时包含背景语音的那些帧，可以假定背景语音远大于环境噪声。由此，在那些帧期间气导麦克风信号的任何方差被认为是由背景语音引起的。结果，方差 σ_v^2 能够由从用户不发声但存在背景语音的那些帧期间气导麦克风信号的值来直接设定。

在步骤 604 处，所识别的用户不发声但存在背景语音的帧用于估算背景语音的备选传感器信道响应 G。具体地，G 被确定为：

$$G=\frac{\sum_{t=1}^{D}(\sigma_u^2|B_t|^2-\sigma_w^2|Y_t|^2)\pm\sqrt{(\sum_{t=1}^{D}(\sigma_u^2|B_t|^2-\sigma_w^2|Y_t|^2))^2+4\sigma_u^2\sigma_w^2|\sum_{t=1}^{D}B_t^*Y_t|^2}}{2\sigma_u^2\sum_{t=1}^{D}B_t^*Y_t} \qquad \text{公式 29}$$

其中，D 是用户不发声但存在背景语音的帧的数目。在公式 29 中，可以假定 G 在话语的所有帧上保持不变，从而不再依赖于时间帧 t。

在步骤 606 处，对背景语音的备选传感器信道响应 G 的值用于确定对干净语

音信号的备选传感器信道响应。具体地，H 如下计算：

$$H = G + \frac{\sum_{t=1}^{T}(\sigma_v^2 |B_t - GY_t|^2 - \sigma_w^2 |Y_t|^2) \pm \sqrt{(\sum_{t=1}^{T}(\sigma_v^2 |B_t - GY_t|^2 - \sigma_w^2 |Y_t|^2))^2 + 4\sigma_v^2\sigma_w^2 |\sum_{t=1}^{T}(B_t - GY_t)^* Y_t|^2}}{2\sigma_v^2 \sum_{t=1}^{T}(B_t - GY_t)^* Y_t}$$

公式 30

在公式 30 中，在 T 上的求和可以用上文结合公式 15－24 讨论的递归指数衰减计算来代替。

当在步骤 606 处确定 H 之后，公式 28 可以用来确定所有帧的干净语音值。在使用公式 28 时，H_t 和 G_t 分别用独立值 H 和 G 代替。另外，在某些实施例中，公式 28 中的 $B_t - GY_t$ 项用 $(1 - \frac{|GY_t|}{|B_t|})B_t$ 来代替，因为发现难以准确地确定背景语音及其对备选传感器的泄漏之间的相位差。

如果在公式 30 中使用该递归指数衰减计算来代替求和，则可以对每一时间帧确定单独的 H_t 值，且可将该值用作公式 28 中的 H_t。

在上述实施例的进一步扩展中，有可能提供对每一时间帧上的背景语音信号的估算。具体地，一旦确定了干净语音值，每一时间帧上的背景语音值可以被确定为：

$$V_t = \frac{1}{\sigma_w^2 + H^* G_u^2}[\sigma_w^2 Y_t + \sigma_u^2 H^* B_t - (\sigma_w^2 + |H|^2 \sigma_u^2) X_t] \qquad \text{公式 31}$$

该可任选步骤在图 6 中的步骤 610 处示出。

在上述实施例中，备选传感器对干净信号的信道响应的先验知识被忽略。在其它实施例中，如果提供了该先验知识，则它能够用来生成对每一时间帧 H_t 上的信道响应的估算，并用来确定干净语音值 X_t。

在该实施例中，对背景语音噪声的信道响应再次被假定为零。从而，气导信号和备选传感器信号的模型与在上述公式 3 和 4 中所示的模型相同。

用于估算每一时间帧上的干净语音值和信道响应 H_t 的公式通过最小化以下目标函数来确定：

$$-\frac{1}{2\sigma_z^2}|Y_t - X_t|^2 - \frac{1}{2\sigma_w^2}|B_t - H_t X_t|^2 - \frac{1}{2\sigma_H^2}|H_t - H_0|^2 \qquad \text{公式 32}$$

