



(12)发明专利

(10)授权公告号 CN 104254819 B

(45)授权公告日 2017.09.08

(21)申请号 201380022351.6

(22)申请日 2013.05.06

(65)同一申请的已公布的文献号

申请公布号 CN 104254819 A

(43)申请公布日 2014.12.31

(30)优先权数据

61/645,818 2012.05.11 US

13/674,690 2012.11.12 US

(85)PCT国际申请进入国家阶段日

2014.10.27

(86)PCT国际申请的申请数据

PCT/US2013/039624 2013.05.06

(87)PCT国际申请的公布数据

W02013/169618 EN 2013.11.14

(73)专利权人 高通股份有限公司

地址 美国加利福尼亚州

(72)发明人 金莱轩 辛钟元 埃里克·维瑟

(74)专利代理机构 北京律盟知识产权代理有限公司 11287

代理人 宋献涛

(51)Int.Cl.

G06F 3/01(2006.01)

G10L 25/48(2013.01)

(56)对比文件

US 2006/0104458 A1,2006.05.18,

WO 2012/001928 A1,2012.01.05,

CN 1890622 A,2007.01.03,

US 2007/0192103 A1,2007.08.16,

GB 2342802 B,2003.04.16,

Harley Hung等.Estimating Dominance in

Multi-Party Meeting Using Speaker

Diarization.《Audio Speech & Language

Processing IEEE Transaction on》.2011,

审查员 杜琳琳

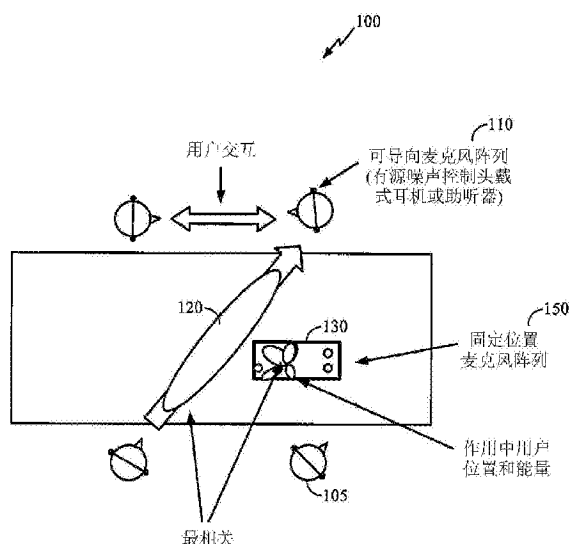
权利要求书2页 说明书21页 附图35页

(54)发明名称

音频用户交互辨识和上下文精炼

(57)摘要

一种跟踪多个参与者之间的社交交互的系统包含:固定波束成形器,其适于输出第一经空间滤波输出,且经配置以从多个可导向波束成形器接收多个第二经空间滤波输出。每一可导向波束成形器输出所述第二经空间滤波输出中与所述参与者中的不同一者相关联的相应一者。所述系统还包含:处理器,其能够确定所述第一经空间滤波输出与所述第二经空间滤波输出中的每一者之间的相似性。所述处理器基于所述第一经空间滤波输出与所述第二经空间滤波输出中的每一者之间的所述相似性确定所述参与者之间的所述社交交互。



1. 一种用于识别用户群组的主动说话用户的设备,所述设备包括:
固定麦克风阵列,其经配置以:
提供第一组经空间滤波输出,所述第一组经空间滤波输出对应于所述用户群组的多个主动说话用户,所述第一组经空间滤波输出指示多个主动说话用户的数目;以及
从多个可导向麦克风阵列接收第二组经空间滤波输出,所述多个可导向麦克风阵列中的每一者经配置以输出所述第二组经空间滤波输出中的各自一者且与所述用户群组的相应用户相关联;
处理器,其经配置以:
确定所述第一组经空间滤波输出与所述第二组经空间滤波输出之间的相似性;
基于所述相似性输出多个说话者识别符ID,所述多个说话者识别符中的每一说话者识别符对应于所述用户群组的不同主动说话用户;
基于所述相似性确定所述多个用户之间的社交交互;以及
基于所述社交交互识别与每一可导向麦克风阵列相关联的参与状态。
2. 根据权利要求1所述的设备,其中所述处理器进一步经配置以基于所述第二组经空间滤波输出中的至少一经空间滤波输出来精炼所述固定麦克风阵列的所述第一组经空间滤波输出。
3. 根据权利要求1所述的设备,其中所述处理器进一步经配置以计算所述固定麦克风阵列的所述第一组经空间滤波输出与来自所述多个可导向麦克风阵列的第一可导向麦克风阵列的特定经空间滤波输出之间的相关。
4. 根据权利要求1所述的设备,其中所述固定麦克风阵列包含在第一移动装置内,且所述多个可导向麦克风阵列的第一可导向麦克风阵列包含在不同于所述第一移动装置的第二移动装置内。
5. 根据权利要求1所述的设备,其中所述处理器进一步经配置以基于以下各项中的至少一者来确定所述相似性:相关,在时域或频域中具有可允许时间调整的最小二乘法拟合,使用线性预测译码LPC或梅尔频率倒谱系数MFCC的基于特征的方法,或使用交叉累积量、经验库贝克-李柏发散或板仓-斋藤距离的基于较高阶的方法。
6. 根据权利要求1所述的设备,其中所述处理器进一步经配置以确定所述用户群组中的至少一者的位置。
7. 根据权利要求1所述的设备,其中所述固定麦克风阵列进一步经配置以:
执行围绕所述固定麦克风阵列的多个到达方向DOA估计;
基于所述多个DOA估计确定来自所述用户群组的第一主动说话用户;以及
产生对应于所述第一主动说话用户的特定经空间滤波输出,所述特定经空间滤波输出包含在所述第一组经空间滤波输出中。
8. 根据权利要求7所述的设备,其中与确定所述相似性有关的所述处理器经配置以:
将对应于所述第一主动说话用户的所述特定经空间滤波输出与来自所述多个可导向麦克风阵列的所述第二组经空间滤波输出中的每一相应经空间滤波相比较。
9. 根据权利要求8所述的设备,其中为了将所述特定经空间滤波输出与所述第二组经空间滤波输出中的每一相应经空间滤波相比较,所述处理器经配置以:
对所述特定经空间滤波输出与所述第二组经空间滤波输出中的每一相应经空间滤波

执行交叉相关；

基于所述交叉相关确定所述特定经空间滤波输出与所述多个可导向麦克风阵列的第一可导向麦克风阵列的相应经空间滤波输出之间的交叉相关峰；以及

确定与所述交叉相关峰相关联的相关角度，其中所述相关角度对应于所述固定麦克风阵列和所述第一可导向麦克风阵列之间的角度。

10. 一种用于识别用户群组的主动说话用户的方法，所述方法包括：

从固定麦克风阵列接收第一组经空间滤波输出，所述第一组经空间滤波输出对应于所述用户群组的多个主动说话用户，所述第一组经空间滤波输出指示多个主动说话用户的数目；

从多个可导向麦克风阵列接收第二组经空间滤波输出，所述多个可导向麦克风阵列中的每一者经配置以输出所述第二组经空间滤波输出中的各自一者且与所述用户群组的相应用户相关联；

由处理器确定所述第一组经空间滤波输出与所述第二组经空间滤波输出之间的相似性；

基于所述相似性输出多个说话者识别符ID，所述多个说话者识别符中的每一说话者识别符对应于所述用户群组的不同主动说话用户；

基于所述相似性确定所述多个用户之间的社交交互；以及

基于所述社交交互识别与每一可导向麦克风阵列相关联的参与状态。

11. 根据权利要求10所述的方法，其进一步包括由所述处理器基于所述第二组经空间滤波输出中的至少一经空间滤波输出来精炼所述固定麦克风阵列的所述第一组经空间滤波输出。

12. 根据权利要求10所述的方法，其中确定所述相似性包括计算所述固定麦克风阵列的所述第一组经空间滤波输出与来自所述多个可导向麦克风阵列的第一可导向麦克风阵列的特定经空间滤波输出之间的相关。

13. 根据权利要求10所述的方法，其中所述处理器进一步基于以下各项中的至少一者来确定所述相似性：在时域或频域中具有可允许时间调整的最小二乘法拟合，使用线性预测译码LPC或梅尔频率倒谱系数MFCC的基于特征的方法，或使用交叉累积量、经验库贝克-李柏发散或板仓-斋藤距离的基于较高阶的方法。

14. 根据权利要求10所述的方法，其进一步包括基于所述相似性确定所述用户群组中的第一用户的位置。

音频用户交互辨识和上下文精炼

[0001] 相关申请案的交叉参考

[0002] 本申请案依据35U.S.C. §119 (e) 的权益主张2012年5月11日申请且标题为“音频用户交互辨识和上下文精炼 (AUDIO USER INTERACTION RECOGNITION AND CONTEXT REFINEMENT)”的第61/645,818号美国临时专利申请案的优先权,且主张2012年11月12日申请且标题为“音频用户交互辨识和上下文精炼 (AUDIO USER INTERACTION RECOGNITION AND CONTEXT REFINEMENT)”的第13/674,690号美国非临时专利申请案的优先权,以上申请案的内容明确地以引用方式并入本文。

背景技术

[0003] 通过在不同时间点确定用户正看着的方向可推导大量有用信息,且此信息可用以增强用户与多种计算系统的交互。因此,常见的是已经采取大量使用基于视觉的方法进行的凝视跟踪研究(即,使用若干不同装置跟踪眼睛)。然而,了解用户的凝视方向仅给出关于用户关注的一个维度的语义信息,且不考虑大部分由语音给出的上下文信息。换句话说,与语音跟踪结合的凝视跟踪的组合将在多种不同用户应用中提供较富含的且较有意义的信息。

