[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
*H04M 1/19* (2006.01)
*H04R 1/00* (2006.01)

[12] 发明专利说明书

专利号 ZL 200510052873.8

[45] 授权公告日 2010 年 1 月 20 日

[11] 授权公告号 CN 100583909C

[22] 申请日 2005.2.24
[21] 申请号 200510052873.8
[30] 优先权
[32] 2004.2.24 [33] US [31] 10/785,768
[73] 专利权人 微软公司
地址 美国华盛顿州
[72] 发明人 M·J·辛克莱尔 黄学东 张正友
[56] 参考文献
US6560468B1 2003.5.6
CN1141548A 1997.1.29
WO2004012477A2 2004.2.5
审查员 贾 杰

[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

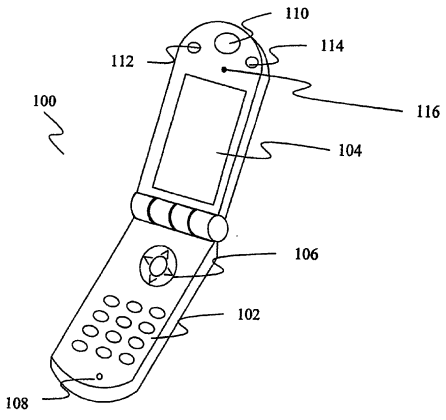
权利要求书 2 页 说明书 17 页 附图 13 页

[54] 发明名称

移动设备上多传感语音增强的装置

[57] 摘要

提供了一种移动设备，它包括一可由用户的手指或大拇指操纵的数字输入、一气导麦克风和一提供指示语音的备选传感器信号的备选传感器。 在某些实施例中，该移动设备也包括一邻近传感器，它提供指示从移动设备到对象的距离的邻近信号。 在某些实施例中，来自气导麦克风的信号、备选传感器信号和邻近信号用于形成干净语音值的估算。 在另外的实施例中，基于干净语音值中的噪声的量产生通过移动设备中的扬声器的声音。 在其它实施例中，通过扬声器产生的信号基于邻近传感器信号。



1．一种移动手持式设备，其特征在于，包括：

一气导麦克风，它将声波转换成表示语音帧的电子麦克风信号；

两个骨导传感器，提供指示所述语音帧的第一和第二电子备选传感器信号，所述两个骨导传感器设置在所述移动手持式设备的扬声器的两侧；以及

一处理器，它使用所述麦克风信号和所述第一和第二备选电子传感器信号中强度较强的一个备选电子传感器信号来估算所述语音帧的干净语音值。

2．如权利要求 1 所述的移动手持式设备，其特征在于，它还包括一选择单元，它选择所述第一备选电子传感器信号和所述第二备选电子传感器信号之一。

3．如权利要求 2 所述的移动手持式设备，其特征在于，所述选择单元基于所述第一备选电子传感器信号和所述第二备选电子传感器信号的幅度，选择所述第一备选电子传感器信号和所述第二备选电子传感器信号之一。

4．如权利要求 1 所述的移动手持式设备，其特征在于，所述扬声器基于所述干净语音值中的噪声的量生成声音。

5．如权利要求 1 所述的移动手持式设备，其特征在于，它还包括一邻近传感器，它产生指示所述移动手持式设备和对象之间的距离的邻近信号。

6．如权利要求 5 所述的移动手持式设备，其特征在于，所述处理器基于所述麦克风信号、所述第一和第二备选电子传感器信号中强度较强的一个备选电子传感器信号和所述邻近信号确定所述干净语音值。

7．如权利要求 5 所述的移动手持式设备，其特征在于，所述扬声器基于所述邻近信号生成声音。

8．如权利要求 1 所述的移动手持式设备，其特征在于，所述两个骨导传感器之一包括一压力转换器，它液压地耦合至一用介质填充的衬垫。

9．如权利要求 8 所述的移动手持式设备，其特征在于，所述移动手持式设备具有左侧和所述左侧对面的右侧，并且其中，所述衬垫具有所述左侧上的第一部分和所述右侧上的第二部分。

10．如权利要求 8 所述的移动手持式设备，其特征在于，所述两个骨导传感器之一还提供一邻近信号。

11．如权利要求 10 所述的移动手持式设备，其特征在于，所述邻近信号包括

由所述压力转换器产生的电信号的直流分量。

12．如权利要求 11 所述的移动手持式设备，其特征在于，所述第一和第二备选电子传感器信号包括由所述压力转换器产生的电信号的交流分量。

# 移动设备上多传感语音增强的装置

## 技术领域

本发明涉及降噪，尤其涉及从由移动手持式设备接收的语音信号中移除噪声。

## 背景技术

诸如移动电话和个人数字助理等提供电话功能或接受语音输入的移动手持式设备通常在诸如繁忙的街道、餐馆、机场和汽车等不利的噪声环境中使用。这些环境中强大的环境噪声使用户的语音变得模糊，并且很难理解一个人在说什么。

尽管开发了试图基于噪声模型移除噪声的噪声滤除系统，然而这些系统尚不能移除所有的噪声。具体地，许多这样的系统发现很难移除在背景中包括其它人的说话的噪声。其一个原因是这些系统及其难以（如果不是不可能的话）确定由麦克风接收的语音信号是来自除使用该移动设备的人之外的其他人。

对于电话头戴式耳机，它通过环绕在用户头部或耳朵的周围定位在用户的头部，开发了通过依赖于头戴式耳机中的附加类型的传感器来提供更健壮的噪声滤除系统。在一个示例中，一骨导传感器被放置在头戴式耳机的一端，并由头戴式耳机的弹力挤压到与覆盖用户头盖骨、耳朵或下颚骨的皮肤接触。该骨导传感器检测头盖骨、耳朵或下颚骨中在用户说话时引起的振动。使用来自骨导传感器的信号，该系统能够更好地识别用户何时在说话，并且结果能够更好地滤除语音信号中的噪声。

尽管这一系统对头戴式耳机能够起较好的作用，其中，骨导传感器和用户之间的接触由头戴式耳机的机械设计来维护，然而这些系统不能直接用于手持式移动设备，因为用户很难将骨导传感器维持在正确的位置，并且这些系统未考虑骨导传感器可能无法保持在正确的位置。

## 发明内容

提供了一种移动设备，包括可由用户的手指或大拇指操纵的数字输入，以及一气导麦克风和提供指示语音的备选传感器信号的备选传感器。在某些实施例中，

移动设备也包括一邻近传感器，它提供指示从移动设备到对象的邻近性的信号。在某些实施例中，来自气导麦克风的信号、备选传感器信号以及邻近信号用于形成对干净语音值的估算。在另外的实施例中，基于干净信号值中的噪声量，通过移动设备中的扬声器产生声音。在其它实施例中，通过扬声器产生的声音基于邻近传感器信号。

附图说明

图 1 是本发明的一个实施例的透视图。

图 2 在用户头部的左侧的位置上示出了图 1 的电话。

图 3 在用户头部的右侧的位置上示出了图 1 的电话。

图 4 是骨导麦克风的框图。

图 5 是本发明的一个替换实施例的透视图。

图 6 是本发明的一个实施例中的备选骨导麦克风的横截面。

图 7 是本发明的一个实施例中的移动设备的框图。

图 8 是本发明的通用语音处理系统的框图。

图 9 是本发明的一个实施例中用于训练降噪参数的框图。

图 10 是使用图 9 的系统训练降噪参数的流程图。

图 11 是本发明的一个实施例中从噪声测试语音信号中标识出干净语音信号的估算的系统的框图。

图 12 是使用图 11 的系统标识干净语音信号的估算的方法的流程图。

图 13 是标识干净语音信号的估算的替换系统的框图。

图 14 是标识干净语音信号的估算的第二替换系统的框图。

图 15 是使用图 14 的系统标识干净语音信号估算的方法的流程图。

图 16 是本发明的移动设备的另一实施例的透视图。

具体实施方式

本发明的实施例提供了一种手持式移动设备，它包含可用于语音检测和噪声滤除的气导麦克风以及备选传感器。图 1 提供了一个示例实施例，其中，手持式移动设备是移动电话 100。移动电话 100 包括键区 102、显示屏 104、光标控制 106、气导麦克风 108、扬声器 110、两个骨导麦克风 112 和 114 以及可任选的邻近传感器 116。

触摸垫 102 允许用户将数字和字母输入到移动电话中。在其它实施例中，触摸垫 102 与显示屏 104 以触摸屏的形式组合。光标控制 106 允许用户加亮并选择显示屏 104 上的信息，并滚动通过大于显示屏 104 的图像和页面。

如图 2 和 3 所示，当移动电话 100 被防止在标准位置用于通过电话对话时，扬声器 110 位于用户左耳 200 或右耳 300 的附近，并且气导麦克风 108 位于用户口部 202 的附近。当电话位于用户的左耳时，如图 2 所示，骨导麦克风 114 接触用户的头盖骨或耳朵，并产生可用于从由气导麦克风 108 接收的语音信号中移除噪声的备选传感器信号。当电话位于用户的右耳时，如图 3 所示，骨导麦克风 112 接触用户的头盖骨或耳朵，并产生可用于从语音信号中移除噪声的备选传感器信号。

可任选邻近传感器 116 指示电话与用户如何接近。如下文进一步讨论的，该信息用于对骨导麦克风在产生干净语音值时的贡献进行加权。一般而言，如果邻近检测器检测到电话就在用户旁边，则骨导麦克风信号被赋予比远离用户某一距离时更大的权值。这一调整反映了这样一个事实：当骨导麦克风与用户接触时，其信号更能够表示用户正在说话。当它远离用户时，它更可疑地为环境噪声。邻近传感器在本发明的实施例中使用，因为用户不总是将电话压向其头部。

图 4 示出了本发明的骨导传感器 400 的一个实施例。在传感器 400 中，一软弹性桥 402 黏附在正常的气导麦克风 406 的横隔膜 404 上。该软桥 402 将来自用户的皮肤接触 408 的振动直接传导到麦克风 406 的横隔膜。横隔膜 404 的移动由麦克风 406 中的转换器转换成电信号。

图 5 提供了本发明的手持式移动设备的一个替换移动电话实施例 500。移动电话 500 包括键区 502、显示屏 504、光标控制 506、气导麦克风 508、扬声器 510 和组合的骨导麦克风和邻近传感器 512。

如图 6 的横截面中所示的，组合的骨导麦克风和邻近传感器 512 包括一软的、填充了介质（用液体或弹性体）衬垫 600，它具有外表面 602，它被设计成当用户将电话紧贴在耳朵上时与用户接触。衬垫 600 形成了为来自扬声器的声音提供了通路的开口周围的环，扬声器位于该开口中或直接在电话 500 内位于开口之下。衬垫 600 不限于这一形状，可对该衬垫使用任何形状。然而，一般而言，如果衬垫 600 包括扬声器 501 的左边和右边部分，则它是较佳的，使得衬垫的至少一个部分与用户接触，而无论用户的哪一耳朵在电话旁边。衬垫的该部分可以是外部连续的，或可以是外部分离的，但是在电话内流畅地连接在一起。

电子压力转换器 604 液压地连接到衬垫 600 中的液体或弹性体中，并将衬垫

600 中的液体的压力转换成导体 606 上的电信号。电子压力转换器 604 的示例包括基于 MEMS 的转换器。一般而言，压力转换器 604 应当具有高频响应。

导线 606 上的电信号包括两个分量：DC 分量和 AC 分量。DC 分量提供了邻近传感器信号，因为当电话被压向用户的耳朵时，衬垫 600 内的静压将高于电话远离用户的耳朵某一距离时的静压。电信号的 AC 分量提供了骨导麦克风信号，因为用户头盖骨、下颚或耳朵的骨头中的振动引起衬垫 600 中的压力波动，它们由压力转换器 604 转换成 AC 电信号。在一个实施例中，将滤波器应用到电信号，以允许信号的 DC 分量和高于最小频率的 AC 能够通过。