通过独立地对 X_t 和 H_t 两个变量取偏导并令结果等于零，该目标函数对于 X_t 和 H_t 被最小化。这提供了下述 X_t 和 H_t 的公式：

$$X_t = \frac{1}{\sigma_w^2 + \sigma_v^2 |H_t|^2}(\sigma_w^2 Y_t + \sigma_v^2 H_t^* B_t) \quad \text{公式 33}$$

$$H_t = \frac{1}{\sigma_w^2 + \sigma_H^2 |X_t|^2}(\sigma_H^2 B_t X_t^* + \sigma_w^2 H_0) \quad \text{公式 34}$$

其中，H_0和σ_H^2分别是备选传感器对干净语音信号的信道响应的先验模型的均值和方差。因为X_t的公式包含H_t，而H_t的公式包含变量X_t，因此公式 33 和 34 必须使用迭代的方式解出。图 7 提供了实现这样一种迭代的流程图。

在图 7 的步骤 700 处，确定信道响应的先验模型的参数。在步骤 702 处，确定对X_t的估算。该估算能够通过使用上述忽略信道响应的先验模型的早先的任何实施例来确定。在步骤 704 处，先验模型的参数和对X_t的初始估算用于使用公式 34 来确定H_t。H_t而后在步骤 706 处用于使用公式 33 更新干净语音值。在步骤 708 处，该过程确定是否需要更多的迭代。如果需要更多的迭代，则该过程回到步骤 704 处，并使用在步骤 706 处确定的所更新的X_t值来更新H_t值。重复步骤 704 和 706，直到在步骤 708 处不需要更多的迭代，此时该过程在步骤 710 处结束。