发明内容

[0004] 使用基于音频的方法确定上下文信息(即,正发送或由用户接收的非视觉信息)。通过朝向特定人或特定声音源导向音频波束可增强接收侧上的音频用户交互。本文描述的技术因此可允许用户较清楚地理解例如对话的上下文。为了实现这些益处,来自一或多个可导向麦克风阵列的输入和来自固定麦克风阵列的输入可用以在正呈现基于音频的上下文信息(或甚至基于视觉的语义信息)的情况下确定一个人正看着谁或一个人相对于正在说话的人注意什么内容。

[0005] 对于各种实施方案,使用两种不同类型的麦克风阵列装置(MAD)。第一类型的MAD是可导向麦克风阵列(本文也称为可导向阵列),其由用户佩戴在关于用户的眼睛的已知定向上,且多个用户可各自佩戴可导向阵列。第二类型的MAD是固定位置麦克风阵列(本文也称为固定阵列),其放置于与用户(其中一或多者正使用可导向阵列)相同的声学空间中。

[0006] 对于某些实施方案,可导向麦克风阵列可为有源噪声控制(ANC)头戴式耳机或助听器的部分。可存在多个可导向阵列,其各自与例如会议或群组中的不同用户或说话者(本文也称为参与者)相关联。在此上下文中,固定麦克风阵列将随后用以使用音频波束来分离在群组会议期间说话和收听的不同人,所述音频波束对应于所述不同人相对于所述固定阵列定位的方向。

[0007] 评估固定阵列的经分离说话者的音频波束与可导向阵列的输出之间的相关或相似性。相关是相似性量度的一个实例,但可使用若干相似性测量或确定技术中的任一者。

[0008] 在一实施方案中,固定阵列的经分离参与者的音频波束与可导向阵列的输出之间的相似性量度可用以跟踪参与者之间的社交交互,包含在不同参与者说话或呈现基于音频

的信息时参与者随着时间的凝视方向。

[0009] 在一实施方案中,固定阵列的经分离参与者的音频波束与可导向阵列的输出之间的相似性量度可用以例如放大目标参与者。此缩放又可能在一个用户(在所述时刻是收听者)正凝视着正提供基于音频的信息(即,说话)的另一个人时导致增强的噪声滤波和放大。

[0010] 在一实施方案中,固定阵列的经分离参与者的音频波束与可导向阵列的输出之间的相似性量度可用以自适应地形成目标参与者的较好波束,从而实际上较好地确定用户中的每一者相对于彼此的物理定向。

[0011] 提供此概述以用简化形式介绍概念的选择,所述概念在以下详细描述中进一步描述。此概述既定不识别所主张标的物的关键特征或本质特征,也既定不用以限制所主张标的物的范围。

附图说明

[0012] 当结合附图阅读时更好地理解前述概述以及以下对说明性实施例的详细描述。为了说明实施例的目的,在图中展示实施例的实例性构造;然而,实施例不限于所揭示的特定方法和手段。在图中:

[0013] 图1是可用以确定上下文信息的各自佩戴可导向麦克风阵列的用户群组连同固定麦克风阵列的图;

[0014] 图2是使用可导向麦克风阵列和固定麦克风阵列确定用户交互的方法的实施方案的操作流程;

[0015] 图3是使用可导向麦克风阵列和固定麦克风阵列确定用户交互的方法的另一实施方案的操作流程;

[0016] 图4是可提供用户身份和用户正观看哪一方向的指示的实例性显示的图;

[0017] 图5是可产生和显示且指示各种用户交互和会议数据的用户接口的图;

[0018] 图6是可产生且显示(例如,在智能电话显示器上)且指示各种用户交互(例如,在会议期间)的用户接口的实例性显示的图;

[0019] 图7是指示相对于各种话题的各种用户交互的实例性显示的图;

[0020] 图8是指示随着时间的各种用户交互的实例性显示的图;

[0021] 图9是指示随着时间的各种用户交互的另一实例性显示的图;

[0022] 图10是使用交叉相关测量相似性的方法的实施方案的操作流程;

[0023] 图11是使用交叉累积量测量相似性的方法的实施方案的操作流程;

[0024] 图12是使用时域最小二乘法拟合测量相似性的方法的实施方案的操作流程;

[0025] 图13是使用频域最小二乘法拟合测量相似性的方法的实施方案的操作流程;

[0026] 图14是使用板仓-斋藤距离测量相似性的方法的实施方案的操作流程;

[0027] 图15是使用基于特征的方法测量相似性的方法的实施方案的操作流程;

[0028] 图16展示实例性用户接口显示;

[0029] 图17展示示范性用户接口显示以展示显示上的合作缩放;

[0030] 图18是用于放大目标参与者的方法的实施方案的操作流程;

[0031] 图19展示具有额外候选观看方向的实例性用户接口显示;

[0032] 图20是用于自适应地精炼目标说话者的波束的方法的实施方案的操作流程;

- [0033] 图21展示相对于麦克风对的平面波传播的远场模型；
- [0034] 图22展示线性阵列中的多个麦克风对；
- [0035] 图23展示四个不同DOA的未缠绕相位延迟对频率的绘图，且图24展示相同DOA的缠绕相位延迟对频率的绘图；
- [0036] 图25展示两个DOA候选的测得相位延迟值和所计算值的实例；
- [0037] 图26展示沿着电视机屏幕的顶部边限布置的线性麦克风阵列；
- [0038] 图27展示计算帧的DOA差的实例；
- [0039] 图28展示计算DOA估计的实例；
- [0040] 图29展示针对每一频率识别DOA估计的实例；
- [0041] 图30展示使用所计算可能性来针对给定频率识别最佳麦克风对和最佳DOA候选的实例；
- [0042] 图31展示可能性计算的实例；
- [0043] 图32展示扬声器电话应用的实例；
- [0044] 图33展示逐对DOA估计到麦克风阵列的平面中的 360° 范围的映射；
- [0045] 图34和35展示DOA估计中的模糊性；
- [0046] 图36展示观测到的DOA的正负号与x-y平面的象限之间的关系；
- [0047] 图37到40展示其中源位于麦克风的平面上方的实例；
- [0048] 图41展示沿着非正交轴的麦克风对的实例；
- [0049] 图42展示使用图41的阵列获得相对于正交x和y轴的DOA估计的实例；
- [0050] 图43和44展示针对两对麦克风阵列(例如,如图45中所示)的逐对正规化波束成形器/空波束成形器(BFNF)的实例；
- [0051] 图46展示逐对正规化最小方差无失真响应(MVDR)BFNF的实例；
- [0052] 图47展示其中矩阵 $A^H A$ 并非条件不良的频率的逐对BFNF的实例；
- [0053] 图48展示导向向量的实例；以及
- [0054] 图49展示如本文描述的源方向估计的集成方法的流程图。

具体实施方式

[0055] 除非由其上下文明确限制,否则术语“信号”在此用以指示其普通意义中的任一者,包含如在电线、总线或其它传输媒体上表示的存储器位置(或存储器位置集合)的状态。除非由其上下文明确限制,否则术语“产生”在此用以指示其普通意义中的任一者,例如计算或以其它方式产生。除非由其上下文明确限制,否则术语“计算”在此用以指示其普通意义中的任一者,例如计算、评估、估计和/或从多个值中选择。除非由其上下文明确限制,否则术语“获得”用以指示其普通意义中的任一者,例如计算、推导、接收(例如,从外部装置)和/或检索(例如,从存储元件阵列)。除非由其上下文明确限制,否则术语“选择”用以指示其普通意义中的任一者,例如识别、指示、应用和/或使用两者或两者以上的集合中的至少一者且少于全部。在本发明描述和权利要求书中使用术语“包括”的情况下,其并不排除其它元件或操作。术语“基于”(如“A基于B”中)用以指示其普通意义中的任一者,包含如下情况:(i)“推导自”(例如,“B是A的前体”),(ii)“至少基于”(例如,“A至少基于B”),以及在特定上下文中适当的情况下,(iii)“等于”(例如,“A等于B”或“A与B相同”)。类似地,术语“响

应于”用以指示其普通意义中的任一者,包含“至少响应于”。

[0056] 对多麦克风音频感测装置的麦克风的“位置”的参考指示所述麦克风的声学敏感面的中心的位置,除非上下文另外指示。根据特定上下文,术语“通道”有时候用以指示信号路径且在其它时候用以指示由此路径载运的信号。除非另外指示,否则术语“系列”用以指示两个或两个以上项目的序列。术语“对数”用以指示基于十的对数,但此运算向其它基数的扩展在本发明的范围内。术语“频率分量”用以指示信号的一组频率或频带当中的一者,例如所述信号的频域表示的样本(或“区间”)(例如,由快速傅立叶变换产生)或所述信号子带(例如,巴克(Bark)尺度或梅尔(mel)尺度子带)。

[0057] 除非另外指示,否则对具有特定特征的设备的操作的任何揭示也明确预期揭示具有类似特征的方法(且反之亦然),且对根据特定配置的设备的操作的任何揭示也明确预期揭示根据类似配置的方法(且反之亦然)。术语“配置”可参考由其特定上下文指示的方法、设备和/或系统来使用。术语“方法”、“过程”、“程序”和“技术”一股地且可互换地使用,除非特定上下文另外指示。术语“设备”和“装置”也一股地且可互换地使用,除非特定上下文另外指示。术语“元件”和“模块”通常用以指示较大配置的一部分。除非由其上下文明确限制,否则术语“系统”在此用以指示其普通意义中的任一者,包含“相互作用以用于共同目的的元件群组”。

[0058] 通过引用文档的一部分的任何并入也应当理解为并入了在所述部分内参考的术语或变量的定义,这些定义在文档中的别处出现的地方,以及在所并入部分中参考的任何图式。除非初始通过定冠词介绍,否则用以修改权利要求元素的序数术语(例如,“第一”、“第二”、“第三”等等)本身并不指示所述权利要求元素相对于另一元素的任何优先级或次序,而是仅使所述权利要求元素区别于具有相同名称(但出于序数术语的使用)的另一权利要求元素。除非通过其上下文明确限制,否则术语“多个”和“集合”中的每一者在本文用以指示大于一的整数量。