尽管上文描述了骨导传感器的这两个示例，然而骨导传感器的其它形式也处于本发明的范围之内。

图 7 所示是本发明的一个实施例中移动设备 700 的框图。移动设备 700 包括微处理器 702、存储器 704、输入/输出（I/O）接口 706 和用于与远程计算机、通信网络或其它移动设备通信的通信接口 708。在一个实施例中，上述组件被耦合在一起，用于通过合适的总线 710 彼此通信。

存储器 704 可以被实现为非易失电子存储器，如具有电池备份模块（未示出）的随机存取存储器（RAM），使得当移动设备 700 的总电源被关闭时，储存在存储器 704 中的信息也不会丢失。或者，存储器 704 的所有或部分可以是易失或非易失可移动存储器。存储器 704 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 704 的另一部分较佳地用于存储，如模拟盘驱动器上的存储。

存储器 704 包括操作系统 712、应用程序 714 以及对象存储 716。在操作过程中，操作系统 712 较佳地由处理器 702 从存储器 704 中执行。在一个较佳实施例中，操作系统 712 是可从微软公司购买的 WINDOWS® CE 品牌的操作系统。操作系统 712 较佳地被设计成用于移动设备，并实现可由应用程序 714 通过一组展现的应用编程接口和方法来使用的数据库特征。对象存储 716 中的对象由应用程序 714 和操作系统 712 至少部分地响应于对所展现的应用编程接口和方法的调用来维护。

通信接口 708 表示允许移动设备 700 发送和接收信息的多种设备和技术。在移动电话环境中，通信接口 708 代表了蜂窝电话网络接口，它与蜂窝电话网络通信以允许呼叫可被放置或接收。可能由通信接口 708 表示的其它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，此处仅举几个例子。移动设备 700 也可直接连接到计算机上，以与其交换数据。在这些情况下，通信接口 708 可以是红外收发器或串行或并行通信连接，所有这些都能够发送流信息。

由处理器 702 执行来实现本发明的计算机可执行指令可以储存在存储器 704 中，或通过通信接口 708 接收。这些指令在计算机可读介质中找到，包括但不限于计算机存储介质和通信介质。

计算机存储介质包括以用于储存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任一方法或技术实现的易失和非易失，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、磁盒、磁带、磁盘存储或其它磁存储设备、或可以用来储存所期望的信息并可访问的任一其它介质。

通信介质通常在诸如载波或其它传输机制的已调制数据信号中包含计算机可读指令、数据结构、程序模块或其它数据，并包括任一信息传送介质。术语“已调制数据信号”指以对信号中的信息进行编码的方式设置或改变其一个或多个特征的信号。作为示例而非局限，通信介质包括有线介质，如有线网络或直接连线连接，以及无线介质，如声学、RF、红外和其它无线介质。上述任一的组合也应当包括在计算机可读介质的范围之内。

输入/输出接口 706 表示到包括扬声器、数字输入 732（如一个或一组按钮、触摸屏、跟踪球、鼠标垫、滚轴或这些组件的组合，它们可由用户的大拇指或手指操纵）、显示屏 734、气导麦克风 736、备选传感器 738、备选传感器 740 和邻近传感器 742 的输入和输出设备的集合的接口。在一个实施例中，备选传感器 738 和 740 是骨导麦克风。上文列出的设备作为示例，不需要在移动设备 700 中都存在。此外，在至少一个实施例中，备选传感器和邻近传感器被组合成单个传感器，它提供邻近传感器信号和备选传感器信号。这些信号可被放置在单独的导线上，或可以是单个导线上的信号的分量。另外，在本发明的范围之内，其它输入/输出设备可以附加到移动设备 700 或在其中找到。

图 8 提供了本发明的实施例的语音处理系统的基本框图。在图 8 中，说话者 800 生成语音信号 802，它由气导麦克风 804 以及备选传感器 806 和备选传感器 807 之一或两者检测。备选传感器的一个示例是骨导传感器，它位于用户的脸部或头盖骨（如颚骨）上，或与其相邻，或在用户的耳朵上，并传感对应于由用户生成的语音的耳朵、头骨或颚骨的振动。备选传感器的另一示例是红外传感器，它被瞄准并检测用户的口部运动。注意，在某些实施例中，仅存在一个备选传感器。气导麦克风 804 是常用于将音频空气波转换成电信号的麦克风的类型。

气导麦克风 804 也接收由一个或多个噪声源 810 生成的噪声 808。根据备选传

感器的类型和噪声级别，噪声 808 也可由备选传感器 806 和 807 检测。然而，在本发明的实施例中，备选传感器 806 和 807 通常比气导麦克风 804 对环境噪声更不敏感。由此，由备选传感器 806 和 807 分别生成的备选传感器信号 812 和 813 一般比由气导麦克风 804 生成的气导麦克风信号 814 包括更少的噪声。

如果有两个备选传感器，如两个骨导传感器，则传感器信号 812 和 813 可任选地被提供给比较/选择单元 815。比较/选择单元 815 比较这两个信号的强度，并选择较强的信号作为其输出 817。较弱的信号不被传递用于进一步处理。对于移动电话环境，如图 1-3 的移动电话，比较/选择单元 815 通常选择由与用户皮肤接触的骨导传感器生成的信号。由此，在图 2 中，来自骨导传感器 114 的信号将被选中，而在图 3 中，来自骨导传感器 112 的信号将被选中。

备选传感器信号 817 和气导麦克风信号 814 被提供给干净信号估算器 816，它通过下文详细描述的过程估算干净信号 818。可任选地，干净信号估算器 816 也接收来自邻近传感器 832 的邻近信号 830，它用于估算干净信号 818。如上所述，在某些实施例中，邻近传感器可以与备选传感器信号相组合。干净信号估算 818 被提供给语音处理 820。干净语音信号 818 可以是经滤波的时域信号或特征域矢量。如果干净信号估算 818 是时域信号，则语音处理 820 可采用收听者、蜂窝电话发送器、语音编码系统或语音识别系统的形式。如果干净语音信号 818 是特征域矢量，则语音处理 820 通常是语音识别系统。

干净信号估算器 816 也产生噪声估算 819，它指示干净语音信号 818 中估算的噪声。噪声估算 819 被提供给侧音生成器 821，它基于噪声估算 819 生成通过移动设备的扬声器的音调。具体地，当噪声估算 819 提高时，侧音生成器 812 提高了侧音的音量。

侧音向用户提供了反馈，指示用户是否将移动设备保持在最佳的位置，以充分利用备选传感器。例如，如果用户未将骨导传感器压紧其头部，则干净信号估算器将接收到较差的备选传感器信号，并且由于较差的备选传感器信号会产生含噪声的干净信号 818。这会导致较响的侧音。当用户将骨导传感器与其头部接触时，备选传感器信号将得到改善，由此降低了干净信号 818 中的噪声，并降低了侧音的音量。由此，用户可基于侧音中的反馈快速地了解如何握住电话以最好地降低干净信号中的噪声。

在一个替换实施例中，侧音是基于来自邻近传感器 32 的邻近传感器信号 803 生成的。当邻近传感器指示电话接触或极接近用户的头部时，侧音音量较低。当邻

近传感器指示电话远离用户的头部时，侧音将更响。

本发明使用若干方法和系统，以利用气导麦克风 814、备选传感器信号 817 和可任选邻近传感器信号 830 估算干净信号。一种系统使用立体声训练数据来训练备选传感器信号的纠正矢量。当这些纠正矢量稍后被添加到测试备选传感器矢量时，它们提供干净信号矢量的估算。本系统的一个进一步扩展是首先跟踪时变失真，然后将该信息结合到纠正矢量的计算和干净信号的估算中。

第二种系统提供了由纠正矢量生成的干净信号估算和通过从气导信号中减去气导测试信号中的当前噪声的估算形成的估算之间的内插。第三种系统使用备选传感器信号来估算语音信号的基音，然后使用所估算的基音来标识干净语音信号的估算。这些系统的每一个在下文单独地讨论。

训练立体声纠正矢量

图 9 和 10 提供了本发明的两个实施例的训练立体声纠正矢量的框图和流程图，它们依赖于纠正矢量来生成干净信号的估算。

标识纠正矢量的方法在图 10 的步骤 1000 开始，其中，将“干净”气导麦克风信号转换成特征矢量序列。为此，图 9 的说话者 900 对气导麦克风 910 说话，后者将音频波转换成电信号。电信号然后由模一数转换器 914 采样，以生成数字值序列，它们由帧构造器 916 组合成值的帧。在一个实施例中，A-D 转换器 914 以 16kHz 和每样值 16 比特对模拟信号进行采样，由此创建了每秒 32 千字节的语音数据，并且帧构造器 916 每 10 毫秒创建包括 25 毫秒数据的新帧。

由帧构造器 916 提供的每一数据帧由特征提取器 918 转换成特征矢量。在一个实施例中，特征提取器 918 形成倒谱特征。这类特征的示例包括 LPC 导出的倒谱和梅尔（Mel）频率倒谱系数。可用于本发明的其它可能的矢量提取模块的示例包括用于执行线性预测编码（LPC）、感知线性预测（PLP）以及听觉模型特征提取的模块。注意，本发明不限于这些特征提取模块，在本发明的环境中也可使用其它模块。

在图 10 的步骤 1002，将备选传感器信号转换成特征矢量。尽管步骤 1002 的转换被示出为在步骤 1000 的转换之后发生，然而在本发明中，转换的任一部分可以在步骤 1000 之前、期间或之后发生。步骤 1002 的转换通过类似于上文相对于步骤 1000 所描述的过程来执行。

在图 9 的实施例中，该过程在备选传感器 902 和 903 检测到与说话者 900 的

语音产生相关联的物理事件开始，如骨振动或面部运动。由于备选传感器 902 和 903 在移动设备上相互隔开，它们不会检测到关于语音产生的相同的值。备选传感器 902 和 903 将物理事件转换成模拟电信号。这些电信号被提供给比较/选择单元 904，它标识这两个信号中较强的一个，并在其输出提供较强的信号。注意，在某些实施例中，仅使用一个备选传感器。在这一情况下，比较/选择单元 904 不存在。

所选择的模拟信号由模一数转换器 905 采样。A/D 转换器 905 的采样特征与上文相对于 A/D 转换器 914 所描述的相同。由 A/D 转换器 905 提供的样值由帧构造器 906 收集成帧，后者以类似于帧构造器 916 的方式运作。样值帧然后由特征提取器 908 转换成特征矢量，后者使用与特征提取器 918 相同的特征提取方法。

备选传感器信号和气导信号的特征矢量被提供给图 9 中的降噪训练器 902。在图 10 的步骤 1004，降噪训练器 920 将备选传感器信号的特征矢量组合成混合分量。这一组合可通过使用最大似然性训练技术将类似的特征矢量组合在一起来完成，或通过将表示语音信号的时间部分的特征矢量组合在一起完成。本领域的技术人员将认识到，可使用组合特征矢量的其它技术，并且上文列出的两个技术仅作为示例来提供。

在图 10 的步骤 1008，降噪训练器 902 然后确定每一混合分量 s 的纠正矢量 $r_s$。在一个实施例中，每一混合分量的纠正矢量使用最大似然性标准来确定。在这一技术中，纠正矢量计算如下：

$$r_s = \frac{\sum_t p(s|b_t)(x_t - b_t)}{\sum_t p(s|b_t)} \qquad \text{公式 1}$$

其中 $x_t$ 是帧 t 的气导矢量值，$b_t$ 是帧 t 的备选传感器矢量值。在公式 1 中：

$$p(s|b_t) = \frac{p(b_t|s)p(s)}{\sum_s p(b_t|s)p(s)} \qquad \text{公式 2}$$

其中，p(s)仅是多个混合分量中的一个，$p(b_t|s)$被模型化为高斯分布：

$$p(b_t|s) = N(b_t;\mu_b,\Gamma_b) \qquad \text{公式 3}$$

它具有均值 $\mu_b$ 和方差 $\Gamma_b$，它们使用期望值最大化（EM）算法来训练，其中，每一次迭代包括以下步骤：

$$\gamma_s(t) = p(s|b_t) \qquad \text{公式 4}$$

$$\mu_s = \frac{\sum_t \gamma_s(t) b_t}{\sum_t \gamma_s(t)} \qquad \text{公式 5}$$

$$\Gamma_s = \frac{\sum_t \gamma_s(t)(b_t - \mu_s)(b_t - \mu_s)^T}{\sum_t \gamma_s(t)} \qquad \text{公式 6}$$