尽管本发明是参考具体实施例而描述的，然而本领域的技术人员可以认识到，可以在形式和细节上进行修改而不背离本发明的精神和范围。

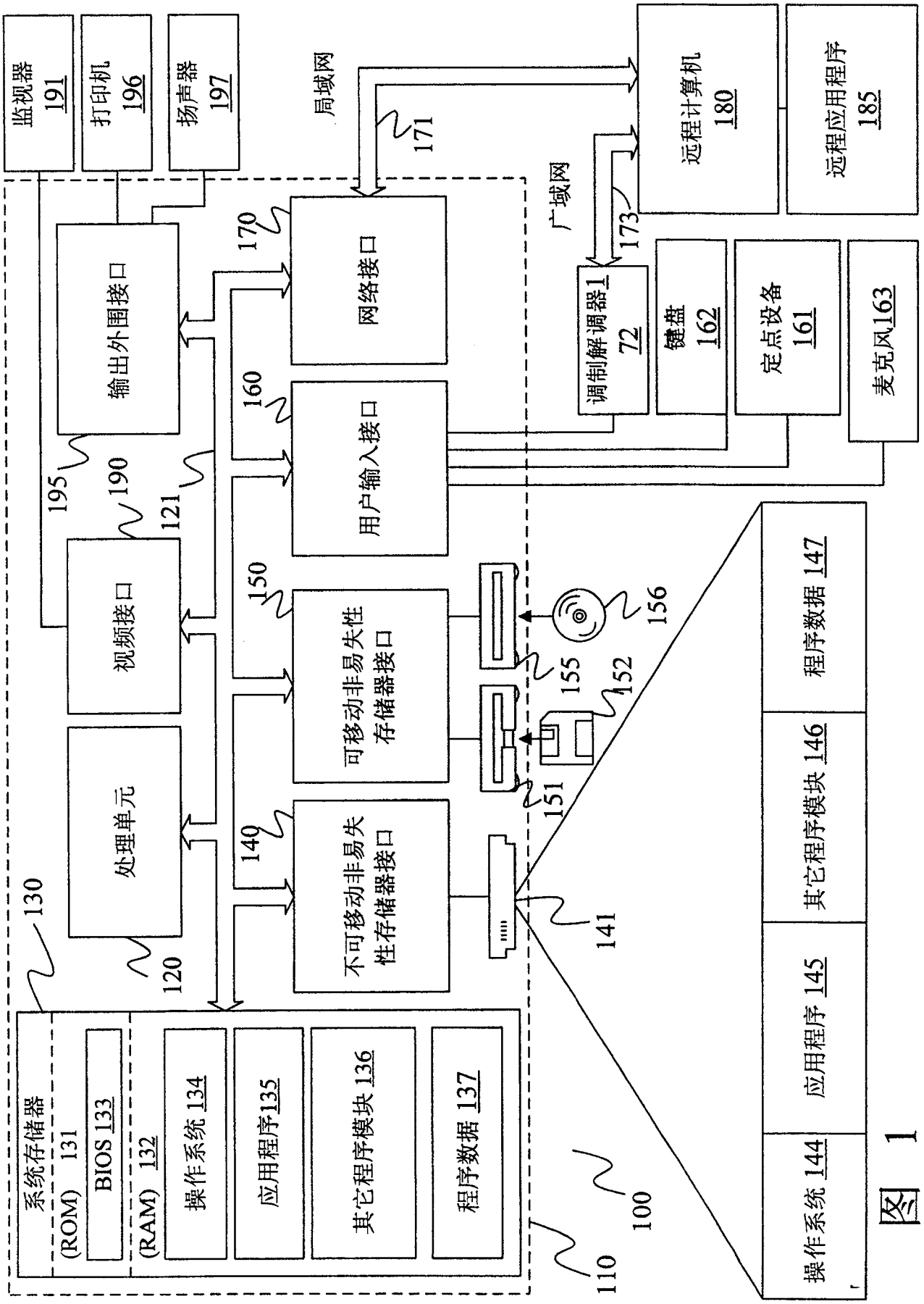

监视器 191
打印机 196
扬声器 197
局域网
171
远程计算机 180
远程应用程序 185
广域网
173
输出外围接口
网络接口
170
调制解调器1 72
键盘 162
定点设备 161
麦克风163
160
用户输入接口
195
190
121
视频接口
150
可移动非易失性存储器接口
处理单元
140
不可移动非易失性存储器接口
155
152
156
151
141
130
120
系统存储器
(ROM) 131
BIOS 133
(RAM) 132
操作系统 134
应用程序135
其它程序模块 136
程序数据 137
110
100
程序数据 147
其它程序模块 146
应用程序 145
操作系统 144