[0059] 本文描述组合基于视觉和听力的方法以使得用户能够朝向人(或声音源)导向,以便使用声音传感器和多种基于位置的计算和所得的交互增强来更清楚地理解在所述时刻正呈现的基于音频的信息(例如,对话的上下文和/或声音源的身份)。

[0060] 举例来说,固定阵列的经分离说话者的音频束与可导向阵列的输出之间的相关或相似性可用以跟踪说话者之间的社交交互。相关仅是相似性量度的一个实例,且可使用任何相似性测量或确定技术。

[0061] 更特定来说,用户(本文也称为说话者或参与者)群组的社交交互或社交网络分析可使用响应于分别与群组的每一用户相关联的固定阵列的经分离说话者的音频束与每一可导向阵列的输出之间的相关或其它相似性量度而产生的连接曲线图来执行和显示。因此,举例来说,可使用会议参与者之间的连接曲线图在参与者的群组会议中执行自动社交网络分析以推导关于谁主动参加呈现或(更一股地)保持用户的注意的呈现的有效性的有用信息。

[0062] 图1是可用以确定上下文信息的各自佩戴可导向麦克风阵列110的用户群组连同与用户在相同空间(例如,房间)中的固定位置麦克风阵列150的图100。如图1所示,房间(或其它经界定空间)中的用户群组的每一用户105佩戴可导向麦克风阵列(例如,作为可包含执行自适应噪声控制(ANC)的能力的头戴式耳机),且固定位置麦克风阵列150位于房间中

(例如,桌子上、电话中等等)。固定位置麦克风阵列150可例如为例如视频游戏平台、平板计算机、笔记本计算机或智能电话等电子装置的部分,或可为独立装置或实施方案。替代地或另外,固定位置麦克风阵列150可包括分布式麦克风阵列(即,分布式麦克风)。

[0063] 佩戴头戴式耳机的用户105可从指向用户的物理视觉(或“观看”)方向的他的可导向(例如,可佩戴)麦克风阵列产生固定波束模式120。如果用户转动他的头,那么用户的波束模式的观看方向也改变。作用中说话者的位置可使用固定麦克风阵列来确定。通过从具有对应于每一作用中说话者的固定麦克风阵列输出的可导向麦克风阵列相关经波束成形输出(或任一类型的空间经滤波输出)或另外确定其相似性,可确定用户正看着(例如,注意着、听着等等)的人的识别。每一头戴式耳机可具有与主处理器(例如,集中式本地或远程计算装置中)通信(例如,经由无线通信链路)以分析头戴式耳机和/或固定阵列之间的波束的相关或相似性的处理器。

[0064] 换句话说,在任一时间时刻的固定波束模式可基于可与固定麦克风阵列输出相关的用户的物理观看方向来形成,进而经由连接曲线图130(例如,在例如手持机、膝上型计算机、平板计算机、计算机、上网本或移动计算装置等任一类型的计算装置的显示器上显示)提供目标用户的社交交互的视觉指示。因此,通过使来自可导向麦克风阵列的经波束成形输出与对应于每一作用中说话用户的固定麦克风阵列输出相关,可执行和显示社交交互或网络分析的跟踪。而且,通过检查来自观看方向可导向麦克风阵列的经波束成形输出与对应于每一作用中说话者的位置固定麦克风阵列输出之间的相似性,用户正看着或注意着的人可被识别且放大。

[0065] 图2是使用可导向麦克风阵列和固定麦克风阵列确定用户交互的方法200的实施方案的操作流程。在210处,可导向麦克风阵列和固定麦克风阵列大致在同时各自接收声音(但可检测小的变化且用以计算用户的相对位置)。在220处,由可导向麦克风阵列和固定麦克风阵列中的每一者产生例如经波束成形输出等经空间滤波输出。在230处,将每一可导向麦克风阵列的经空间滤波输出与固定麦克风阵列的经空间滤波输出进行比较。可使用用于确定相似性或相关的任何已知技术。在240处,从230获得的相似性或相关信息可用以确定和/或显示用户交互信息,如本文进一步描述。

[0066] 图3是使用可导向麦克风阵列和固定位置麦克风阵列确定用户交互的方法300的另一实施方案的操作流程。多个用户中的每一者具有可导向立体麦克风阵列,例如ANC头戴式耳机,其具有对应于每一此用户的视觉凝视方向的已知定向。在305处可导向阵列(ANC头戴式耳机中)中的每一者提供固定边射波束成形,其中在310处在用户观看方向上(即,在可导向阵列的用户正观看的方向上)产生经波束成形输出(或任一类型的经空间滤波输出)。

[0067] 具有相关联处理器的固定麦克风阵列(例如在智能电话中)在围绕固定麦克风阵列的三个维度(3D)中在320处执行到达方向(DOA)估计,且在325处分离作用中说话者。在370处确定作用中说话者的数目,且在380处产生每一作用中说话者的单独输出(由例如识别号识别)。在一实施方案中,可在330处执行作用中说话者的说话者辨识和标记。

[0068] 在340处测量固定阵列的经分离说话者与可导向阵列的输出之间的相似性。使用测得的相似性和DOA估计和说话者ID,可在350处产生和显示用户交互的视觉化(具有说话者身份(ID)或参与者ID)。在360处可将每一用户的观看方向作为例如智能电话坐标提供到固定阵列。

[0069] 可产生连接曲线图(也称为交互曲线图),其显示例如(a)谁正在讲话和/或听着谁和/或看着谁,(b)谁正在支配和/或领导群组的讨论,和/或(c)谁已厌烦、未参与和/或安静。可执行实时会议分析以帮助会议和未来会议的效率。例如会议时间、地点(例如,会议位置)、说话者身份或参与者身份、会议话题或标的以及参与者数目等信息例如可显示且在分析中使用。

[0070] 图4是可提供用户身份和用户正观看哪一方向的指示的实例性显示器403的图400。连同用户正观看的方向(参与者观看方向410)一起显示用户身份(参与者ID 406)。在例如会议期间,参与者观看方向410的此显示可产生且提供到关注方,例如会议管理者或领导者或监督者,使得关注方可看见在会议的各个时间所述参与者正看着谁。虽然在图403中仅展示一个参与者ID 406和参与者观看方向410,但这既定不是限制性的。关注方可接收一个以上参与者的此信息,且此信息可取决于实施方案在一或多个显示器上同时显示。用于在显示器403上显示而产生的数据可存储在存储器中且在较晚时间检索和显示,以及实时显示。

[0071] 图5是可在显示器418上产生和显示且指示各种用户交互和会议数据的用户接口的图415。可产生和显示(例如,在会议期间实时)各种类型的信息,例如正在讲话的参与者的识别符(ID) 420、正在听着的参与者的ID 422和/或未参与的参与者的ID 424(例如,在所述时刻不在听、在超过预定时间量或会议的至少一百分比的时间中不在听、看着除了正在讲话的参与者之外的某个地方或看着另一预定位置或方向等等)。在例如会议期间,此显示4108可产生且提供到关注方,例如会议管理者或领导者或监督者。

[0072] 可在显示418上显示额外数据,例如会议时间426、会议位置428、会议长度430(即,持续时间)、会议话题432以及会议参与者数目434。可显示此数据中的一些或全部。另外或替代地,取决于实施方案可显示其它数据,例如所有参与者的ID和如本文进一步描述可产生的其它统计数据。用于在显示器418上显示而产生的信息和数据可存储在存储器中且在较晚时间检索和显示,以及实时显示。

[0073] 应注意,即使参与者仅在会议中听着(且未说话),她也将是参与的,因为所述参与者的麦克风(可导向麦克风阵列)将仍在她正听着时观看的方向上拾取声音。因此,即使参与者不说话,也将仍存在与她的收听相关联的待分析的声音。

[0074] 可产生和显示用户接口(例如,在智能电话显示器或其它计算装置显示器上,例如与手持机、膝上型计算机、平板计算机、计算机、上网本或移动计算装置相关联的显示器),其指示在会议期间的各种用户交互。图4是可产生且显示(例如,在智能电话显示器443上)且指示各种用户交互(例如,在会议期间)的用户接口440的实例性显示的图。在此实例中,每一箭头454的方向指示谁正看着谁(在此实例中仅展示一个箭头454,但取决于实施方案和特定时间的用户交互可展示多个此类箭头)。每一箭头的厚度指示交互的相对强度(例如,基于连接时间等等)。从一个人或到一个人无箭头指示所述用户未涉及于群组会议中。针对用户可显示百分比数字,其指示针对群组会议的参与比率。可显示指示符448以识别会议的领导者,且可确定和显示百分比450、452以分别展示讨论中有多少被引导到一个人,且讨论中有多少是来自所述人。在一实施方案中,可使用颜色或突出显示来指示参与者群组的领导者。

[0075] 在图6的实例中,约翰(John)和马克(Mark)正进行很多交互,如相对大的厚箭头

446指示。玛丽(Mary)是安静的。可执行实时会议分析(例如上文相对于图4和5以及本文别处所描述)以帮助会议的效率。举例来说,因为看上去玛丽不在对话中,所以约翰可鼓励玛丽参与(例如,通过向玛丽问问题)。

[0076] 可在一时间周期上(例如,一个月、一年等等)累积社交交互绘图以评估例如群组动态性或话题动态性。图7是指示相对于各种话题464的各种用户交互的实例性显示462的图460。此信息可在一或多个会议期间俘获,存储在一存储器(或多个存储器)中,且在较晚时间(例如,在历史数据分析期间)以一或多个格式显示。此处,每一参与者ID 466连同其针对各种话题464的参与比率468一起列出。

[0077] 因此举例来说,简(Jane)在关于“设计”的会议中具有20%参与比率,在关于“代码走查”的会议中具有40%参与比率,且在关于“文档化”的会议中具有10%参与比率。此数据可用以确定哪些参与者最适合或关注例如特定话题,或哪些参与者可能需要相对于特定话题的更多鼓励。参与比率可经确定且基于本文描述的一或多个数据项目,例如在会议中说话的时间量、在会议中注意的时间量、在会议中收听的时间量等等。虽然在图7中展示百分比,但可使用任何相对测量、编号或指示系统或技术来识别参与水平或比率中的相对强度和/或弱度。