公式 4 是 EM 算法中的 E 步骤，它使用先前估算的参数。公式 5 和公式 6 是 M 步骤，它使用 E 步骤的结果更新参数。

该算法的 E 步骤和 M 步骤反复，直到确定了模型参数的稳定值。这些参数然后用于评估公式 1，以形成纠正矢量。纠正矢量和模型参数然后被储存在降噪参数存储 922 中。

在步骤 1008 对每一混合分量确定了纠正矢量之后，训练本发明的降噪系统的过程完成。一旦对每一混合分量确定了纠正矢量，该矢量可在本发明的降噪技术中使用。下文讨论使用就纠正矢量的两个单独的降噪技术。

使用纠正矢量和噪声估算的降噪

基于纠正矢量和噪声估算降低含噪声的语音信号中的噪声的系统和方法分别在图 11 的框图和图 12 的流程图中示出。

在步骤 1200，由气导麦克风 1104 检测的测试信号被转换成特征矢量。由麦克风 1104 接收的音频测试信号包括来自说话者 1100 的语音和来自一个或多个噪声源 1102 的加性噪声。由麦克风 1104 检测到的音频测试信号被转换成电信号，它被提供给模一数转换器 1106。

A-D 转换器 1106 将来自麦克风 1104 的模拟信号转换成一系列数字值。在若干实施例中，A-D 转换器 1106 以 16kHz 和每样值 16 比特对模拟信号进行采样，由此创建了每秒 32 千字节的语音数据。这些数字值被提供给帧构造器 1108，在一个实施例中，帧构造器 1108 将值组合成 25 毫秒的帧，其起始处相隔 10 毫秒。

由帧构造器 1108 创建的数据帧被提供给特征提取器 1110，它从每一帧中提取特征。在一个实施例中，该特征提取器不同于用于训练纠正矢量的特征提取器 908 和 918。具体地，在这一实施例中，特征提取器 1110 产生功率谱值而非倒谱值。所提取的特征被提供给干净信号估算器 1122、语音检测单元 1126 和噪声模型训练器 1124。

在步骤 1202，与说话者 1100 的语音产生相关联的物理事件，如骨振动或面部运动被转换成特征矢量。尽管在图 12 中被示出为单独的步骤，然而本领域的技术人员将认识到，该步骤可以与步骤 1200 同时完成。在步骤 1202 中，由备选传感器 1112 和 1114 之一或两者检测物理事件。备选传感器 1112 和 1114 基于物理事件生成模拟电信号。模拟信号被提供给比较和选择单元 1115，它选择较大幅度的信号作为其输出。注意，在某些实施例中，仅提供了一个备选传感器。在这一实施例中，

比较和选择单元 1115 是不需要的。

所选择的模拟信号由模一数转换器 1116 转换成数字信号，并且所得的数字样值由帧构造器 1118 组合成帧。在一个实施例中，模一数转换器 1116 和帧构造器 1118 以类似于模一数转换器 1106 和帧构造器 1108 的方式运作。

数字值的帧被提供给特征提取器 1120，它使用用于训练纠正矢量的相同的特征提取技术。如上所述，这类特征提取模块的示例包括用于执行线性预测编码（LPC）、LPC 导出倒谱、感知线性预测（PLP）、听觉模型特征提取以及梅尔频率倒谱系数（MFCC）特征提取的模块。然而，在许多实施例中，使用了产生倒谱特征的特征提取技术。

特征提取模块产生特征矢量流，其每一个都与语音信号的一个单独帧相关联。该特征矢量流被提供给干净信号估算器 1122。

来自帧构造器 1118 的值帧也被提供给特征提取器 1121，在一个实施例中，特征提取器 1121 提取每一帧的能量。每一帧的能量值被提供给语音检测单元 1126。

在步骤 1204，语音检测单元 1126 使用备选传感器信号的能量特征来确定何时可能存在语音。该信息被传递到噪声模型训练器 1124，在步骤 1206，它试图在没有语音的时间段中对噪声建模。

在一个实施例中，语音检测单元 1126 首先搜索帧能量值序列以找出能量中的峰值。它然后搜索峰值之后的谷值。该谷值的能量被称为能量分隔符 d。为确定帧是否包含语音，然后确定帧能量 e 与能量分隔符 d 之间的比值 k，如下：k=e/d。然后确定该帧的语音置信度 q，如下：

$$q=\begin{cases}0 & : \quad k<1\\ \dfrac{k-1}{\alpha-1} & : \quad 1\le k\le\alpha\\ 1 & : \quad k>\alpha\end{cases} \qquad \text{公式 7}$$

其中，$\alpha$ 定义了两种状态之间的转移，在一个实施例中被设为 2。最后，将其 5 个相邻帧（包括其本身）的平均置信度值用作该帧的最终置信度值。

在一个实施例中，使用固定的阈值来确定是否存在语音，使得如果置信度值超过阈值，则该帧被认为包含语音，如果置信度值不超过阈值，则该帧被认为包含非语音。在一个实施例中，使用了 0.1 的阈值。

对于由语音检测单元 1126 检测的每一非语音帧，噪声模型训练器 1124 在步骤 1206 更新噪声模型 1125。在一个实施例中，噪声模型 1125 是高斯模型，它具有均值 $\mu_n$ 和方差 $\Sigma_n$。该模型基于非语音的最近帧的移动窗。用于从窗中的非语音

帧中确定均值和方差的技术在本领域中是众所周知的。

参数存储 922 中的纠正矢量和模型参数以及噪声模型 1125，连同备选传感器的特征矢量 $b$ 和含噪声的气导麦克风信号的特征矢量 $S_y$ 一起被提供给干净信号估算器 1122。在步骤 1208，干净信号估算器 1122 基于备选传感器特征矢量、纠正矢量和备选传感器的模型参数估算干净语音信号的初始值。具体地，干净信号的备选传感器估算计算如下：

$$\hat{x} = b + \sum_s p(s|b) r_s \qquad \text{公式 8}$$

其中，$\hat{x}$ 是倒谱域中的干净信号估算，$b$ 是备选传感器特征矢量，$p(s|b)$ 是使用上文的公式 2 来确定的，$r_s$ 是混合分量 $s$ 的纠正矢量。由此，公式 8 中的干净信号估算通过将备选传感器特征矢量添加到纠正矢量的加权和来形成，其中，加权基于给定备选传感器特征矢量时混合分量的概率。

在步骤 1210，通过将初始备选传感器干净语言估算与从含噪声的气导麦克风矢量和噪声模型中形成的干净语音估算相组合，初始备选传感器干净语音估算被净化。这得到一经净化的干净语音估算 1128。为将初始干净信号估算的倒谱值与含噪声的气导麦克风的功率谱特征矢量相组合，使用以下公式将倒谱值转换成功率谱域：

$$\hat{S}_{x|b} = e^{C^{-1}\hat{x}} \qquad \text{公式 9}$$

其中，$C^{-1}$ 是离散余弦反变换，$\hat{S}_{x|b}$ 是基于备选传感器的干净语音的功率谱估算。

一旦已将来自备选传感器的初始干净信号估算放入功率谱域中，它可与含噪声的气导麦克风矢量和噪声模型相组合，如下：

$$\hat{S}_x = (\Sigma_n^{-1} + \Sigma_{x|b}^{-1})^{-1}[\Sigma_n^{-1}(S_y - \mu_n) + \Sigma_{x|b}^{-1}\hat{S}_{x|b}] \qquad \text{公式 10}$$

其中，$\hat{S}_x$ 是功率谱域中经净化的干净信号估算，$S_y$ 是含噪声的气导麦克风特征矢量，$(\mu_n, \Sigma_n)$ 是先验噪声模型（见 1124）的均值和协方差，$\hat{S}_{x|b}$ 是基于备选传感器的初始干净信号估算，$\Sigma_{x|b}$ 是给定备选传感器的测量时干净信号的条件概率分布的协方差矩阵。$\Sigma_{x|b}$ 可被计算如下。设 $J$ 表示公式 9 的右侧的函数的雅各比行列式。设 $\Sigma$ 是 $\hat{x}$ 的协方差矩阵，则 $\hat{S}_{x|b}$ 的协方差为

$$\Sigma_{x|b} = J\Sigma J^T \qquad \text{公式 11}$$

在一简化的实施例中，公式 10 被重写成以下公式：

$$\hat{S}_x = \alpha(f)(S_y - \mu_n) + (1 - \alpha(f))\hat{S}_{x|b} \qquad \text{公式 12}$$

其中，$\alpha(f)$是时间和频带的函数。例如，如果备选传感器的频带达3KHz，对低于3KHz的频带选择$\alpha(f)$为0。基本上，对于低频带，来自备选传感器的初始干净信号估算是可信的。

对于高频带，来自备选传感器的初始干净信号估算并不可靠。直观上，当在当前帧上频带的噪声较小时，选择较大的$\alpha(f)$，使得对这一频带可从气导麦克风中取出更多的信息。否则，通过选择较小的$\alpha(f)$使用来自备选传感器的更多的信息。在一个实施例中，使用来自备选传感器的初始干净信号估算的能量来确定每一频带的噪声级别。设$E(f)$表示频带$f$的能量。设$M = Max_f E(f)$，作为$f$的函数，$\alpha(f)$被定义如下：

$$\alpha(f)=\begin{cases}\dfrac{E(f)}{M} & : \quad f \geq 4K \\ \dfrac{f-3K}{1K}\alpha(4K) & : \quad 3K < f < 4K \\ 0 & : \quad f \leq 3K\end{cases} \qquad \text{公式 13}$$

其中，使用了线性内插从3K过渡到4K，以确保$\alpha(f)$的平滑性。

在一个实施例中，移动设备与用户头部的邻近性被结合到$\alpha(f)$的确定中。具体地，如果邻近传感器832产生最大距离值$D$和当前距离值$d$，则公式13可被修改为：

$$\alpha(f)=\begin{cases}\beta\dfrac{E(f)}{M}+(1-\beta)\dfrac{d}{D} & : \quad f \geq 4K \\ \dfrac{f-3K}{1K}\alpha(4K) & : \quad 3K < f < 4K \\ 0 & : \quad f \leq 3K\end{cases} \qquad \text{公式 14}$$

其中，$\beta$在0到1之间，并基于哪一矢量、能量或邻近性被认为能够提供气导麦克风的噪声模型或备选传感器的纠正矢量将提供干净信号的最佳估算的最佳指示来选择。

如果$\beta$被设为0，则$\alpha(f)$不再是频率相关的，并简单地变为：

$$\alpha=\frac{d}{D} \qquad \text{公式 15}$$

功率谱域中经净化的干净信号估算可用于构造维纳（Wiener）滤波器，以对含噪声的气导麦克风信号进行滤波。具体地，设置维纳滤波器h，使得：

$$H=\frac{\hat{S}_x}{S_y} \qquad \text{公式 16}$$

该滤波器然后可被应用于时域含噪声的气导麦克风信号，以产生经降噪的或

干净的时域信号。经降噪的信号可被提供给收听者或可应用于语音识别器。

注意，公式 12 提供了经净化的干净信号估算，它是两个因子的加权和，其中一个是来自备选传感器的干净信号估算。该加权和可被扩充以包括附加备选传感器的附加因子。由此，可使用一个以上备选传感器来生成干净信号的独立估算。这多个估算然后可使用公式 12 来组合。

在一个实施例中，经净化的干净信号估算中的噪声也被估算。在一个实施例中，该噪声被认为是 0 均值的高斯型，其协方差被确定如下：

$$\Sigma_x = (\Sigma_n^{-1} + \Sigma_{x|b}^{-1})^{-1} = \Sigma_n \Sigma_{x|b} / (\Sigma_n + \Sigma_{x|b})$$

其中，$\Sigma_n$ 是气导麦克风中的噪声的方差，$\Sigma_{x|b}$ 是来自备选传感器的估算中的噪声的方差。具体地，如果备选传感器不与皮肤表面较好地接触，则 $\Sigma_{x|b}$ 较大。接触的程度可通过使用附加邻近传感器或分析备选传感器来测量。对于后者，观察到如果接触良好，则备选传感器几乎不产生高频响应（大于 4KHz），则用低频能量（小于 3KHz）与高频能量之比来测量接触。该比值越高，接触越好。