图 1

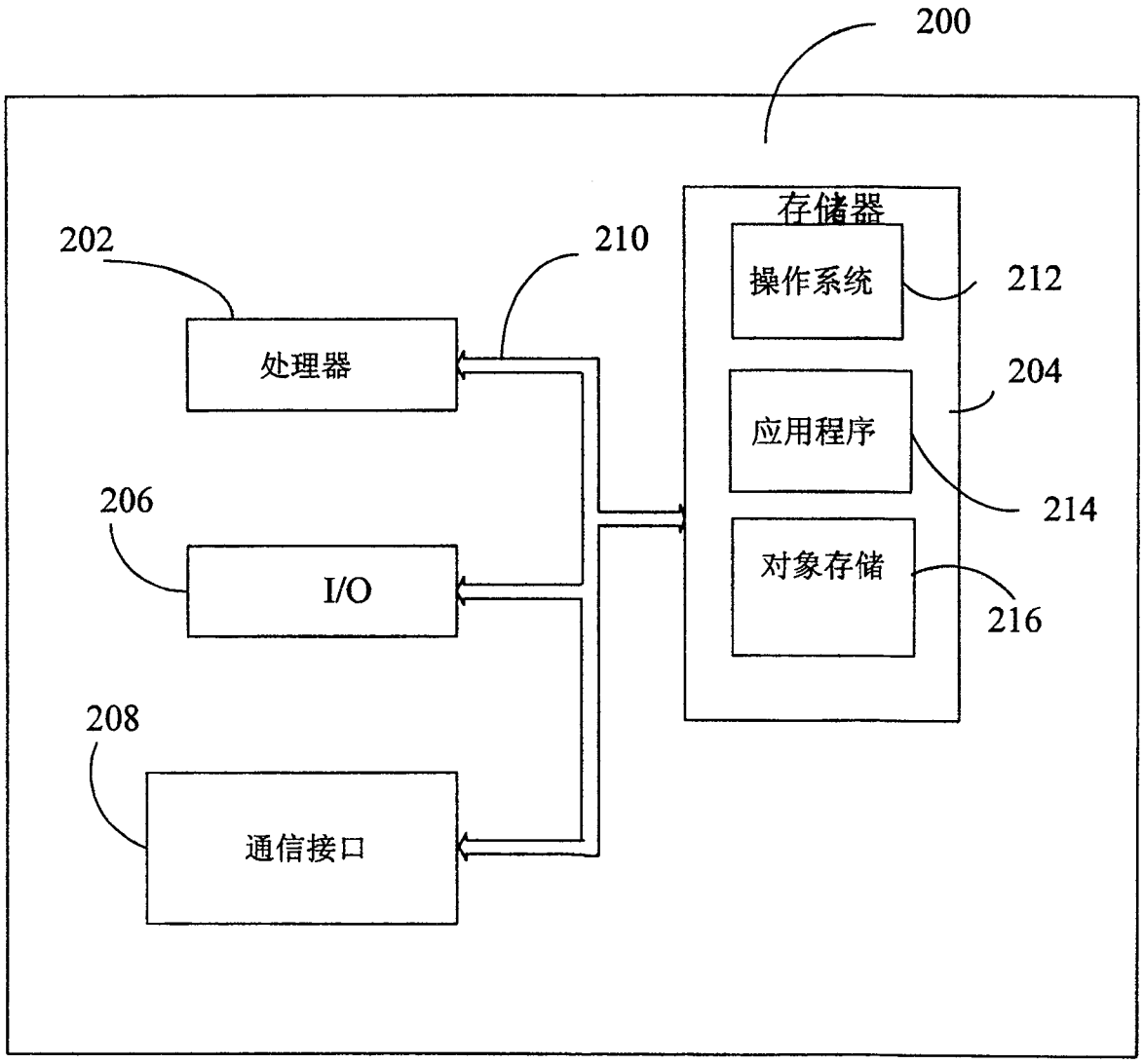

200
存储器
202
210
操作系统
212
处理器
204
应用程序
214
206
I/O
对象存储
216
208
通信接口

图 2