[0078] 图460中的“L”用作实例性指示符以指示哪一用户在某一话题中参与最多,进而指示例如所述话题的潜在领导者。可使用任何指示符,例如颜色、突出显示或特定符号。在此实例中,约翰最多地参与设计,简最多地参与代码走查,且玛丽最多地参与文档化。因此,他们可被识别为相应话题中的潜在领导者。

[0079] 另外,可针对一或多个会议参与者产生具有交互历史的个人时间线。因此,不仅可俘获、分析在会议期间的单个快照或时间周期且显示(在实时或较晚离线显示)与其有关的信息,而且可存储(例如,在例如智能电话等计算装置或例如手持机、膝上型计算机、平板计算机、计算机、上网本或移动计算装置等任一类型的计算装置的存储器中)、分析和显示(例如,在例如智能电话等计算装置或例如手持机、膝上型计算机、平板计算机、计算机、上网本或移动计算装置等任一类型的计算装置的日历或其它显示中)随着时间的历史。

[0080] 图8是可用于例如在一或多个会议之后的历史分析的指示随着时间的各种用户交互的实例性显示472的图470。此处,连同例如会议日期和会议话题等信息一起提供用户识别符474。随着时间476提供此显示472上的信息478。其针对每一周期或时刻展示信息478,例如在所述周期或时刻用户正看着谁、在那时用户是否正在说话,以及在所述周期或时刻正看着所述用户的会议参与者的百分比。此信息478可在会议期间的预定时间(例如,每分钟、每5分钟等等)确定,或经确定为例如特定时间周期上的平均值或其它经加权确定。此信息仅作为实例提供且无意为限制性的,可产生和显示额外或替代信息作为信息478。

[0081] 图8中显示的信息可用于会议分析和用户分析。因此,在图8中,可确定当简不在说话时用户简通常看着玛丽或马克,但当简在说话时简看着约翰。图8还指示当简不在说话时,看着简的参与者的百分比为零,但此百分比在简说话时增加。

[0082] 还可产生、存储、分析和显示交互统计数据。举例来说,可跟踪且显示人之间的交互演进。可使用随着时间的回归加权(例如, $0.9 \times \text{历史数据} + 0.1 \times \text{当前数据}$),使得随着数据变得较旧,其变得较不相关,其中最当前数据被加权最高(或反之亦然)。以此方式,用户可能看见他或其他人正在与哪些人联网多于其他人。可将额外统计数据作为分析中的因数

以提供更准确的交互信息。举例来说,可与会议、历史和/或参与者交互数据一起使用(组合)从电子邮件交换或其它通信获得的交互信息以提供额外(例如,较准确)交互信息。

[0083] 图9是指示随着时间的各种用户交互的另一实例性显示482的图480。此处,连同交互尺度488和时间周期一起识别用户简。图480展示其它用户ID 484和过去的月份列表486。在此实例中的交互尺度范围是从0到10,其中0表示无交互且10表示在月份486中的每一者中所识别用户与简之间的非常强的交互。此信息可产生且作为历史数据提供,且例如由会议参与者或领导者或监督者用来查看和分析随着时间的各种用户交互,以例如看谁在何时与谁最强地交互。

[0084] 作为另一实例,可执行在线学习监视以确定远程位点的学生是否正在主动参与。同样,还预期用于具有参与者交互的视频游戏的应用,其中可存在在可能的声音事件位置当中用户正看着何处的立即辨识。

[0085] 图10是方法500的实施方案的操作流程,且使用交叉相关作为示范性量度,但可使用任何相似性测量技术。在503处,固定麦克风阵列提供作用中说话者的数目N和作用中说话者的经分离语音信号。一个信号(声音)由固定麦克风阵列接收。固定麦克风阵列的输出包括波束,每一参与者对应于一个波束。因此,单独输出与每一参与者相关联。在510处,可导向麦克风阵列提供用户的观看方向。对于每一用户,个别用户的输出与从固定麦克风阵列输出的波形(或其它经空间滤波输出)中的每一者相关。

[0086] 在515处可使用此信息产生位置映射。可利用与用户何时转向某人且看着他们有关的信息。如图示可使用例如506处所示的众所周知的经典相关方程式,其中E等于期望值且c为相关值。每当存在最大峰时,就是强相关角度。在一实施方案中,可使用物理约束或系统复杂性来预定最大可允许时间移位。举例来说,仅当佩戴可导向阵列的用户在作用中时,可测量和使用可导向麦克风与固定麦克风之间的时间延迟。应注意,常规帧长度20ms对应于几乎7米。角度 θ 是作用中说话者相对于收听用户定位的相对角度。在513处可确定固定阵列与可导向阵列之间的角度 θ 。

[0087] 图11是测量相似性的方法520的实施方案的操作流程,且使用交叉累积量作为示范性量度,但可使用任何相似性测量技术。在523处,固定麦克风阵列提供作用中说话者的数目N和作用中说话者的经分离语音信号。一个信号(声音)由固定麦克风阵列接收。固定麦克风阵列的输出包括波束,每一参与者对应于一个波束。因此,单独输出与每一参与者相关联。在530处,可导向麦克风阵列提供用户的观看方向。对于每一用户,个别用户的输出与从固定麦克风阵列输出的波形(或其它经空间滤波输出)中的每一者相关。

[0088] 在525处可使用此信息产生位置映射。可利用与用户何时转向某人且看着他们有关的信息。如图示可使用526处所示的众所周知的经典累积量方程式,其中E等于期望值且c为相关值。每当存在最大峰时,就是强相关角度。角度 θ 是作用中说话者相对于收听用户定位的相对角度。在513处可确定固定阵列与可导向阵列之间的角度 θ 。

[0089] 应注意,可使用任何相似性或相关技术。关于可能的相似性量度,实际上可使用任何距离度量,例如(但不限于)以下众所周知的技术:(1)具有可允许时间调整的最小二乘法拟合:时域或频域;(2)基于特征的方法:使用线性预测译码(LPC)或梅尔频率倒谱系数(MFCC);和(3)基于较高阶的方法:交叉累积量、经验库贝克-李柏发散(Kullback-Leibler Divergence)或板仓-斋藤(Itakura-Saito)距离。

[0090] 图12是使用时域最小二乘法拟合测量相似性的方法540的实施方案的操作流程,且图13是使用频域最小二乘法拟合测量相似性的方法550的实施方案的操作流程。使用时域最小二乘法拟合的方法540类似于上述图11的方法520,不同的是替代于使用526的累积量方程,可如所示使用542处所示的时域方程式。类似地,方法550类似于图11的方法520,但不是使用能量正规化,而是使用快速傅立叶变换(FFT)结合552处所示的频域方程式。

[0091] 图14是使用板仓-斋藤距离测量相似性的方法560的实施方案的操作流程。此技术类似于图13的FFT技术,但使用562处所示的方程式。图15是使用基于特征的方法测量相似性的方法570的实施方案的操作流程。如573和575处所示执行且结合图10的其它操作503、510、513和515以及572处所示的方程式来执行特征提取。

[0092] 在一实施方案中,固定麦克风阵列的经分离说话者的音频波束与可导向麦克风阵列的输出之间的相关或相似性可用以放大目标说话者。此类型的合作缩放可提供用于放大所要说话者的用户接口。

[0093] 换句话说,可执行合作放大,其中为具有多个装置的多个用户提供用户接口以用于通过仅看着目标说话者来放大目标说话者。可经由头戴式耳机或手持机在目标人处产生波束成形,使得可组合多个装置的所有可用资源以用于合作缩放,进而增强目标人的观看方向。

[0094] 举例来说,用户可看着目标人,且可通过使用头戴式耳机或手持机(无论哪一者较靠近目标人)在目标人处产生波束成形。这可通过使用包含具有两个麦克风的隐藏相机的装置来实现。当多个装置的多个用户看着目标人时,相机可视觉上聚焦于所述人。另外,装置可通过使用(例如,全部)可用麦克风在听觉上聚焦(即,放大)所述人来增强目标人的观看方向。

[0095] 另外,可通过消除其它说话者且增强目标人的话音来在听觉上放大目标人。所述增强也可使用头戴式耳机或手持机(无论哪一者较靠近目标人)来完成。

[0096] 图16中展示示范性用户接口显示600。所述显示(例如,在智能电话显示器610或其它显示装置上显示)展示作用中用户位置620和相关联能量630。图17展示示范性用户接口显示,用以展示显示器上的合作缩放,其中说话者1如显示660中所示从初始显示650放大。

[0097] 图18是用于放大目标人的方法700的实施方案的操作流程。如图3中,在710处可导向阵列705(在ANC头戴式耳机中)提供固定边射波束成形,其中在用户观看方向上(即,在可导向阵列的用户正观看的方向上)产生经波束成形输出。具有相关联处理器的固定麦克风阵列707(例如在智能电话中)在围绕固定麦克风阵列的三个维度中执行DOA估计,且在720处分离作用中说话者。确定作用中说话者的数目,且产生每一作用中说话者的单独输出(由例如识别号识别)。

[0098] 在一实施方案中,可在730处执行作用中说话者的说话者辨识和标记。在750处确定固定阵列的经分离说话者与可导向阵列的输出之间的相关或相似性。使用相关或相似性测量和说话者的ID,可在760处检测、定位和放大目标用户。

[0099] 可用例如具有两个麦克风的隐藏相机等装置代替用户,且仅通过看着目标人,便可以用通过听觉以及视觉的缩放来聚焦于目标人。

[0100] 具有多个装置的摄像机应用是预期的。观看方向是已知的,且其它装置的所有可用麦克风可用以增强观看方向源。

[0101] 在一实施方案中,固定阵列的经分离说话者的音频波束与可导向阵列的输出之间的相关或相似性可用以自适应地形成目标说话者的较好波束。以此方式,可自适应地精炼固定麦克风波束成形器,使得固定波束成形器可自适应地产生新观看方向。