在某些实施例中，干净信号估算中的噪声用于生成如上文相对于图 6 所描述的侧音。当经净化的干净信号估算中的噪声增加时，侧音的音量也提高，以鼓励用户将备选传感器放置在更好的位置，使得增强处理得以改进。例如，侧音鼓励用户将骨导传感器压向其头部，使得增强处理得以改进。

使用纠正矢量而没有噪声估算的降噪

图 13 提供了本发明中估算干净语音值的替换系统的框图。图 13 的系统类似于图 11 的系统，除干净语音值的估算在不需要气导麦克风或噪声模型的情况下形成之外。

在图 13 中，与产生语音的说话者 1300 相关联的物理事件由备选传感器 1302、模一数转换器 1304、帧构造器 1306 和特征提取器 1308 以与上述图 11 的备选传感器 1114、模一数转换器 1116、帧构造器 1117 和特征提取器 1118 相似的方式转换成特征矢量。注意，尽管在图 13 中仅示出了一个备选传感器，然而如图 11 中一样，可使用附加的备选传感器，外加图 11 中讨论的比较和选择单元。

来自特征提取器 1308 的特征矢量以及降噪参数 922 被提供给干净信号估算器 1310，它使用上文的公式 8 和 9 确定干净信号值 1312 $\hat{S}_{x|b}$ 的估算。

功率谱域中干净信号估算 $\hat{S}_{x|b}$ 可用于构造维纳滤波器，以对含噪声的气导麦克风信号进行滤波。具体地，设置维纳滤波器 H，使得：

$$H=\frac{\hat{S}_{x|b}}{S_y} \qquad \text{公式 17}$$

该滤波器可被应用于时域含噪声的气导麦克风信号，以产生经降噪的或干净信号。经降噪的信号可被提供给收听者或被应用于语音识别器。

或者，公式 8 中计算的倒谱域中的干净信号估算 $\hat{x}$ 可被直接应用于语音识别系统。

使用基音跟踪的降噪

生成干净语音信号估算的一个替换技术在图 14 的框图和图 15 的流程图中示出。具体地，图 14 和 15 的实施例通过使用备选传感器然后使用基音将含噪声的气导麦克风信号分解成谐波分量和随机分量，来标识语音信号的基音，从而确定了干净信号估算。由此，含噪声的信号被表示为：

$$y=y_h+y_r \qquad \text{公式 18}$$

其中，$y$ 是含噪声的信号，$y_h$ 是谐波分量，$y_r$ 是随机分量。使用谐波分量和随机分量的加权和来形成表示经降噪的语音信号的经降噪的特征矢量。

在一个实施例中，谐波分量被模型化为谐波上相关的正弦和，使得：

$$y_h=\sum_{k=1}^{K}a_k\cos(k\omega_0 t)+b_k\sin(k\omega_0 t) \qquad \text{公式 19}$$

其中，$\omega_0$ 是基频或基音频率，K 是信号中的谐波总数。

由此，为标识谐波分量，必须确定基音频率和幅度参数 $\{a_1a_2\ldots a_kb_1b_2\ldots b_k\}$ 的估算。

在步骤 1500，收集含噪声的语音信号，并将其转换成数字样值。为此，气导麦克风 1404 将来自说话者 1400 和一个或多个加性噪声源 1402 的音频波转换成电信号。电信号然后由模一数转换器 1406 转换，以生成一数字值序列。在一个实施例中，A-D 转换器 1406 以 16kHz 和每样值 16 比特对模拟信号进行采样，由此创建了每秒 32 千字节的语音数据。在步骤 1502，数字样值由帧构造器 1408 组合成帧。在一个实施例中，帧构造器 1408 每 10 毫秒创建包括 25 毫秒数据的新帧。

在步骤 1504，与语音产生相关联的物理事件由备选传感器 1444 检测。在本实施例中，能够检测谐波分量的备选传感器，如骨导传感器最适合用作备选传感器 1444。注意，尽管步骤 1504 被示出为与步骤 1500 分离，然而本领域的技术人员将认识到，这些步骤可以同时执行。另外，尽管在图 14 中仅示出了一个备选传感器，

然而可如图11中一样使用附加的备选传感器，外加图11中所述的比较和选择单元。

由备选传感器 1444 生成的模拟信号由模一数转换器 1446 转换成数字样值。数字样值然后在步骤 1506 由帧构造器 1448 组合成帧。

在步骤 1508，备选传感器信号的帧由基音跟踪器 1450 用于标识语音的基音频率或基频。

可使用多种可用基音跟踪系统的任一种来确定基音频率的估算。在许多这样的系统中，候选基音用于标识备选传感器信号的片断中心之间的间隔。对于每一候选基音，确定连续的语音片断之间的相关。一般而言，提供最佳相关的候选基音是该帧的基音频率。在某些系统中，使用附加信息来净化基音选择，如信号的能量和/或期望基音跟踪。

给定来自基音跟踪器 1450 的基音的估算，在步骤 1510，气导信号矢量可被分解成谐波分量和随机分量。为此，公式 19 被重写为：

$$\mathbf{y} = \mathbf{Ab} \qquad \text{公式 20}$$

其中，$\mathbf{y}$ 是含噪声的语音信号的 N 个样值的矢量，$\mathbf{A}$ 是 $N \times 2K$ 矩阵，由以下公式给出：

$$\mathbf{A} = [\mathbf{A}_{\cos}\mathbf{A}_{\sin}] \qquad \text{公式 21}$$

其元素为

$$\mathbf{A}_{\cos}(k,t) = \cos(k\omega_0 t) \qquad \mathbf{A}_{\sin}(k,t) = \sin(k\omega_0 t) \qquad \text{公式 22}$$

并且 $b$ 是 $2K \times 1$ 的矢量，由以下公式给出：

$$\mathbf{b}^T = [a_1 a_2 \ldots a_k b_1 b_2 \ldots b_k] \qquad \text{公式 23}$$

则振幅系数的最小二乘解为：

$$\hat{\mathbf{b}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} \qquad \text{公式 24}$$

使用 $\hat{\mathbf{b}}$，含噪声的语音信号的谐波分量的估算可以确定如下：

$$\mathbf{y}_h = \mathbf{A}\hat{\mathbf{b}} \qquad \text{公式 25}$$

随机分量的估算则被计算如下：

$$\mathbf{y}_r = \mathbf{y} - \mathbf{y}_h \qquad \text{公式 26}$$

由此，使用上述公式 20-26，谐波分解单元 1410 能够产生谐波分量样值 1412 的矢量 $\mathbf{y}_h$ 和随机分量样值的矢量 $\mathbf{y}_r$。

在将帧的样值分解成谐波和随机样值之后，在步骤 1512 对谐波分量确定一比例缩放参数或权值。该比例缩放参数被用作经降噪的语音信号的计算的一部分，如下文进一步讨论的。在一个实施例中，比例缩放参数计算如下：

$$\alpha_h = \frac{\sum_i y_h(i)^2}{\sum_i y(i)^2}$$ 公式 27

其中，$\alpha_h$ 是比例缩放参数，$y_h(i)$ 是谐波分量样值矢量 $\mathbf{y}_h$ 中的第 $i$ 个样值，$y(i)$ 是该帧的含噪声的语音信号的第 $i$ 个样值。在公式 27 中，分子是谐波分量的每一样值的能量之和，分母是含噪声的语音信号的每一样值的能量之和。由此，比例缩放参数是该帧的谐波能量与该帧的总能量之比。

在替换实施例中，比例缩放参数使用概率有声一无声检测单元来设置。这一单元提供了语音的特定帧为有声而非无声的概率，这意味着该帧中的声带共振。帧来自语音的有声区域的概率可直接用作比例缩放参数。

在确定了比例缩放参数之后，或正在被确定时，在步骤 1514 确定谐波分量样值矢量和随机分量样值矢量的梅尔频谱。这涉及令每一样值矢量通过一离散傅立叶变换（DFT），以产生谐波分量频率值矢量 1422 和随机分量频率值矢量 1420。由频率值矢量表示的功率谱然后由梅尔加权单元 1424 使用沿梅尔标度应用的一系列三角加权函数来平滑。这可得到谐波分量梅尔频谱矢量 1428 $\mathbf{Y}_h$ 和随机分量梅尔频谱矢量 1426 $\mathbf{Y}_r$。

在步骤 1516，将谐波分量和随机分量的梅尔频谱组合成加权和，以形成经降噪的梅尔频谱估算。这一步骤由加权和计算器 1430 使用以上确定的比例缩放因子在以下公式中执行：

$$\hat{\mathbf{X}}(t) = \alpha_h(t)\mathbf{Y}_h(t) + \alpha_r\mathbf{Y}_r(t)$$ 公式 28

其中，$\hat{\mathbf{X}}(t)$ 是经降噪的梅尔频谱的估算，$\mathbf{Y}_h(t)$ 是谐波分量梅尔频谱，$\mathbf{Y}_r(t)$ 是随机分量梅尔频谱，$\alpha_h(t)$ 是以上确定的比例缩放因子，$\alpha_r$ 是随机分量的固定的比例缩放因子，在一个实施例中，它被设为 1，时间下标 t 用于强调谐波分量的比例缩放因子是对每一帧确定的，而随机分量的比例缩放因子保持不变。注意，在其它实施例中，随机分量的比例缩放因子可对每一帧确定。

在步骤 1516 计算了经降噪的梅尔频谱之后，确定梅尔频谱的对数 1432，然后在步骤 1518 将其应用于离散余弦变换 1434。这产生了梅尔频率倒谱系统（MFCC）特征矢量 1436，它表示经降噪的语音信号。

对含噪声的信号的每一帧产生一单独的经降噪的 MFCC 特征矢量。这些特征矢量可用于任何期望的目的，包括语音增强和语音识别。对于语音增强，MFCC 特征矢量可被转换到功率谱域，并可用含噪声的气导信号来形成维纳滤波器。

尽管特别参照使用骨导传感器作为备选传感器来讨论本发明，然而可使用其

它备选传感器。例如，在图 16 中，本发明的移动设备使用红外传感器 1600，它一般瞄准用户的脸部，尤其是口部，并生成指示用户的面部运动中对应于语音的变化。由红外传感器 1600 生成的信号可用作上述技术中的备选传感器信号。