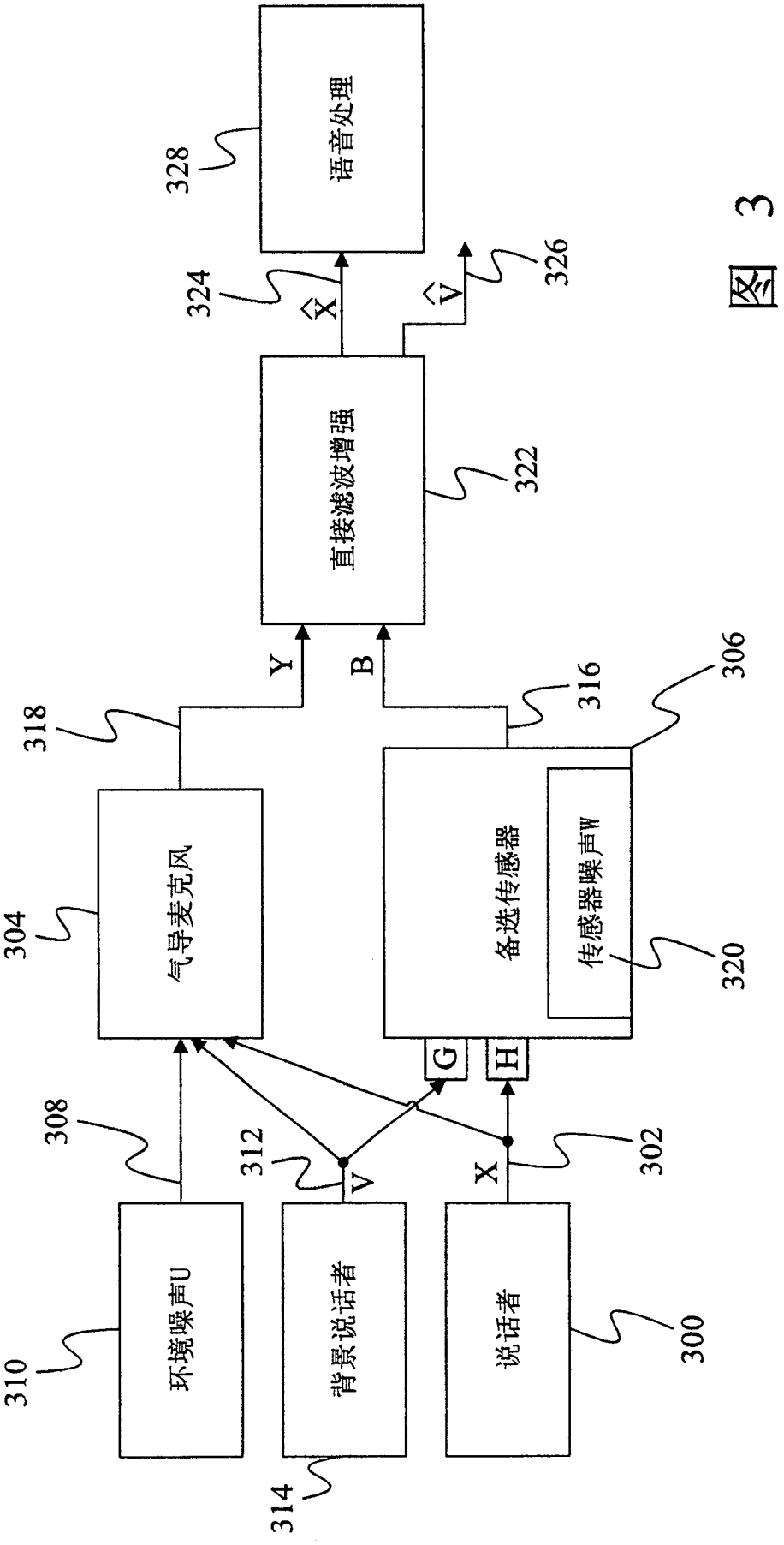

语音处理
328
324
326
直接滤波增强
322
Y
B
318
316
306
气导麦克风
304
备选传感器
传感器噪声W
320
G
H
308
312
302
V
X
环境噪声U
背景说话者
说话者
310
314
300