[0102] 举例来说,头戴式耳机麦克风阵列的波束成形器输出可用作参考以精炼固定麦克风阵列的波束成形器的观看方向。头戴式耳机波束成形器输出与当前固定麦克风阵列波束成形器输出之间的相关或相似性可同头戴式耳机波束成形器输出与具有稍微移动的观看方向的固定麦克风阵列波束成形器输出之间的相关或相似性进行比较。

[0103] 图19展示具有额外候选观看方向810的实例性用户接口显示800。通过利用如图19所示的头戴式耳机波束成形器输出与原始固定麦克风波束成形器输出820之间的相关或相似性,可产生固定波束成形器的新候选观看方向。使用此技术,头戴式耳机麦克风波束成形器输出可用作参考以精炼固定麦克风波束成形器的观看方向。举例来说,图19中的说话者1可正在说话,且在他说话时,可自适应地形成新候选观看方向。

[0104] 图20是用于自适应地精炼目标说话者的波束的方法900的实施方案的操作流程。如图3中,在910处可导向阵列905(例如,在ANC头戴式耳机中)提供固定边射波束成形,其中在用户观看方向上(即,在可导向阵列的用户正观看的方向上)产生经波束成形输出。具有相关联处理器的固定麦克风阵列907(例如在智能电话中)在围绕固定麦克风阵列的三个维度中执行DOA估计,且在920处分离作用中说话者。确定作用中说话者的数目,且产生每一作用中说话者的单独输出(由例如识别号识别)。如同图18,在950处确定固定阵列的经分离说话者与可导向阵列的输出之间的相关或相似性。

[0105] 继续图20,在960处,使用所确定相关或相似性来增加作用中用户的DOA附近的角度分辨率,且再次执行作用中说话者的分离。使用增加的角度分辨率和可导向阵列的输出,在970处,确定固定阵列的经分离说话者与可导向阵列的输出之间的另一相关或相似性量度。在980处可随后使用此相关或相似性量度来放大目标说话者。

[0106] 为在背景噪声和混响下充分稳健的同时多个声音事件的音频信号的每一帧估计三维到达方向(DOA)的方法是一个挑战。通过最大化可靠频率区间的数目可获得稳健性。可能希望此方法适合于任意形状的麦克风阵列几何形状,使得可避免麦克风几何形状上的特定约束。如本文描述的逐对1D方法可适当地并入到任何几何形状中。

[0107] 针对此一股扬声器电话应用或远场应用可实施解决方案。可实施此方法以在无麦克风放置约束的情况下操作。也可实施此方法以使用上至奈奎斯特频率且下至较低频率的可用频率区间(例如,通过支持具有较大麦克风间距离的麦克风对的使用)来跟踪源。并非限于用于跟踪的单个对,可实施此方法以选择所有可用对当中的最佳对。此方法可用以支持甚至远场情境(多达三到五米或更大的距离)中的源跟踪,且提供高得多的DOA分辨率。其它潜在特征包含获得作用中源的确切2D表示。为了最佳结果,可能希望每一源是稀疏宽带音频源,且每一频率区间大部分由不超过一个源支配。

[0108] 对于由一对麦克风在特定DOA上从点源直接接收的信号,相位延迟对于每一频率分量不同且还取决于麦克风之间的间距。在特定频率区间处的相位延迟的观测值可计算为复FFT系数的虚项与复FFT稀疏的实项的比率的反正切。如图21所示,在特定频率 f 处的相位延迟值 $\Delta\phi_f$ 可涉及在作为 $\Delta\phi_f = 2\pi f \frac{d \sin \theta}{c}$,的远场(即,平面波)假设下的源DOA,其中 d 表示

麦克风之间的距离(以m计), θ 表示相对于与阵列轴正交的方向的到达角度(以弧度计), f 表示频率(以Hz计),且 c 表示声音速度(以m/s计)。对于不具有混响的单个点源的理想情况,相位延迟与频率的比率 $\Delta\varphi/f$ 将在所有频率上具有相同值 $2\pi \frac{d \sin \theta}{c}$ 。

[0109] 此方法在实践中受到麦克风对的空间混叠频率限制,所述频率可界定为信号的波长为麦克风之间的距离 d 的两倍的频率。空间混叠造成相位缠绕,其对可用以为特定麦克风对提供可靠相位延迟测量的频率范围施加上限。图23展示针对四个不同DOA的未缠绕相位延迟对频率的绘图,且图24展示针对相同DOA的缠绕相位延迟对频率的绘图,其中以黑体展示每一绘图的初始部分(即,直到第一缠绕发生)。通过对测得相位进行解缠绕来延伸相位延迟测量的有用频率范围的尝试通常是不可靠的。

[0110] 替代于相位解缠绕,建议的方法针对DOA候选库存中的每一者将测得的(例如,缠绕)相位延迟与缠绕相位延迟的预计算值进行比较。图25展示包含(有噪声)测得相位延迟值(灰色)和库存的两个DOA候选(实线和虚线)的相位延迟值的角度对频率绘图的此实例,其中相位经缠绕于 π 到 $-\pi$ 的范围。随后通过针对每一DOA候选 θ_i 计算在频率分量 f 的范围上第 i DOA候选的相位延迟值 $\Delta\varphi_{i,f}$ 与观测相位延迟值 $\Delta\varphi_{ob,f}$ 之间的对应误差 e_i ,且识别对应于最小误差的DOA候选值,来确定最佳匹配于所观测信号的DOA候选。在一个实例中,误差 e_i 表达为 $\|\Delta\varphi_{ob,f} - \Delta\varphi_{i,f}\|_f^2$,即作为在频率分量的所要范围或其它集合 F 上观测与候选相位延迟值之间的差平方的和

$$[0111] \quad e_i = \sum_{f \in F} (\Delta\varphi_{ob,f} - \Delta\varphi_{i,f})^2$$

[0112] 每一DOA候选 θ_i 的相位延迟值 $\Delta\varphi_{i,f}$ 可在运行时间之前(例如,在设计或制造期间)根据 c 和 d 的已知值以及频率分量 f 的所要范围来计算,且在装置的使用期间从存储装置检索。此预计算库存可经配置以支持所要角度范围和分辨率(例如,均匀分辨率,例如一、二、五或十度,或所要不均匀分辨率)和所要频率范围和分辨率(也可为均匀或不均匀的)。

[0113] 可能希望在尽可能多的频率区间上计算误差 e_i ,以增加针对噪声的稳健性。举例来说,可能希望误差计算包含来自超越空间混叠频率的频率区间的项。在特定应用中,最大频率区间可受到其它因素限制,所述因素可包含可用存储器、计算复杂性、在高频率下刚性体的强反射等等。

[0114] 语音信号通常在时间频率域中是稀疏的。如果源在频域中不相连,那么可同时跟踪两个源。如果源在时域中不相连,那么可在同一频率跟踪两个源。可能希望阵列包含至少等于在任何一个时间待区分的不同源方向的数量数目的麦克风。麦克风可为全向的(例如,对于蜂窝式电话或专用会议装置可为典型的)或定向的(例如,对于例如机顶盒等装置可为典型的)。

[0115] 此多通道处理一股适用于例如用于扬声器电话应用的源跟踪。此技术可用以计算所接收多通道信号的帧的DOA估计。此方法可在每一频率区间计算每一候选角度相对于观测角度的误差,其由相位延迟指示。所述频率区间处的目标角度是具有最小误差的候选。在一个实例中,随后在频率区间上将误差求和以获得所述候选的可能性量度。在另一实例中,在所有频率区间上最经常发生的目标DOA候选中的一或多者被识别为给定帧的DOA估计(或

多个估计)。

[0116] 可应用此方法以获得瞬时跟踪结果(例如,具有小于一个帧的延迟)。延迟取决于FFT大小和重叠程度。举例来说,对于具有50%重叠和16kHz取样频率的512点FFT,所得的256样本延迟对应于十六毫秒。此方法可用以支持通常多达两到三米或甚至多达五米的源-阵列距离的源方向区分。

[0117] 误差也可视为方差(即,个别误差偏离预期值的程度)。时域接收信号到频域中的转换(例如,通过应用FFT)具有使每一区间中的频谱平均化的作用。此平均化在使用子带表示(例如,梅尔尺度或巴克尺度)的情况下甚至更明显。另外,可能希望对DOA估计执行时域平滑(例如,通过作为回归平滑器应用,例如一阶无限脉冲响应滤波器)。

[0118] 可能希望减少误差计算操作的计算复杂性(例如,通过使用例如二元树等搜索策略,和/或应用已知信息,例如从一或多个先前帧的DOA候选选择)。

[0119] 即使可在相位延迟方面测量方向信息,通常也希望获得指示源DOA的结果。因此,可能希望在DOA方面而非相位延迟方面计算误差。

[0120] 在DOA方面的误差 e_i 的表达式可通过假定观测缠绕相位延迟的表达式为DOA的函数来推导,例如

$$[0121] \quad \Psi_{f_wr}(\theta) = \text{mod}\left(-2\pi f \frac{d \sin \theta}{c} + \pi, 2\pi\right) - \pi,$$

[0122] 等效于作为DOA函数的未缠绕相位延迟的对应表达式,例如

$$[0123] \quad \Psi_{f_un}(\theta) = -2\pi f \frac{d \sin \theta}{c},$$

[0124] 由于相位缠绕所致的近似不连续性除外。误差 e_i 可随后表达为

$$[0125] \quad e_i = \|\Psi_{f_wr}(\theta_{ob}) - \Psi_{f_wr}(\theta_i)\|_f^2 \equiv \|\Psi_{f_un}(\theta_{ob}) - \Psi_{f_un}(\theta_i)\|_f^2,$$

[0126] 其中在频率 f 下观测与候选相位延迟之间的差在DOA方面表达为

$$[0127] \quad \Psi_{f_un}(\theta_{ob}) - \Psi_{f_un}(\theta_i) = \frac{-2\pi f d}{c} (\sin \theta_{ob_f} - \sin \theta_i).$$