尽管参考特定的实施例描述了本发明，然而本领域的技术人员将认识到，可以在不脱离本发明的精神和范围的情况下在形式和细节上作出改变。

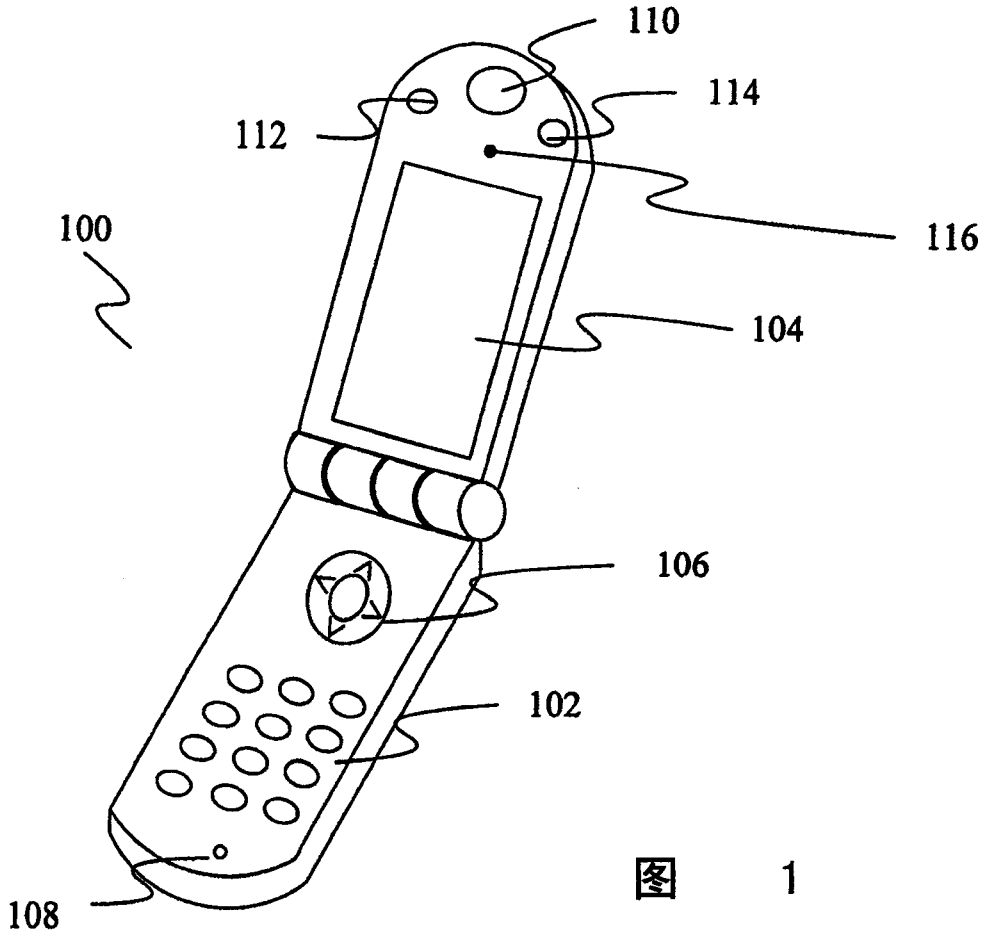

110
114
112
116
100
104
106
102
108

图 1

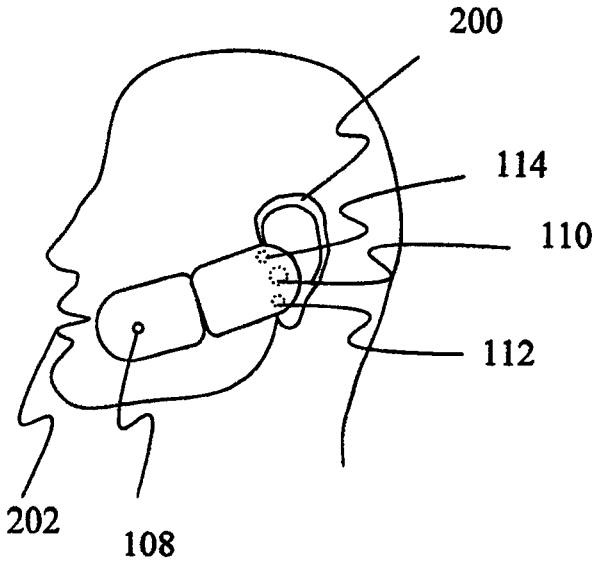

200
114
110
112
202
108

图 2

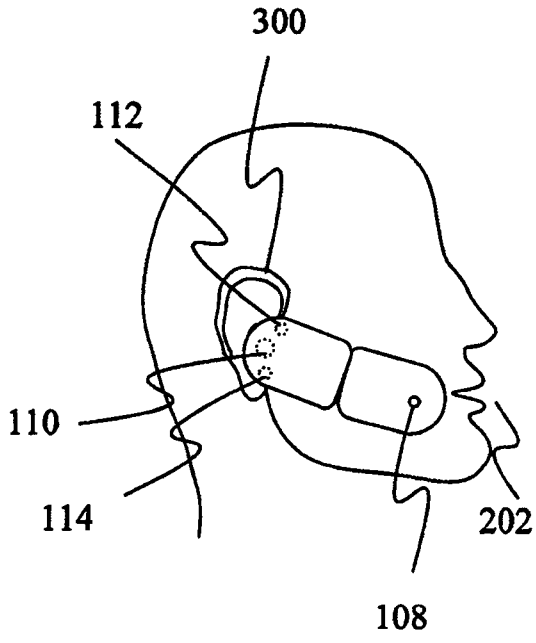

300
112
110
114
202
108

图 3

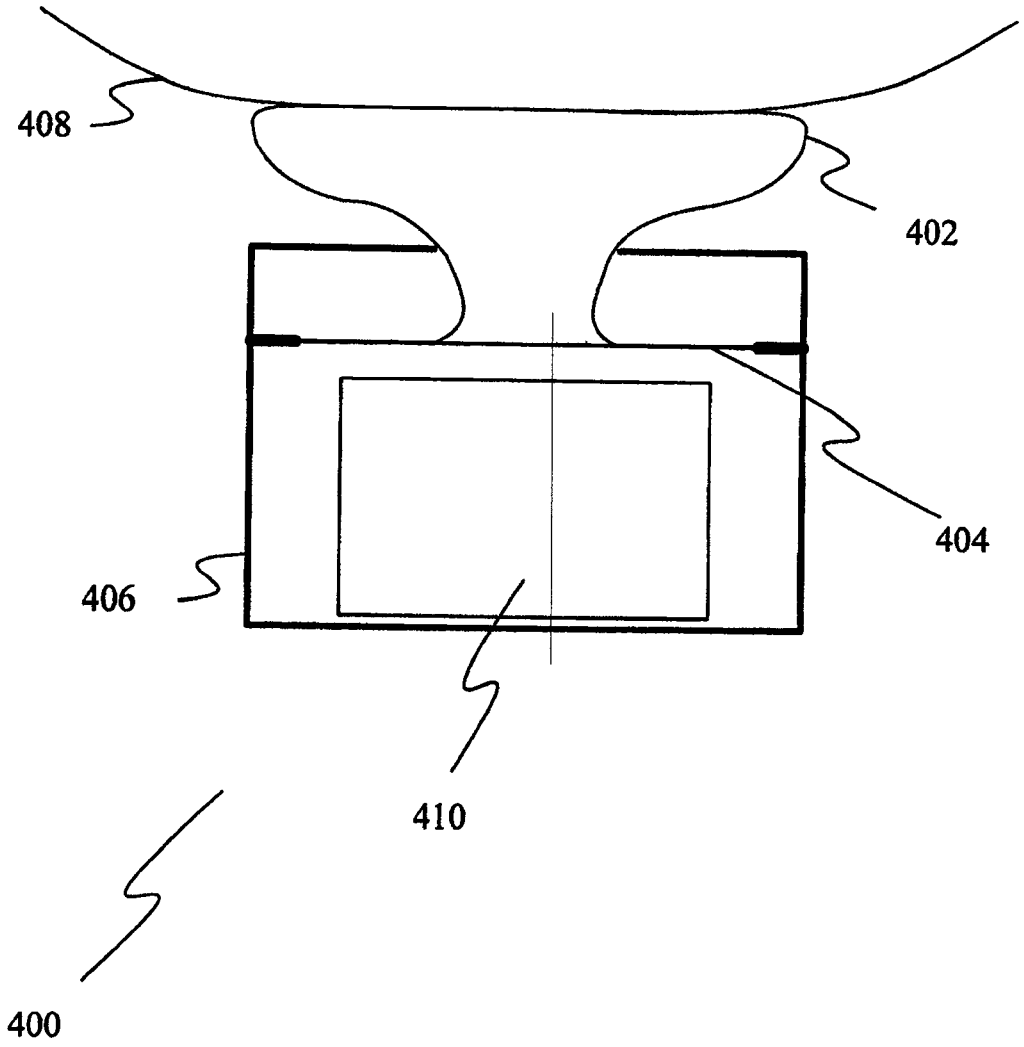



图 4

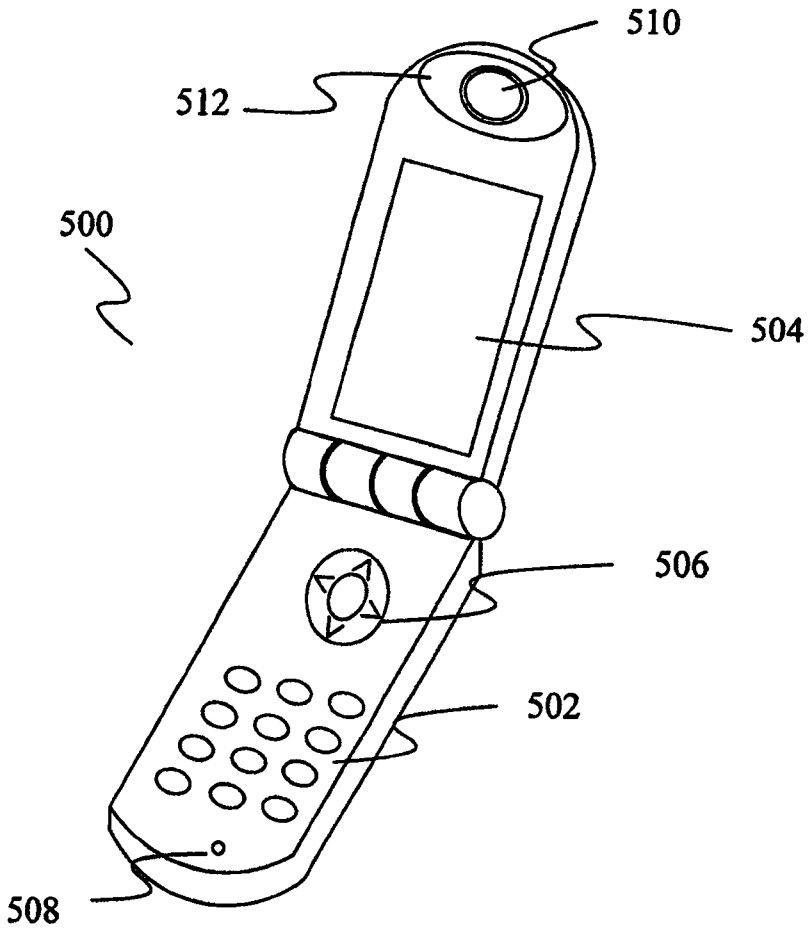



图 5

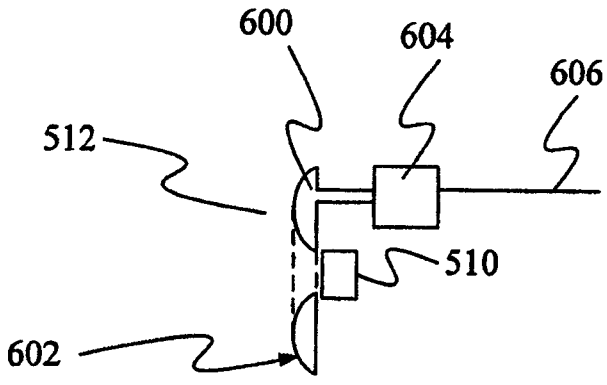



图 6

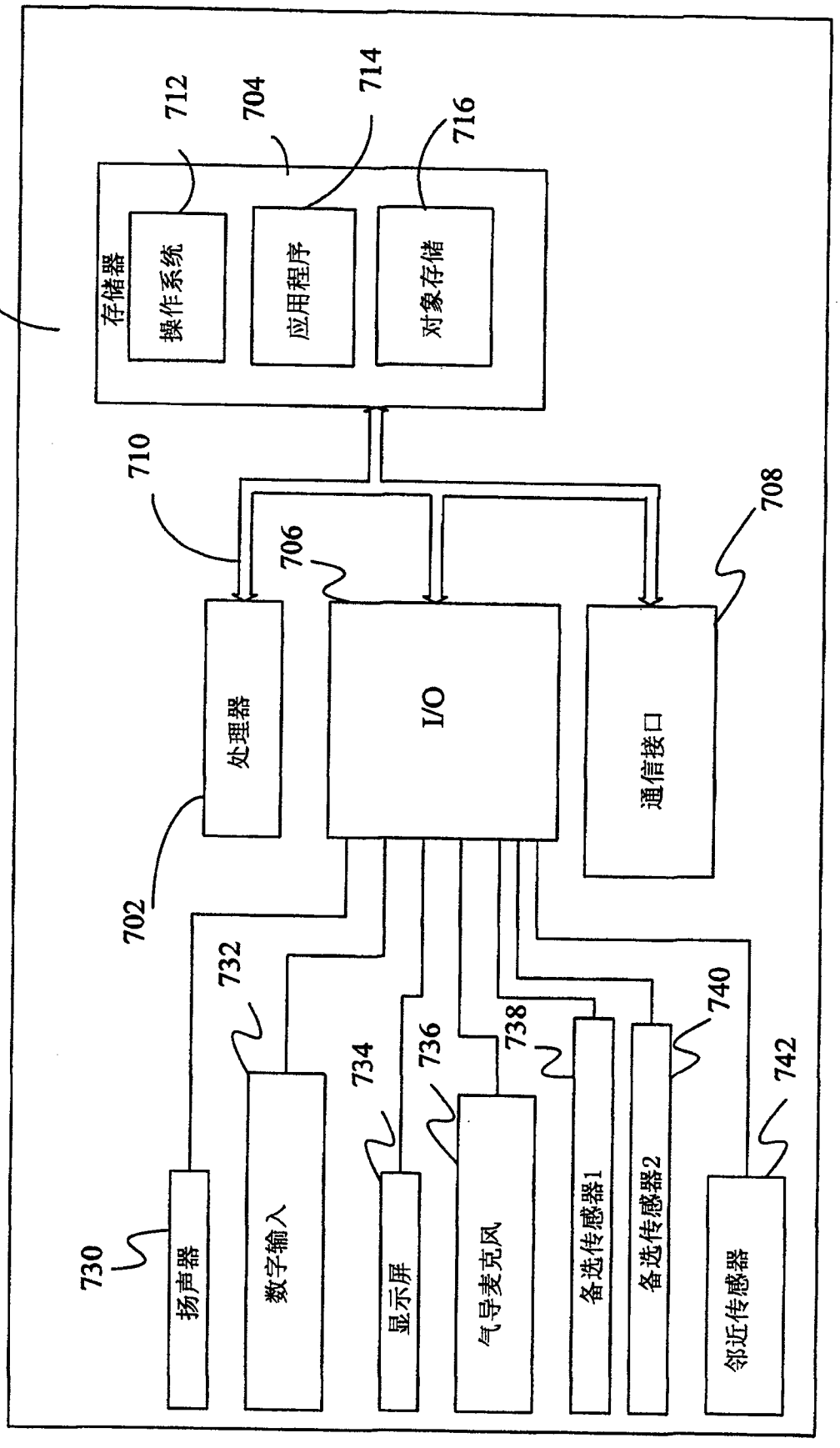

700
712
704
714
716
存储器
操作系统
应用程序
对象存储
710
706
708
处理器
I/O
通信接口
702
732