图 3

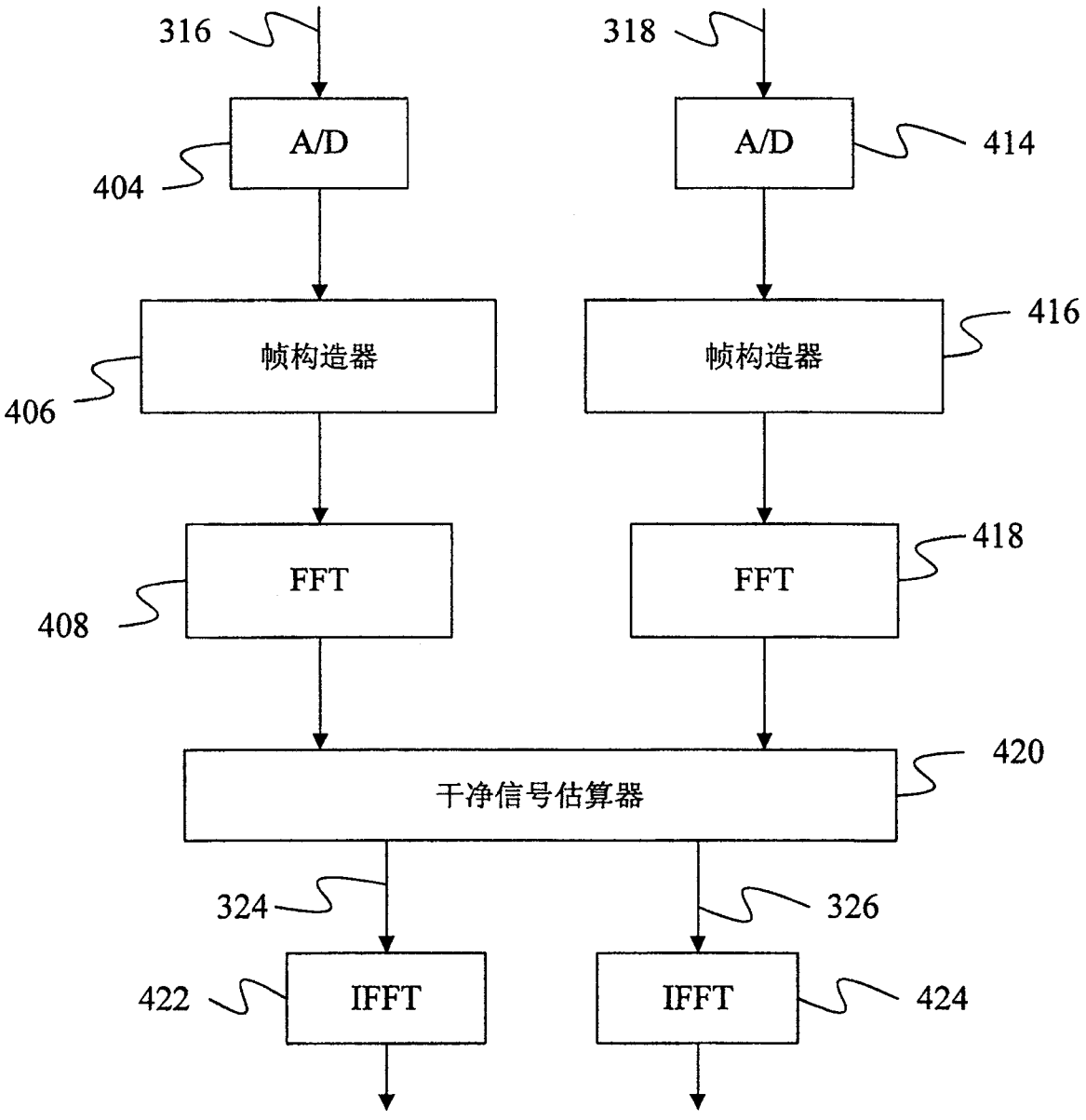

316
318
A/D
A/D
404
414
帧构造器
帧构造器
406
416
FFT
FFT
408
418
干净信号估算器
420
324
326
IFFT
IFFT
422
424

图 4

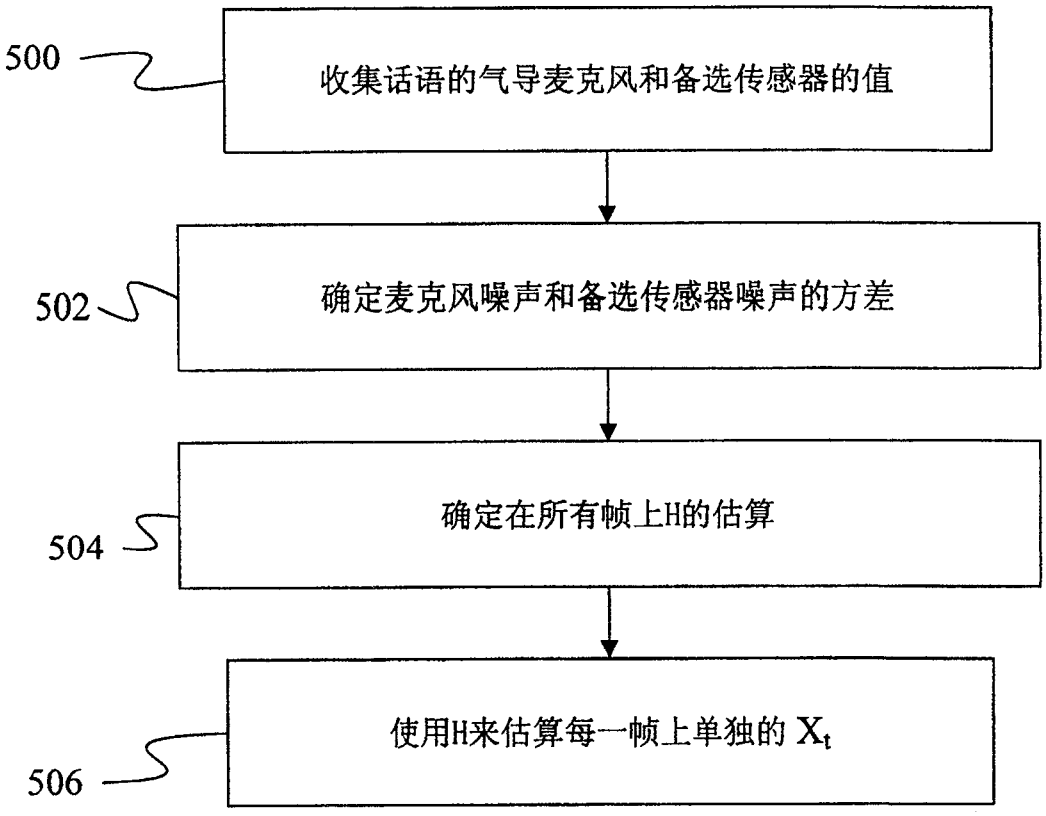

500
收集话语的气导麦克风和备选传感器的值
502
确定麦克风噪声和备选传感器噪声的方差
504
确定在所有帧上H的估算
506
使用H来估算每一帧上单独的 X_t