[0128] 执行泰勒级数展开以获得以下一阶近似:

$$[0129] \quad \frac{-2\pi f d}{c} (\sin \theta_{ob_f} - \sin \theta_i) \approx (\theta_{ob_f} - \theta_i) \frac{-2\pi f d}{c} \cos \theta_i,$$

[0130] 其用以获得在频率 f 下观测的DOA θ_{ob_f} 与DOA候选 θ_i 之间的差的表达式:

$$[0131] \quad (\theta_{ob_f} - \theta_i) \cong \frac{\Psi_{f_un}(\theta_{ob}) - \Psi_{f_un}(\theta_i)}{\frac{2\pi f d}{c} \cos \theta_i}.$$

[0132] 此表达式可在假定观测缠绕相位延迟与未缠绕相位延迟的等效的情况下用以在DOA方面表达误差 e_i :

$$[0133] \quad e_i = \|\theta_{ob} - \theta_i\|_f^2 \cong \frac{\|\Psi_{f_wr}(\theta_{ob}) - \Psi_{f_wr}(\theta_i)\|_f^2}{\left\|\frac{2\pi f d}{c} \cos \theta_i\right\|_f^2},$$

[0134] 其中 $[\Psi_{f_wr}(\theta_{ob}), \Psi_{f_wr}(\theta_i)]$ 的值界定为 $[\Delta\varphi_{ob_f}, \Delta\varphi_{i_f}]$ 。

[0135] 为了避免在端射方向($\theta = \pm 90^\circ$)除以零,可能希望改为使用二阶近似来执行此展开,如下:

$$[0136] \quad |\theta_{ob} - \theta_i| \cong \begin{cases} |-C/B|, & \theta_i = 0 \text{ (边射)} \\ \left| \frac{-B + \sqrt{B^2 - 4AC}}{2A} \right|, & \text{其它情况} \end{cases},$$

[0137] 其中 $A = (\pi f d \sin \theta_i) / c$, $B = (-2\pi f d \cos \theta_i) / c$, 且

[0138] $C = -(\Psi_{f_un}(\theta_{ob}) - \Psi_{f_un}(\theta_i))$ 。

[0139] 如以上一阶实例中,此表达式可在假定观测缠绕相位延迟与未缠绕相位延迟的等效的情况下用以在DOA方面表达误差 e_i 作为观测和候选缠绕相位延迟值的函数。

[0140] 如图27中所示,所接收信号的给定帧的观测与候选DOA之间的差可以此方式在所接收麦克风信号(例如, $\forall f \in F$)的多个频率 f 中的每一者处且针对多个DOA候选 θ_i 中的每一者计算。如图28中证明,给定帧的DOA估计可通过将所述帧中的所有频率区间上的每一候选的平方差求和以获得误差 e_i 且选择具有最小误差的DOA候选来确定。替代地,如图29中证明,此些差可用以识别每一频率处的最佳匹配(即,最小平方差)DOA候选。帧的DOA估计可随后确定为所有频率区间上的最频繁DOA。

[0141] 如图31所示,可针对每一候选角度 i 和每一帧 k 的频率集合 F 中的每一者计算误差项。可能希望在计算的DOA差或误差方面指示源活动的可能性。此可能性的一个实例 L 可针对特定帧、频率和角度表达为

$$[0142] \quad L(i, f, k) = \frac{1}{|\theta_{ob} - \theta_i|_{f,k}^2} \quad (1)$$

[0143] 对于表达式(1),在特定频率处的极好匹配可造成对应可能性支配所有其它可能性。为了减少此易感性,可能希望包含规则化项 λ ,如以下表达式中:

$$[0144] \quad L(i, f, k) = \frac{1}{|\theta_{ob} - \theta_i|_{f,k}^2 + \lambda} \quad (2)$$

[0145] 语音往往在时间和频率两者上稀疏,使得在频率集合 F 上的和可包含来自噪声支配的区间的结果。可能希望包含偏置项 β ,如以下表达式中:

$$[0146] \quad L(i, f, k) = \frac{1}{|\theta_{ob} - \theta_i|_{f,k}^2 + \lambda} - \beta \quad (3)$$

[0147] 可随着频率和/或时间变化的偏置项可基于噪声的假定分布(例如,高斯)。另外或替代地,偏置项可基于噪声的初始估计(例如,来自仅噪声初始帧)。另外或替代地,可例如由话音活动检测模块指示,基于来自仅噪声帧的信息动态地更新偏置项。

[0148] 频率特定可能性结果可投影到(帧,角度)平面上以获得每帧的DOA估计 $\theta_{est,k} = \max_i \sum_{f \in F} L(i, f, k)$,其对噪声和混响是稳健的,因为仅目标支配性频率区间贡献于估计。在此求和中,其中误差较大的项具有接近零的值,且因此变为对估计较不重要。如果一方向源在一些频率区间中是支配性的,那么在那些频率区间处的误差值将针对所述角度较接近于零。而且,如果另一方向源在其它频率区间中是支配性的,那么在所述其它频率区间处的

误差值将针对其它角度较接近于零。

[0149] 基于方向成员关系(例如,用于话音活动检测),可能性结果还可投影到(帧,频率)平面上以指示每频率区间的可能性信息。此可能性可用以指示语音活动的可能性。另外或替代地,此信息可例如用以通过根据到达方向对帧和/或频率分量进行分类来支持所接收信号的时间和/或频率选择性掩蔽。

[0150] 角度图表示类似于谱图表示。可通过在每一帧处绘制在每一频率下的当前DOA候选的可能性来获得角度图。

[0151] 具有大间距的麦克风对通常不适合于高频率,因为空间混叠在针对此对的低频率下开始。然而,如本文描述的DOA估计方法允许使用超越相位缠绕开始的频率且甚至高达奈奎斯特频率(即,取样率的一半)的相位延迟测量。通过放松空间混叠约束,此方法使得能够使用具有较大麦克风间间距的麦克风对。由于与具有小麦克风间距离的阵列相比,具有大麦克风间距离的阵列通常在低频率下提供较好的方向性,因此较大阵列的使用通常也将有用相位延迟测量的范围延伸到较低频率中。

[0152] 本文描述的DOA估计原理可延伸到线性阵列中的多个麦克风对(例如,如图22中所示)。用于远场情境的此应用的一个实例是沿着电视机或其它大格式视频显示屏幕的边限布置的线性麦克风阵列(例如,如图26中所示)。可能希望将此阵列配置为在麦克风之间具有不均匀(例如,对数)间距,如图22和26的实例中。

[0153] 对于远场源,线性阵列的多个麦克风对将具有基本上相同的DOA。因此,一个选项是将DOA估计为来自阵列中的两个或两个以上对的DOA估计的平均值。然而,平均化方案可受到所述对中的甚至单个一对的失配影响,其可降低DOA估计准确性。替代地,可能希望从阵列的两对或两对以上麦克风当中选择用于每一频率的最佳麦克风对(例如,在所述频率下给出最小误差 e_i 的对),使得针对不同频带可选择不同麦克风对。在麦克风对的空间混叠频率下,误差将较大。因此,此方法将往往在频率接近于其缠绕频率时自动避开麦克风对,因此避免DOA估计中的相关不确定性。对于较高频率区间,在麦克风之间具有较短距离的对将通常提供较好估计且可自动偏好,而对于较低频率区间,在麦克风之间具有较大距离的对将通常提供较好估计且可自动偏好。在图22中所示的四麦克风实例中,不同的六对麦克风是可能的(即, $\binom{4}{2} = 6$)。

[0154] 在一个实例中,对于每一轴的最佳对是通过针对每一频率 f 计算 $P \times I$ 值来选择,其中 P 是对数目, I 是库存的大小,且每一值 e_{pi} 是观测角度 θ_{pf} (对于对 p 和频率 f)与候选角度 θ_{if} 之间的平方绝对差。对于每一频率 f ,选择对应于最低误差值 e_{pi} 的对 p 。此误差值还指示在频率 f 下的最佳DOA候选 θ_i (如图30中所示)。

[0155] 由麦克风对接收的信号可如本文描述经处理以提供相对于麦克风对的轴线在多达180度的范围上提供估计DOA。所要角度跨度和分辨率在所述范围内可为任意的(例如,均匀(线性)或不均匀(非线性),限于所关注的选定扇区等等)。另外或替代地,所要频率跨度和分辨率可为任意的(例如,线性、对数、梅尔尺度、巴克尺度等等)。

[0156] 在图22中所示的模型中,来自麦克风对的在0与 ± 90 度之间的每一DOA估计指示相对于与所述对的轴线正交的平面的角度。此估计描述围绕所述对的轴线的圆锥,且沿着此圆锥的表面的源的实际方向是不确定的。举例来说,来自单个麦克风对的DOA估计不指示

所述源是否在麦克风对前方或后方。因此,虽然在线性阵列中可使用两个以上麦克风来改善一频率范围上的DOA估计性能,但由线性阵列支持的DOA估计范围通常限于180度。

[0157] 本文描述的DOA估计原理也可延伸到二维(2D)麦克风阵列。举例来说,2D阵列可用以将源DOA估计范围延伸到多达完整360°(例如,提供与例如雷达和生物医学扫描等应用中相似的范围)。此阵列可在例如扬声器电话应用中使用以甚至针对电话相对于一或多个源的任意放置也支持良好性能。

[0158] 甚至对于远场点源,2D阵列的多个麦克风对也通常将不共享相同DOA。举例来说,相对于阵列的平面的源高度(例如,在z轴上)可在2D跟踪中起重要作用。图32展示扬声器电话应用的实例,其中由麦克风轴线界定的x-y平面平行于电话放置于其上的表面(例如,桌面)。在此实例中,源是从沿着x轴的位置说话但在z轴的方向上偏移的人(例如,说话者的嘴部高于桌面)。相对于由麦克风阵列界定的x-y平面,源的方向是沿着x轴,如图32中所示。沿着y轴的麦克风对将源的DOA估计为从x-z平面的零度。然而由于说话者高于x-y平面的高度,沿着x轴的麦克风对将源的DOA估计为从x轴的30°(即,从y-z平面的60度),而不是沿着x轴。图34和35展示与此DOA估计相关联的干扰圆锥的两个视图,其造成相对于麦克风轴线的估计说话者方向中的模糊性。