734
736
738
740
742
730
扬声器
数字输入
显示屏
气导麦克风
备选传感器1
备选传感器2
邻近传感器

图 7

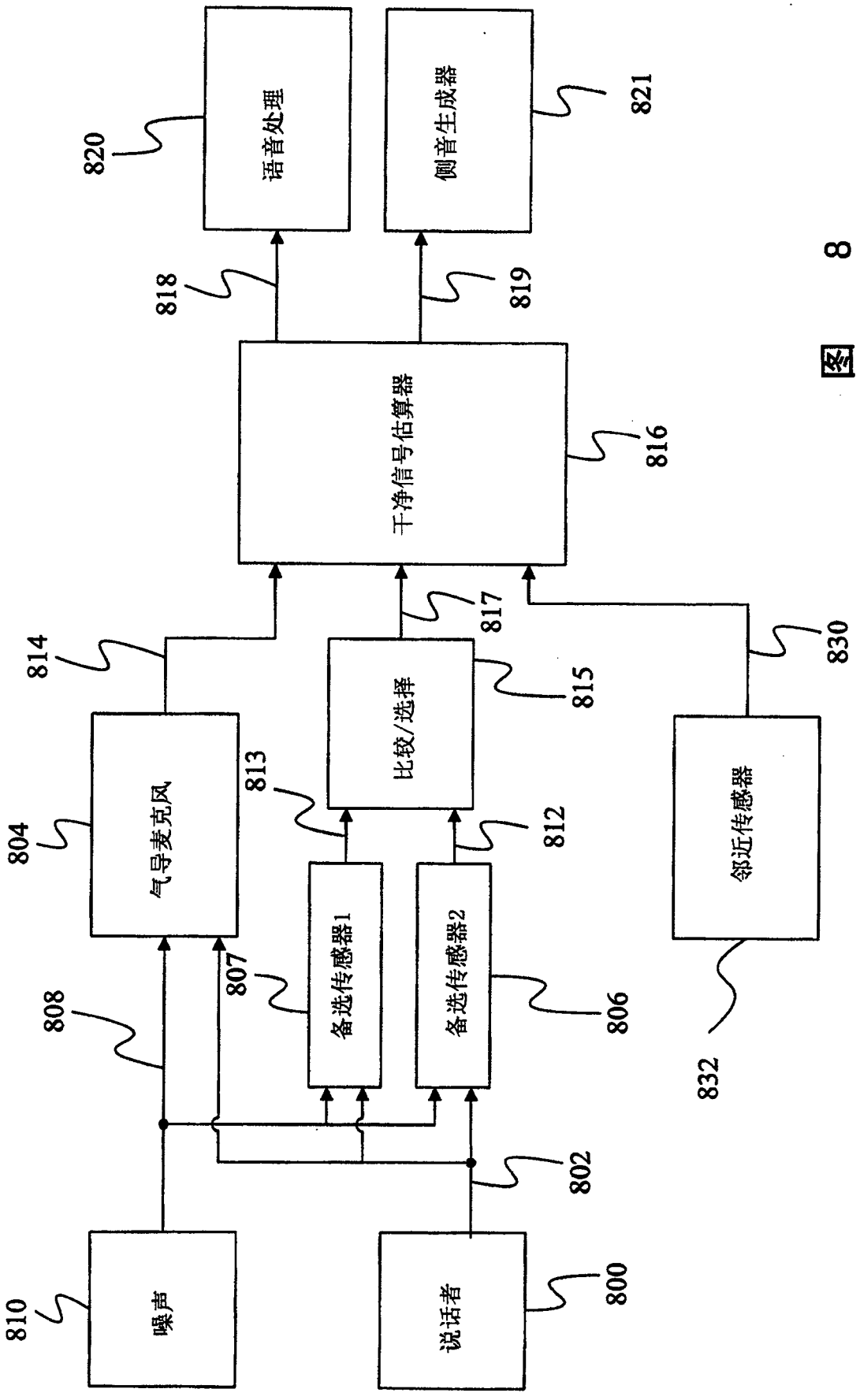

820
语音处理
821
侧音生成器
818
819
干净信号估算器
816
814
817
830
比较/选择
815
813
812
804
气导麦克风
备选传感器1
备选传感器2
邻近传感器
808
807
806
832
802
810
噪声
说话者
800

图 8

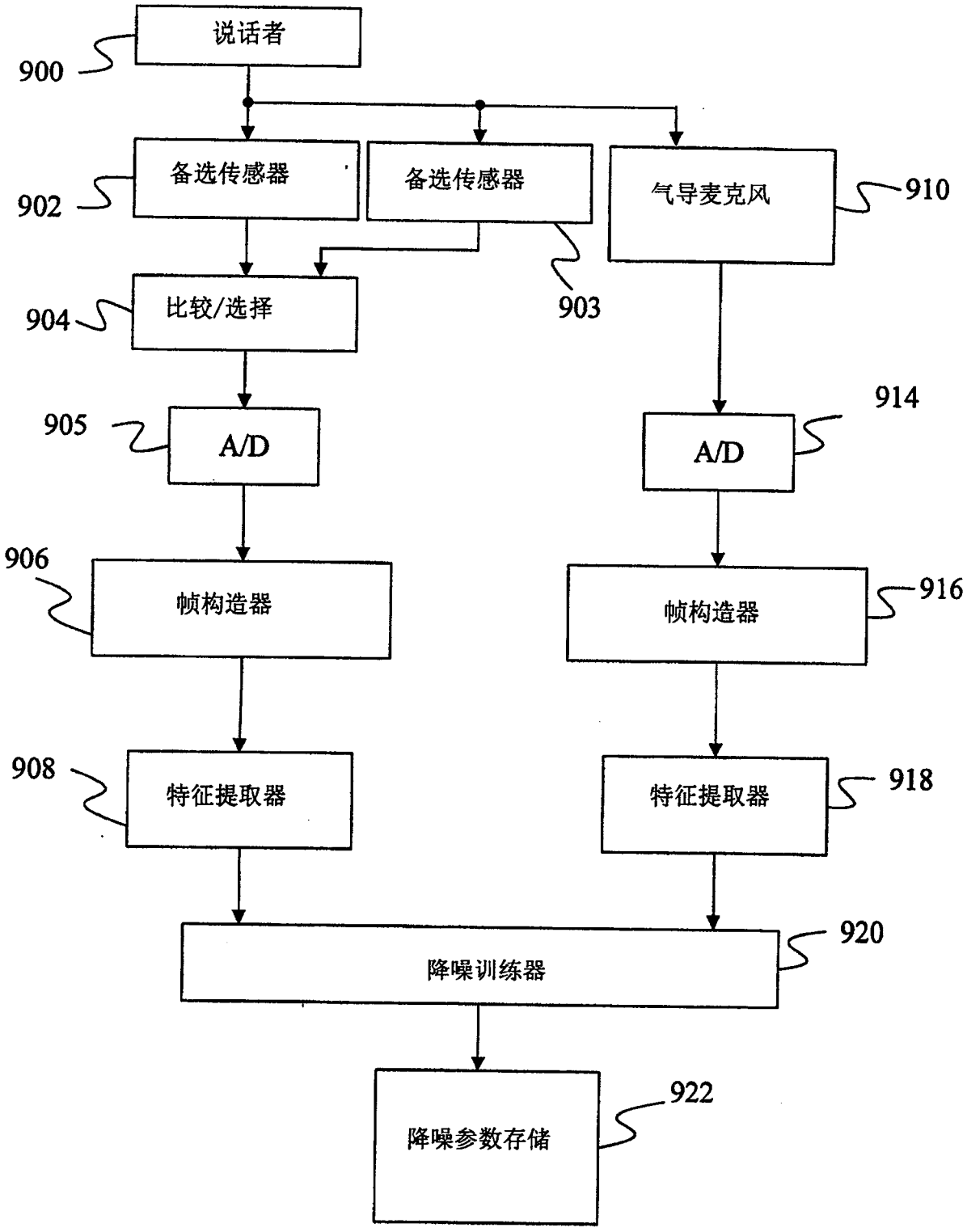

说话者
900
备选传感器
902
备选传感器
903
气导麦克风
910
比较/选择
904
A/D
905
A/D
914
帧构造器
906
帧构造器
916
特征提取器
908
特征提取器
918
降噪训练器
920
降噪参数存储
922

图 9

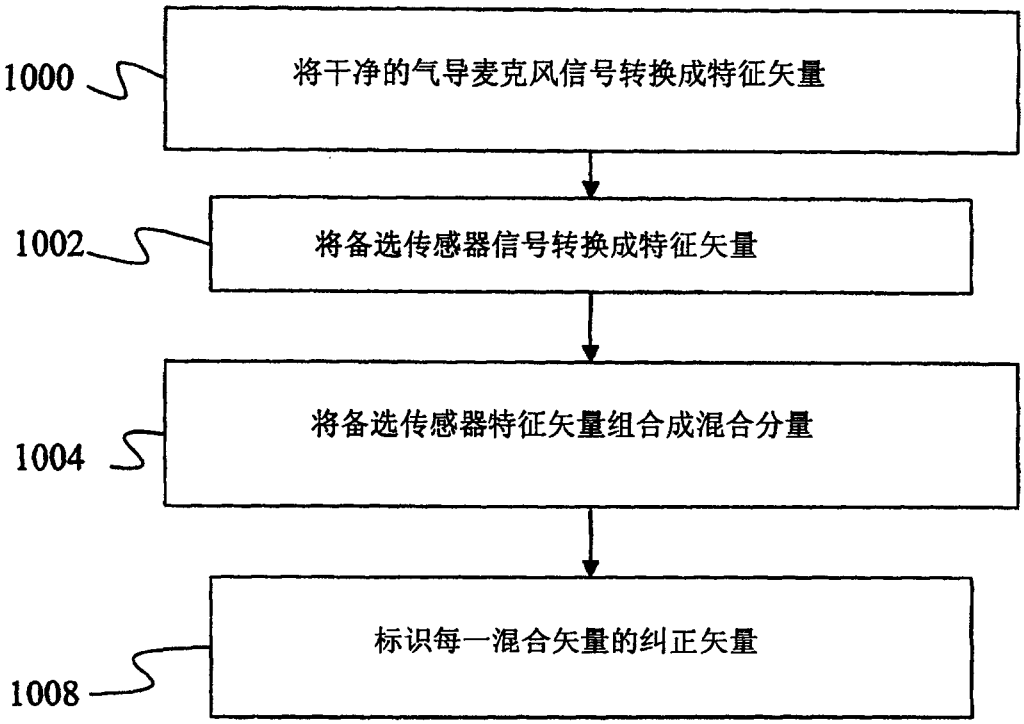

1000
将干净的气导麦克风信号转换成特征矢量
1002
将备选传感器信号转换成特征矢量
1004
将备选传感器特征矢量组合成混合分量
1008
标识每一混合矢量的纠正矢量