图 5

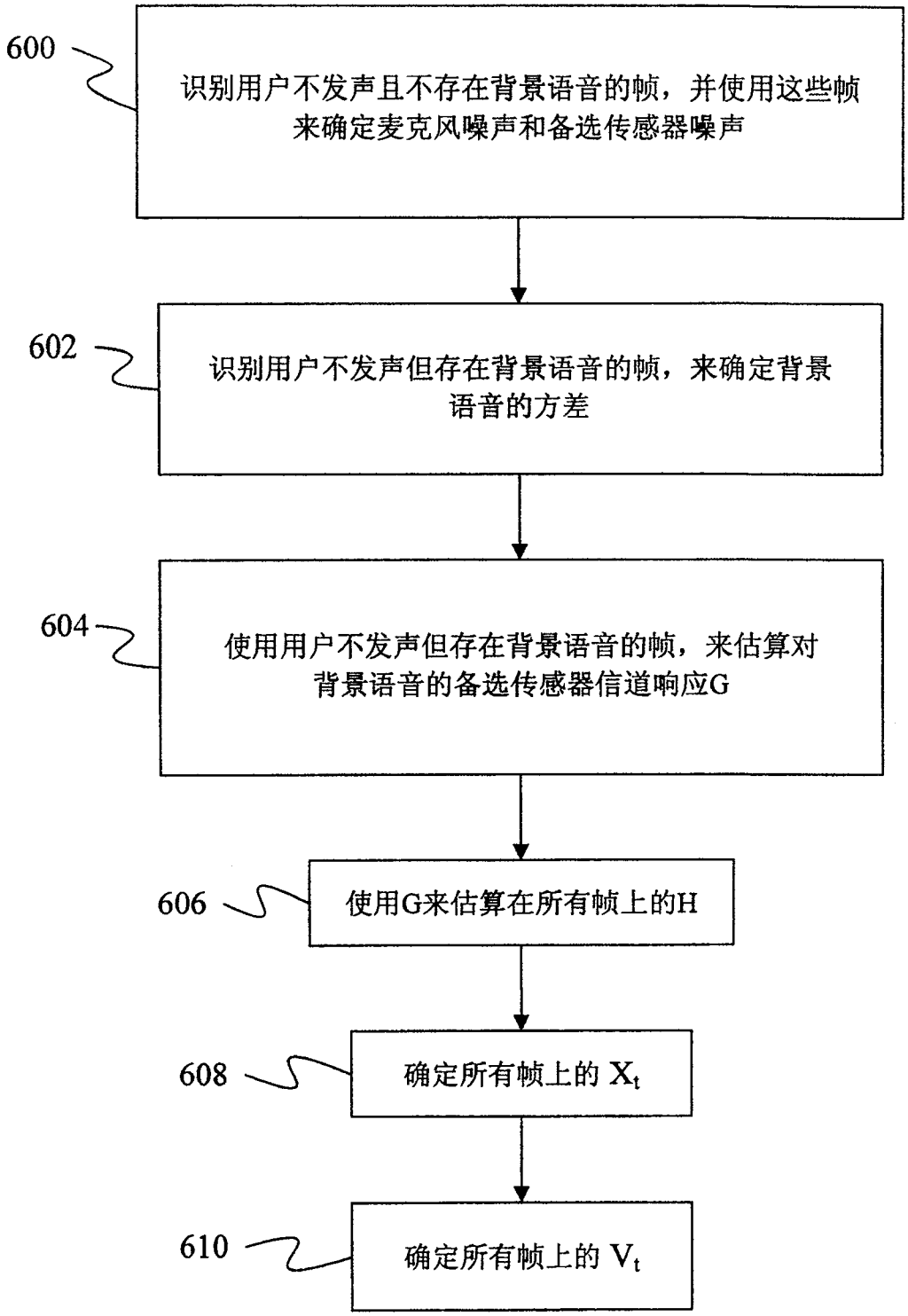

600
识别用户不发声且不存在背景语音的帧，并使用这些帧来确定麦克风噪声和备选传感器噪声
602
识别用户不发声但存在背景语音的帧，来确定背景语音的方差
604
使用用户不发声但存在背景语音的帧，来估算对背景语音的备选传感器信道响应G
606
使用G来估算在所有帧上的H
608
确定所有帧上的 X_t
610
确定所有帧上的 V_t