[0159] 例如以下表达式

$$[0160] \quad \left[\tan^{-1} \left(\frac{\sin \theta_1}{\sin \theta_2} \right), \tan^{-1} \left(\frac{\sin \theta_2}{\sin \theta_1} \right) \right], \quad (4)$$

[0161] 其中 θ_1 和 θ_2 分别为对1和2的估计DOA,可用以将DOA的所有对投影到其中定位三个麦克风的平面中的360°范围。此投影可用以使得能够在麦克风阵列周围的360°范围上跟踪作用中说话者的方向,无论高度差如何。应用以上表达式以将图32的DOA估计(0°,60°)投影到x-y平面中产生

$$[0162] \quad \left[\tan^{-1} \left(\frac{\sin 0^\circ}{\sin 60^\circ} \right), \tan^{-1} \left(\frac{\sin 60^\circ}{\sin 0^\circ} \right) \right] = (0^\circ, 90^\circ),$$

[0163] 其可映射到如图33中所示的270°的组合方向估计(例如,方位角)。

[0164] 在定性使用情况下,源将位于未投影到麦克风轴线上的方向上。图37到40展示其中源位于麦克风的平面上方的此实例。在此实例中,源信号的DOA穿过点 $(x, y, z) = (5, 2, 5)$ 。图37展示从+z方向所见的x-y平面,图38和40展示从麦克风MC30的方向所见的x-z平面,且图39展示从麦克风MC10的方向所见的y-z平面。图37中的阴影区域指示由y轴麦克风对MC20-MC30观测到的与DOA θ_1 相关联的干扰圆锥CY,且图38中的阴影区域指示由x轴麦克风对MC10-MC20观测到的与DOA θ_2 相关联的干扰圆锥CX。在图39中,阴影区域指示圆锥CY,且虚线圆指示圆锥CY与穿过源且与x轴正交的平面的交点。此圆上指示其与圆锥CY的交点的两个点是源的候选位置。同样,在图40中,阴影区域指示圆锥CX,虚线圆指示圆锥CY与穿过源且与y轴正交的平面的交点,且此圆上指示其与圆锥CX的交点的两个点是源的候选位置。可见在此2D情况下,相对于源是否高于或低于x-y平面仍有模糊性。

[0165] 对于图37到40中所示的实例,由x轴麦克风对MC10-MC20观测到的DOA是 $\theta_2 = \tan^{-1}(-5/\sqrt{25+4}) \approx -42.9^\circ$,且由y轴麦克风对MC20-MC30观测到的DOA是 $\theta_1 = \tan^{-1}(-2/\sqrt{25+25}) \approx -15.8^\circ$ 。使用表达式(4)将这些方向投影到x-y平面中产生

分别相对于x和y轴的所要角度的量值 ($21.8^\circ, 68.2^\circ$) ,其对应于给定源位置 $(x, y, z) = (5, 2, 5)$ 。观测到的角度的正负号指示源位于其中的x-y象限,如图36中所示。

[0166] 事实上,2D麦克风阵列几乎给出3D信息,除了上下混淆。举例来说,由麦克风对MC10-MC20和MC20-MC30观测到的到达方向也可用以估计源的仰角相对于x-y平面的量值。如果d表示从麦克风MC20到源的向量,那么向量d到x轴、y轴和x-y平面上的投影的长度可分别表达为 $d \sin(\theta_2)$, $d \sin(\theta_1)$ 和 $d\sqrt{\sin^2(\theta_1) + \sin^2(\theta_2)}$ 。仰角的量值可随后估计为 $\hat{\theta}_h = \cos^{-1} \sqrt{\sin^2(\theta_1) + \sin^2(\theta_2)}$ 。

[0167] 虽然图32到33和37到40的特定实例中的麦克风对具有正交轴,但应注意,对于具有非正交轴的麦克风对,表达式(4)可用以将DOA估计投影到那些非正交轴,且从所述点直接获得相对于正交轴的组合方向估计的表示。图41展示麦克风阵列MC10-MC20-MC30的实例,其中对MC20-MC30的轴线1位于x-y平面中且相对于y轴偏斜一偏斜角 θ_0 。

[0168] 图42展示以来自图41中所示的阵列的观测 (θ_1, θ_2) 获得x-y平面中相对于正交轴x和y的组合方向估计的实例。如果d表示从麦克风MC20到源的向量,那么向量d到x轴和轴线1上的投影的长度可分别表达为 $d \sin(\theta_2)$ 和 $d \sin(\theta_1)$ 。向量 (x, y) 表示向量d到x-y平面上的投影。x的估计值是已知的,且其保留以估计y的值。

[0169] y的估计可使用向量 (x, y) 到轴线1上的投影 $p_1 = (d \sin\theta_1 \sin\theta_0, d \sin\theta_1 \cos\theta_0)$ 来执行。观测到向量 (x, y) 与向量 p_1 之间的差正交于 p_1 ,将y计算为

$$[0170] \quad y = d \frac{\sin \theta_1 - \sin \theta_2 \sin \theta_0}{\cos \theta_0}。$$

[0171] x-y平面中相对于正交的x和y轴的所要到达方向可随后分别表达为

[0172]

$$\left(\tan^{-1} \left(\frac{y}{x} \right), \tan^{-1} \left(\frac{x}{y} \right) \right) = \left(\tan^{-1} \left(\frac{\sin \theta_1 - \sin \theta_2 \sin \theta_0}{\sin \theta_2 \cos \theta_0} \right), \tan^{-1} \left(\frac{\sin \theta_2 \cos \theta_0}{\sin \theta_1 - \sin \theta_2 \sin \theta_0} \right) \right)。$$

[0173] 到2D阵列的DOA估计的延伸通常良好地适于且足以用于扬声器电话应用。然而,到N维阵列的进一步延伸也是可能的,且可以直接方式执行。对于其中一个目标是支配性的跟踪应用,可能希望选择N对以用于表示N个维度。一旦以特定麦克风对获得2D结果,便可利用另一可用对来增加自由度。举例来说,图37到42说明使用来自x-y平面中的不同麦克风对的观测DOA估计来获得投影到x-y平面中的源方向的估计。以相同方式,来自x轴麦克风对和z轴麦克风对(或x-z平面中的其它对)的观测DOA估计可用以获得投影到x-z平面中的源方向的估计,且对于y-z平面或与麦克风中的三者或三者以上相交的任一其它平面是相同的。

[0174] 来自不同维度的DOA误差的估计可用以例如使用例如以下表达式来获得组合可能性估计

$$[0175] \quad \frac{1}{\max(|\theta - \theta_{0,1}|_{f,1}^2, |\theta - \theta_{0,2}|_{f,2}^2) + \lambda} \text{ 或 } \frac{1}{\text{mean}(|\theta - \theta_{0,1}|_{f,1}^2, |\theta - \theta_{0,2}|_{f,2}^2) + \lambda}，$$

[0176] 其中 $\theta_{0,i}$ 表示为对i选择的DOA候选。使用不同误差当中的最大者可能需要优先于接近于干扰圆锥中的仅一者的估计而促进对接近于两个观测的干扰圆锥的估计的选择,且因此可能指示错误的峰。此组合结果可用以获得如本文所述的(帧,角度)平面和/或如本文

所述的(帧,频率)绘图。

[0177] 本文描述的DOA估计原理可用以支持多个说话者当中的选择。举例来说,多个源的位置可与特定说话者的手动选择(例如,按下特定按钮以选择特定对应用户)或特定说话者的自动选择(例如,通过说话者辨识)组合。在一个此类应用中,电话经配置以辨识其所有者的话音,且优先于其它源的方向而自动选择对应于所述话音的方向。

[0178] 源DOA可容易地在1D中界定,例如从 -90° 到 90° 。从任意相对位置处的两个以上麦克风,建议使用如上所述的1D的直接延伸,例如2D中的两对情况中的 (θ_1, θ_2) 、3D中的三对情况中的 $(\theta_1, \theta_2, \theta_3)$ 等等。

[0179] 关键问题是如何将空间滤波应用于成对1D DOA估计的此组合。在此情况下,通过增强用于每一对的导向向量可应用如图43中所示的波束成形器/空波束成形器(BFNF)。在此图中, A^H 表示A的共轭转置, x 表示麦克风通道,且 y 表示经空间滤波通道。使用如图43中所示的伪逆运算 $A^+ = (A^H A)^{-1} A^H$ 允许使用非方阵。对于例如如图45中说明的三麦克风情况(即,两个麦克风对),行数目 $2 \times 2 = 4$ 而非3,使得额外的行使矩阵为非方阵。

[0180] 由于图43中所示的方法是基于稳健的1D DOA估计,因此不需要麦克风几何形状的完全了解,且也不需要同时使用所有麦克风的DOA估计。此方法良好地适合于与如本文所述的基于角度图的DOA估计一起使用,但也可使用任何其它1D DOA估计方法。图44展示如图43中所示的BFNF的实例,其还包含正规化因数以防止在空间混叠频率下的不良条件的逆。

[0181] 图46展示逐对(PW)正规化MVDR(最小方差无失真响应)BFNF的实例,其中获得导向向量(阵列流形向量)的方式不同于常规方法。在此情况下,由于两个对之间的麦克风的共享而消除了共同通道。通过测量或通过使用sinc函数的理论计算来获得噪声相干矩阵 Γ 。应注意,图43、44和46的实例可经一股化于任意数目的源N,使得 $N \leq M$,其中M是麦克风的数目。