图 10

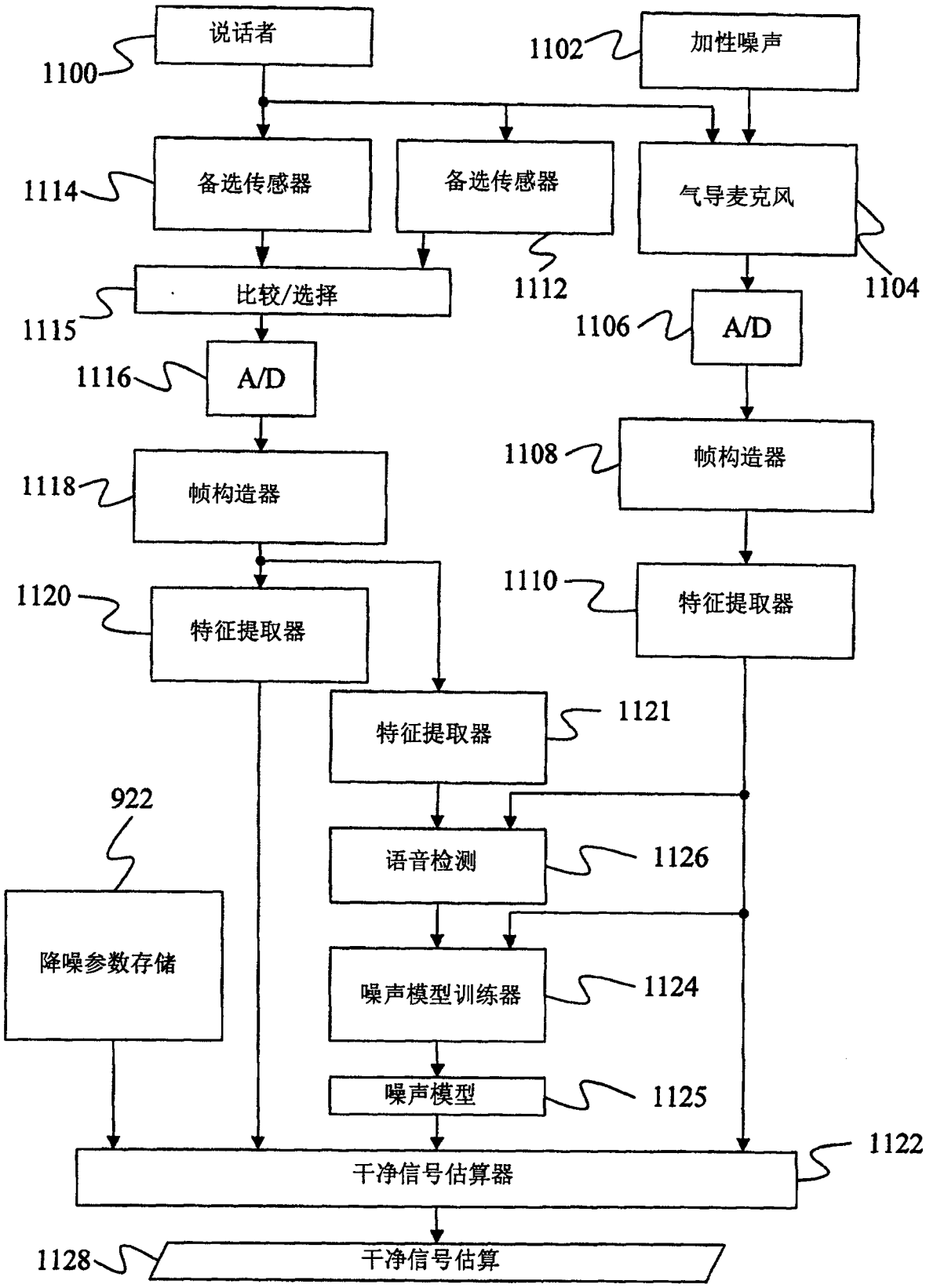

说话者
1100
加性噪声
1102
备选传感器
1114
备选传感器
1112
气导麦克风
1104
比较/选择
1115
A/D
1106
A/D
1116
帧构造器
1108
帧构造器
1118
特征提取器
1110
特征提取器
1120
特征提取器
1121
922
语音检测
1126
降噪参数存储
噪声模型训练器
1124
噪声模型
1125
干净信号估算器
1122
干净信号估算
1128

图 11

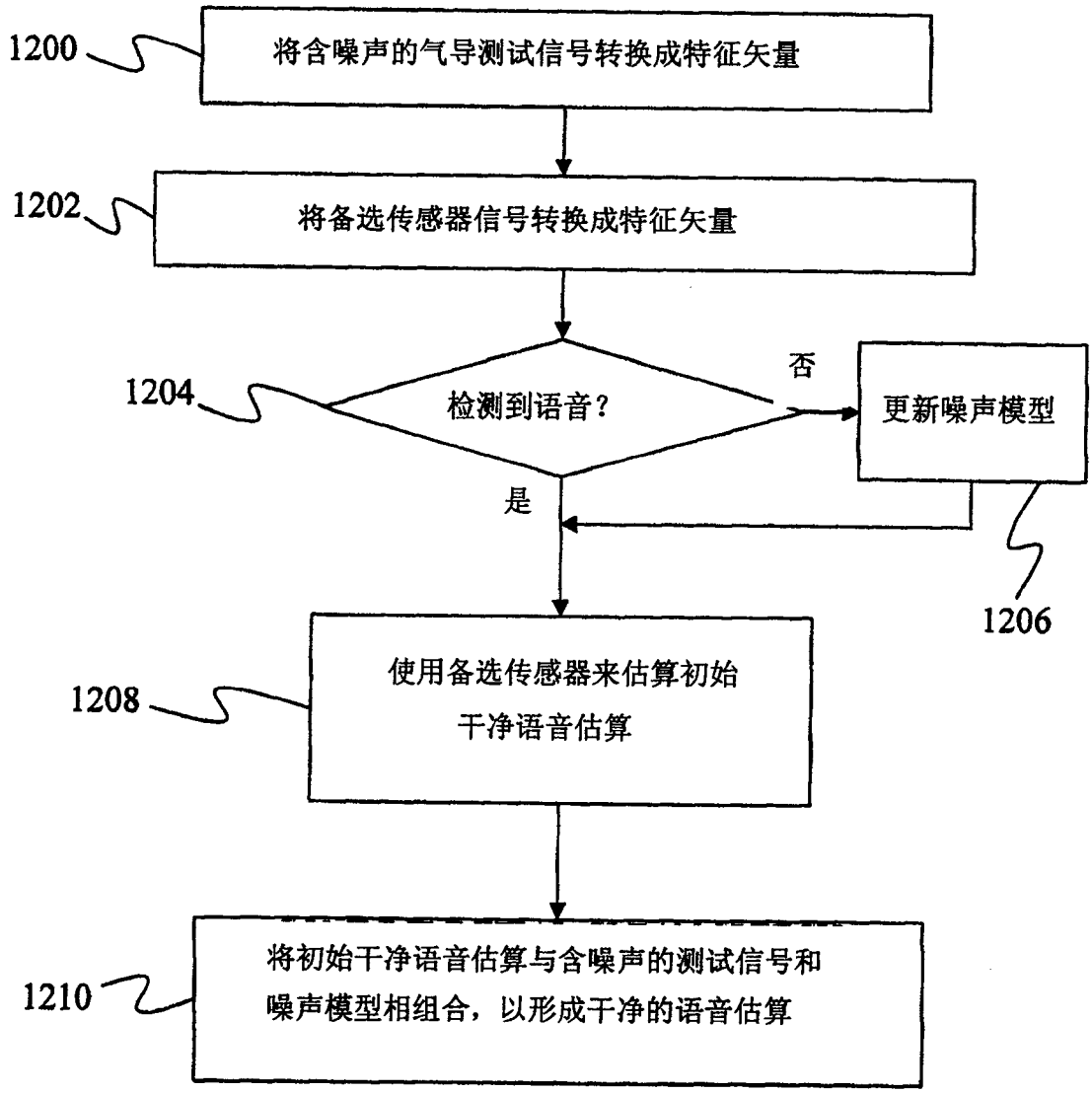

1200
将含噪声的气导测试信号转换成特征矢量
1202
将备选传感器信号转换成特征矢量
1204
检测到语音？
否
更新噪声模型
1206
是
1208
使用备选传感器来估算初始
干净语音估算
1210
将初始干净语音估算与含噪声的测试信号和
噪声模型相组合，以形成干净的语音估算

图 12

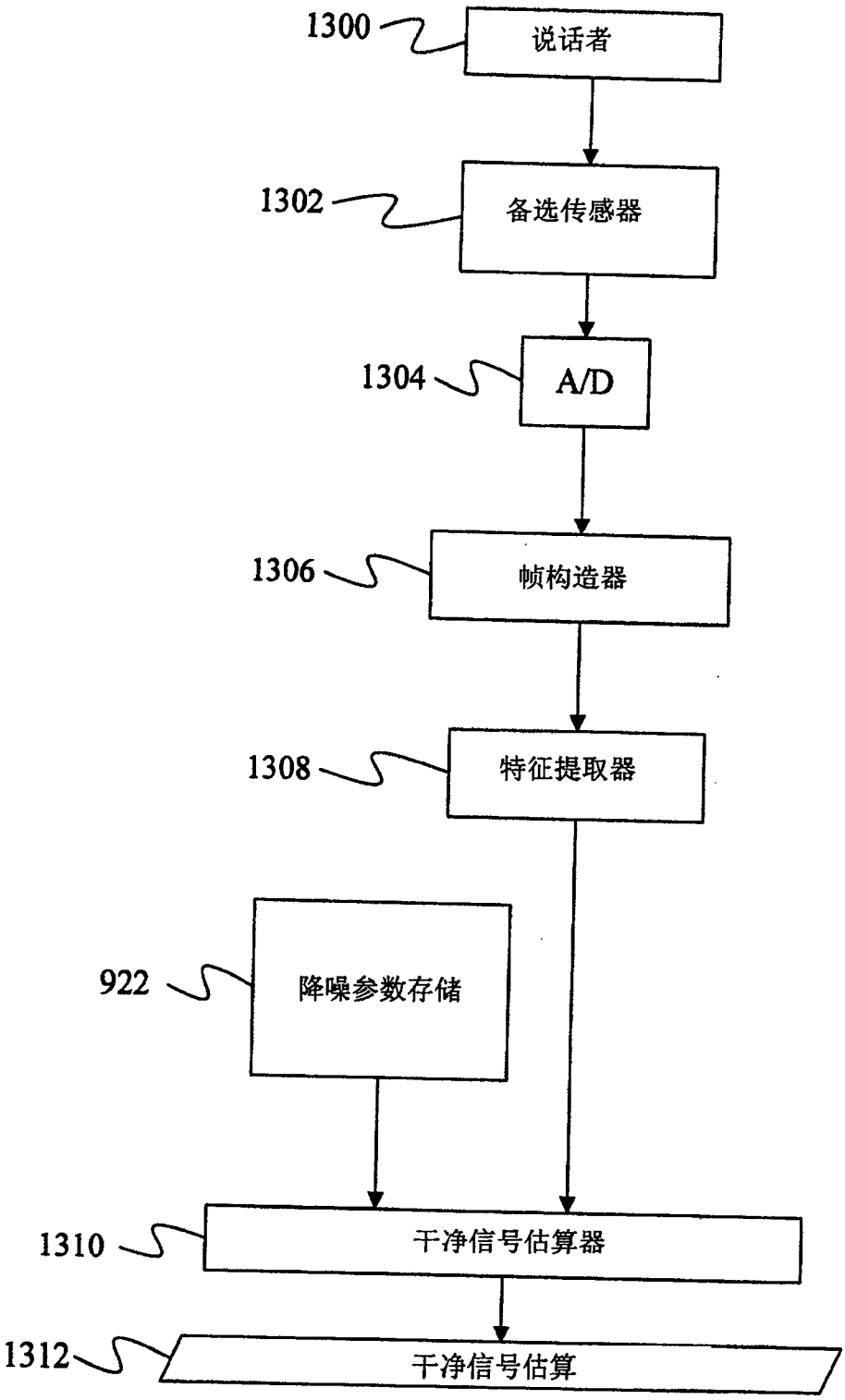

1300
说话者
1302
备选传感器
1304
A/D
1306
帧构造器
1308
特征提取器
922
降噪参数存储
1310
干净信号估算器
1312
干净信号估算