图 6

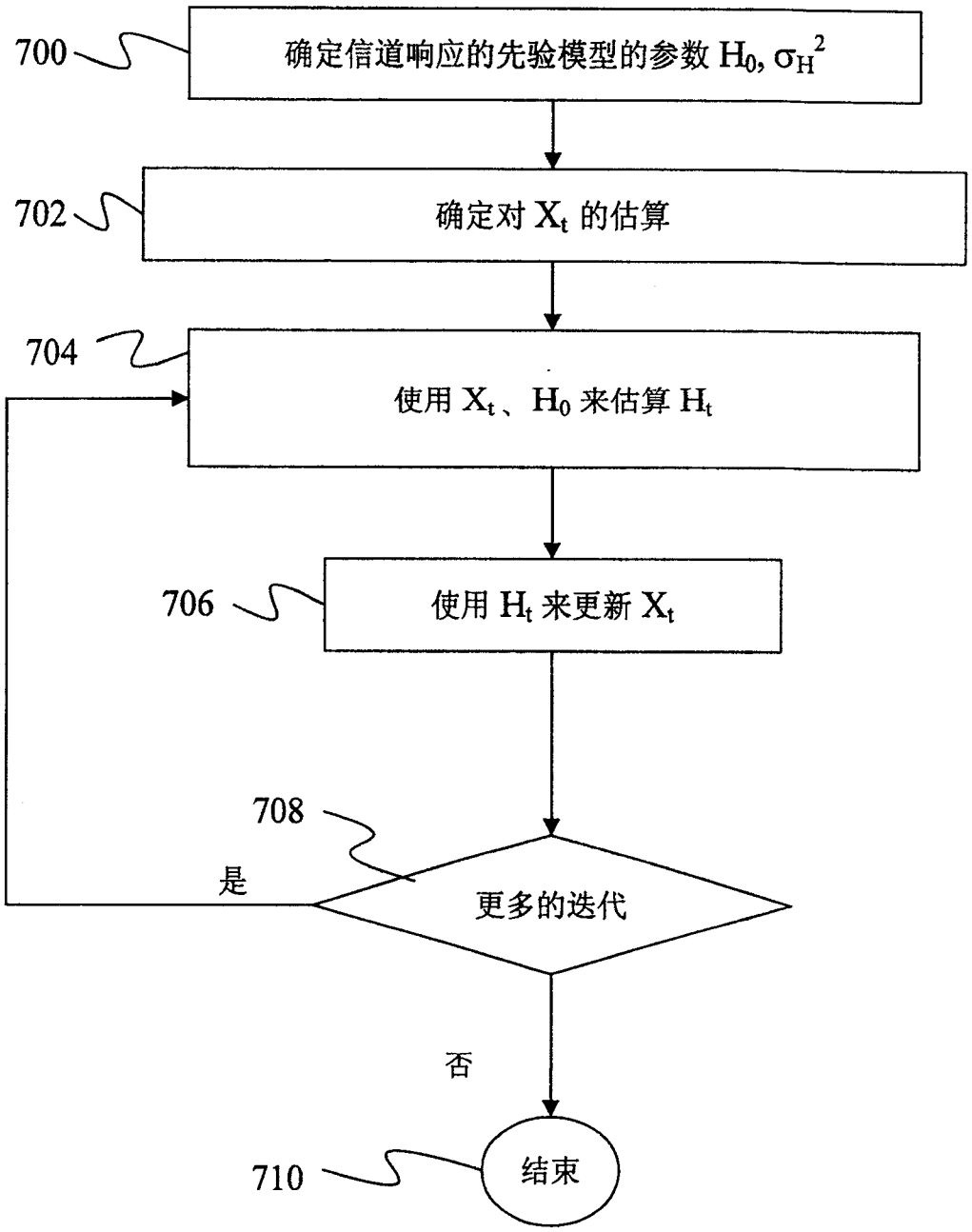

700
确定信道响应的先验模型的参数 H_0, σ_H^2
702
确定对 X_t 的估算
704
使用 X_t、H_0 来估算 H_t
706
使用 H_t 来更新 X_t
708
是
更多的迭代
否
710
结束

图 7