图 13

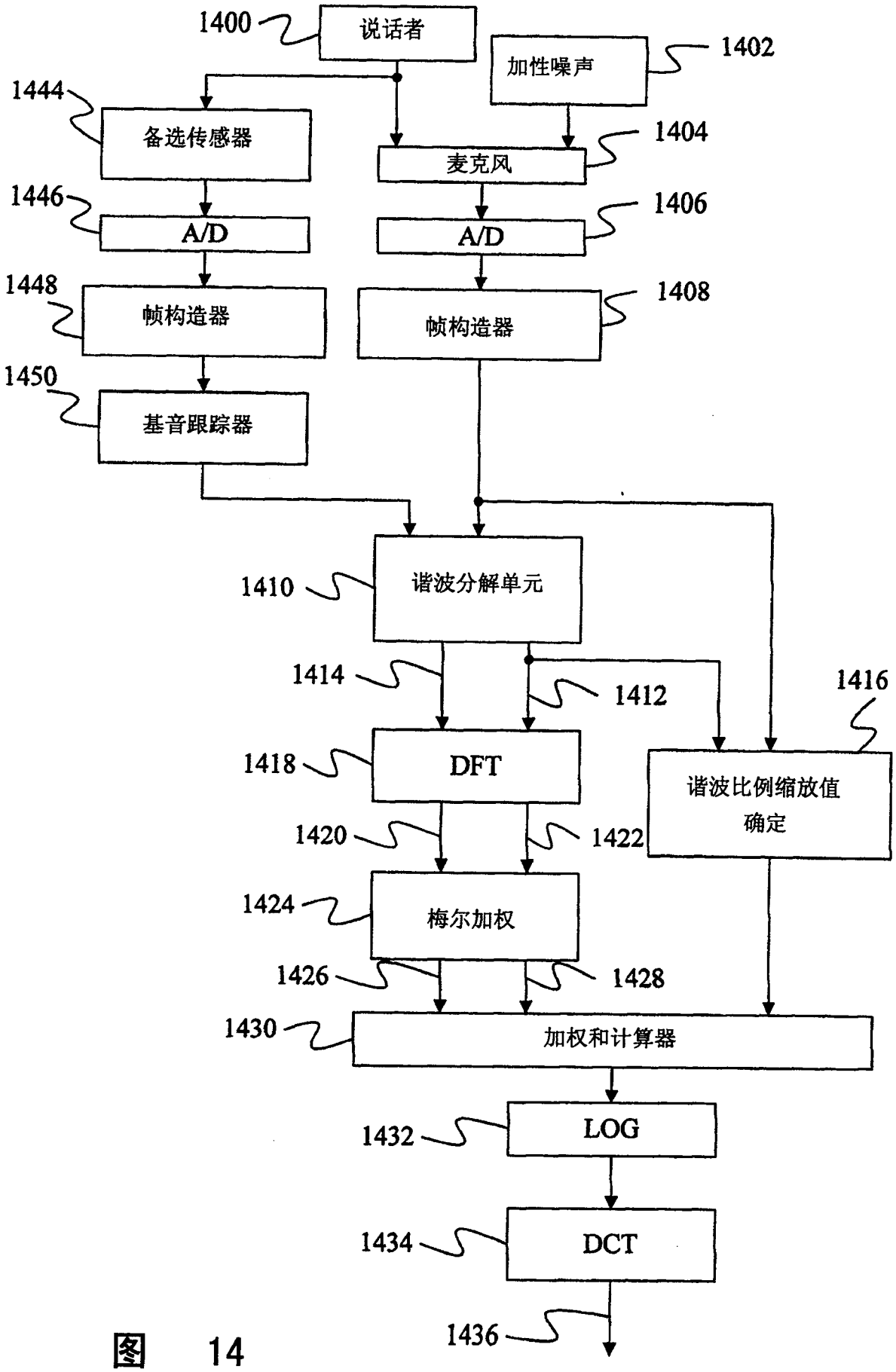

1400
说话者
加性噪声
1402
1444
备选传感器
麦克风
1404
1446
A/D
A/D
1406
1448
帧构造器
帧构造器
1408
1450
基音跟踪器
1410
谐波分解单元
1414
1412
1416
1418
DFT
谐波比例缩放值确定
1420
1422
1424
梅尔加权
1426
1428
1430
加权和计算器
1432
LOG
1434
DCT
1436

图 14

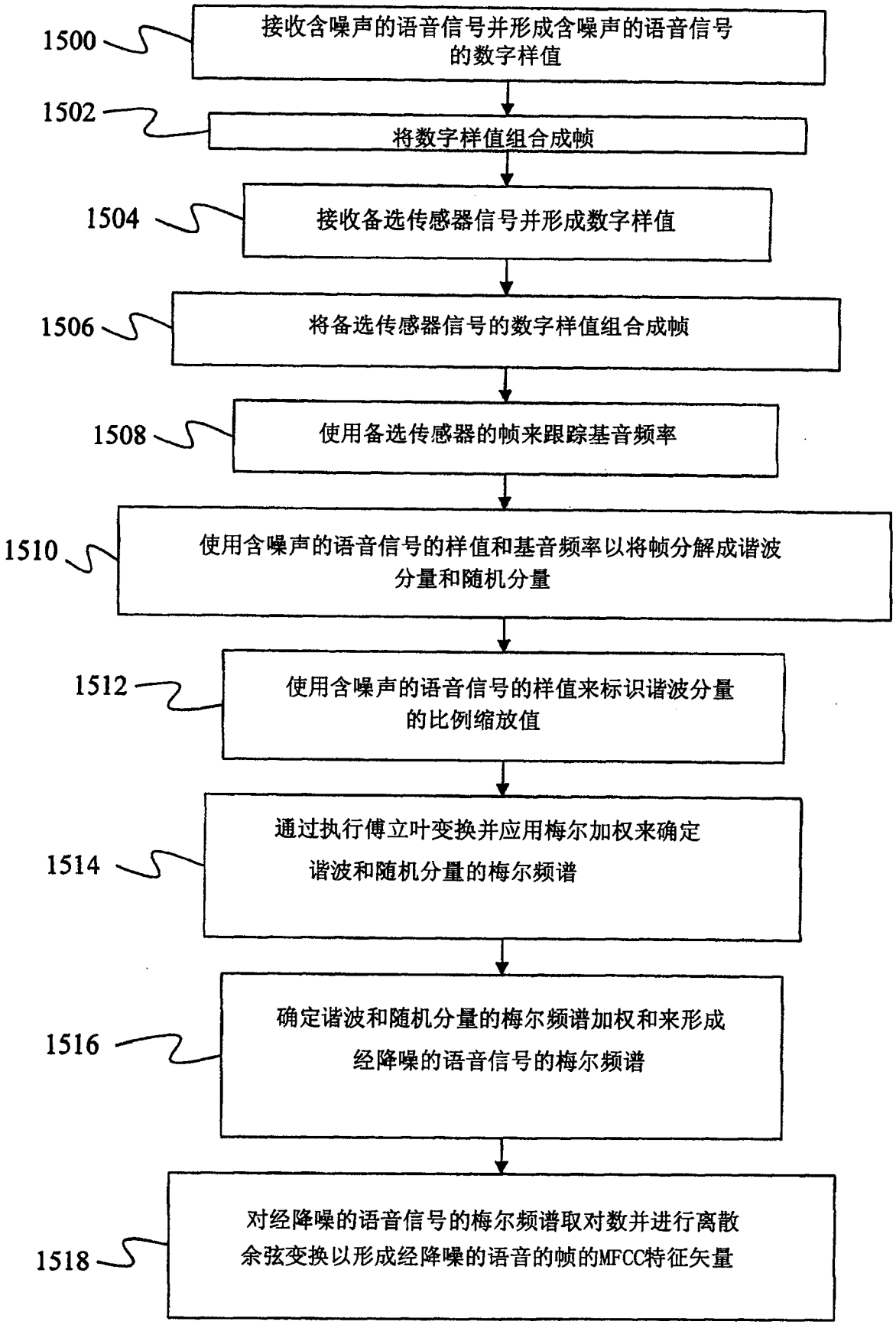

1500
接收含噪声的语音信号并形成含噪声的语音信号的数字样值
1502
将数字样值组合成帧
1504
接收备选传感器信号并形成数字样值
1506
将备选传感器信号的数字样值组合成帧
1508
使用备选传感器的帧来跟踪基音频率
1510
使用含噪声的语音信号的样值和基音频率以将帧分解成谐波分量和随机分量
1512
使用含噪声的语音信号的样值来标识谐波分量的比例缩放值
1514
通过执行傅立叶变换并应用梅尔加权来确定谐波和随机分量的梅尔频谱
1516
确定谐波和随机分量的梅尔频谱加权和来形成经降噪的语音信号的梅尔频谱
1518
对经降噪的语音信号的梅尔频谱取对数并进行离散余弦变换以形成经降噪的语音的帧的MFCC特征矢量

图 15

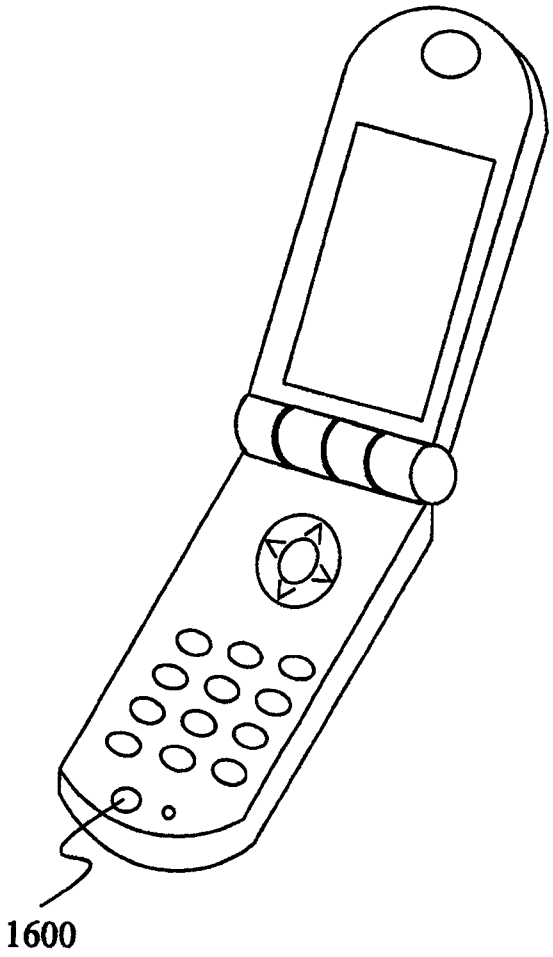



图 16