[0182] 图47展示在矩阵 $A^H A$ 并未条件不良的情况下可使用的另一实例,其可使用矩阵的条件数或行列式来确定。如果矩阵条件不良,那么可能希望绕过用作源通道的频率区间的—个麦克风信号,同时继续应用方法以对其中矩阵 $A^H A$ 并未条件不良的其它频率区间进行空间滤波。此选项节省了用于计算用于正规化的分母的计算。图43到47中的方法证明了可在每一频率区间处独立应用的BFNF技术。使用如本文所述的用于每一频率和麦克风对的DOA估计来构造导向向量。举例来说,用于DOA θ_i ,频率 f 和麦克风数目 m (1或2)的对 p 和源 n 的导向向量的每一元素可计算为

$$[0183] \quad d_{p,m}^n = \exp\left(\frac{-j\omega f_s(m-1)l_p}{c} \cos \theta_i\right),$$

[0184] 其中 l_p 指示对 p 的麦克风之间的距离, ω 指示频率区间数,且 f_s 指示取样频率。图48展示如图45中所示的阵列的导向向量的实例。

[0185] PWBFBF方案可用于抑制干扰者直到可用自由度的直接路径(不具有平滑轨迹假设的瞬时抑制,使用方向性掩蔽的额外噪声抑制增益,使用带宽延伸的额外噪声抑制增益)。象限框架的单通道后处理可用于静态噪声和噪声参考处置。

[0186] 可能希望获得瞬时抑制,但也提供例如音乐噪声等假象的最小化。可能希望最大地使用可用自由度来用于BFNF。可在所有频率上固定一个DOA,或可准许频率上的稍微失配的对准。可仅使用当前帧,或可实施前馈网络。可针对高达奈奎斯特速率的范围中的所有频

率(例如,不良条件的频率除外)设定BFNF。可使用自然掩蔽方法(例如,以获得平稳自然无缝的激进性过渡)。

[0187] 图49展示如本文描述的集成方法的一个实例的流程图。此方法包含用于相位延迟估计的库存匹配任务,用以获得DOA误差方差值的方差计算任务,维度匹配和/或对选择任务,以及用以将选定DOA候选的DOA误差方差映射到源活动可能性估计的任务。逐对DOA估计结果也可用以跟踪一或多个作用中说话者,执行逐对空间滤波操作,和或执行时间和/或频率选择性掩蔽。活动可能性估计和/或空间滤波操作也可用以获得噪声估计以支持单通道噪声抑制操作。

[0188] 本文揭示的方法和设备可一般应用于任何收发和/或音频感测应用中,尤其是这些应用的移动或另外便携式实例。举例来说,本文揭示的配置的范围包含驻留在经配置以采用码分多址(CDMA)空中接口的无线电话通信系统中的通信装置。然而,所属领域的技术人员将了解,具有如本文描述的特征的方法和设备可驻留在采用所属领域的技术人员已知的广泛多种技术的各种通信系统中的任一者中,例如采用有线和/或无线(例如,CDMA、TDMA、FDMA和/或TD-SCDMA)发射信道上的IP语音(VoIP)的系统。

[0189] 明确预期且在此揭示,本文揭示的通信装置可适于在包交换(例如,经布置以根据例如VoIP等协议载运音频发射的有线和/或无线网络)和/或电路交换的网络中使用。还明确预期且在此揭示,本文揭示的通信装置可适于在窄带译码系统(例如,对大约四或五千赫的音频频率范围进行编码的系统)中使用和/或在宽带译码系统(例如,对大于五千赫的音频频率进行编码的系统)中使用,包含全带宽带译码系统和分带宽带译码系统。

[0190] 可与如本文描述的通信装置的发射器和/或接收器一起使用或适于与其一起使用的编解码器的实例包含:增强型可变速率编解码器,如标题为“增强型可变速率编解码器,用于宽带扩展频谱数字系统的语音服务选项3、68和70(Enhanced Variable Rate Codec, Speech Service Options 3,68,and 70for Wideband Spread Spectrum Digital Systems)”第三代合作伙伴计划2(3GPP2)文档C.S0014-C,v1.0(2007年2月,在www-dot-3gpp-dot-org在线可用)中描述;可选择模式声码器语音编解码器,如标题为“用于宽带扩展频谱通信系统的可选择模式声码器(SMV)服务选项(Selectable Mode Vocoder(SMV) Service Option for Wideband Spread Spectrum Communication Systems)”的3GPP2文档C.S0030-0,v3.0(2004年1月,在www-dot-3gpp-dot-org在线可用)中描述;自适应多速率(AMR)语音编解码器,如文档ETSI TS 126 092V6.0.0(欧洲电信标准协会(ETSI),法国索菲亚安提波利斯企业邮政编码,2004年12月)中描述;以及AMR宽带语音编解码器,如文档ETSI TS 126 192V6.0.0(ETSI,2004年12月)中描述。此编解码器可例如用以从所接收无线通信信号恢复经再生音频信号。

[0191] 提供所描述配置的呈现以使得所属领域的技术人员能够制作或使用本文揭示的方法和结构。本文展示和描述的流程图、框图和其它结构仅为实例,且这些结构的其它变体也在本发明的范围内。对这些配置的各种修改是可能的,且在此呈现的一股原理也可应用于其它配置。因此,本发明既定不限于上文展示的配置,而是应被赋予与本文以任何方式(包含所申请的形成原始发明的一部分的所附权利要求书中)揭示的原理和新颖特征一致的最广范围。

[0192] 所属领域的技术人员将了解,可使用多种不同技艺和技术中的任一者来表示信息

和信号。举例来说,贯穿以上描述可参考的数据、指令、命令、信息、信号、位和符号可通过电压、电流、电磁波、磁场或磁粒子、光场或光粒子或其任一组合来表示。

[0193] 用于如本文揭示的配置的实施的重要设计要求可包含最小化处理延迟和/或计算复杂性(通常以每秒百万指令或MIPS测量),尤其是针对计算密集型应用,例如经压缩音频或视听信息的重放(例如,根据例如本文识别的实例中的一者的压缩格式编码的文件或流)或用于宽带通信的应用(例如,在高于八千赫的取样率下的话音通信,例如12、16、32、44.1、48或192kHz)。

[0194] 如本文揭示的设备(例如,经配置以执行如本文描述的技术的任一装置)可以被视为适合于既定应用的硬件与软件和/或与固件的任一组合来实施。举例来说,此设备的元件可制造为例如驻留在同一芯片上或芯片组中的两个或两个以上芯片当中的电子和/或光学装置。此装置的一个实例是例如晶体管或逻辑门等逻辑元件的固定或可编程阵列,且这些元件中的任一者可实施为一或多个此类阵列。这些元件中的任何两个或两个以上或甚至全部可在相同的一或多个阵列内实施。此一或多个阵列可在一或多个芯片内(例如,在包含两个或两个以上芯片的芯片组内)实施。

[0195] 本文揭示的设备的各种设施方案的一或多个元件可整体或部分地实施为经布置以在一或多个固定或可编程逻辑元件阵列上执行的一或多个指令集,所述逻辑元件阵列例如为微处理器、嵌入式处理器、IP核心、数字信号处理器、FPGA(现场可编程门阵列)、ASSP(专用标准产品)和ASIC(专用集成电路)。如本文揭示的设备的实施方案的各种元件中的任一者也可体现为一或多个计算机(例如,包含经编程以执行一或多个指令集或指令序列的一或多个阵列的机器,也称为“处理器”),且这些元件中的任何两个或两个以上或甚至全部可在相同的此类一或多个计算机内实施。

[0196] 如本文揭示的处理器或用于处理的其它装置可制造为例如驻留在同一芯片上或芯片组中的两个或两个以上芯片当中的一或多个电子和/或光学装置。此装置的一个实例是例如晶体管或逻辑门等逻辑元件的固定或可编程阵列,且这些元件中的任一者可实施为一或多个此类阵列。此一或多个阵列可在一或多个芯片内(例如,在包含两个或两个以上芯片的芯片组内)实施。这些阵列的实例包含固定或可编程逻辑元件阵列,例如微处理器、嵌入式处理器、IP核心、DSP、FPGA、ASSP和ASIC。如本文揭示的处理器或用于处理的其它装置也可体现为一或多个计算机(例如,包含经编程以执行一或多个指令集或指令序列的一或多个阵列的机器)或其它处理器。如本文描述的处理器可用以执行不直接与本文描述的实施方案的程序相关的任务或其它指令集,例如与其中嵌入处理器的装置或系统(例如,音频感测装置)的另一操作相关的任务。如本文揭示的方法的部分还可由音频感测装置的处理器执行,且所述方法的另一部分在一或多个其它处理器的控制下执行。

[0197] 所属领域的技术人员将了解,结合本文揭示的配置描述的各种说明性模块、逻辑块、电路和测试以及其它操作可实施为电子硬件、计算机软件或两者的组合。这些模块、逻辑块、电路和操作可以通用处理器、数字信号处理器(DSP)、ASIC或ASSP、FPGA或其它可编程逻辑装置、离散门或晶体管逻辑、离散硬件组件或其经设计以产生如本文揭示的配置的任一组合来实施或执行。举例来说,此配置可至少部分地实施为硬连线电路、制造于专用集成电路中的电路配置,或者加载到非易失性存储装置中的固件程序或作为机器可读代码从数据存储媒体加载或加载到其中的软件程序,此代码是可由例如通用处理器或其它数字信号

处理单元等逻辑元件阵列执行的指令。通用处理器可为微处理器,但在替代例中,处理器可为任何常规处理器、控制器、微控制器或状态机。处理器也可实施为计算装置的组合,例如DSP与微处理器的组合、多个微处理器、结合DSP核心的一或多个微处理器,或任何其它此类配置。软件模块可驻留于非暂时性存储媒体中,例如RAM(随机存取存储器)、ROM(只读存储器)、例如快闪RAM等非易失性RAM(NVRAM)、可擦除可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)、寄存器、硬盘、可装卸式磁盘、或CD-ROM,或此项技术中已知的任一其它形式的存储媒体中。说明性存储媒体耦合到处理器,使得处理器可从存储媒体读取信息和向存储媒体写入信息。在替代方案中,存储器媒体可与处理器成一体式。处理器和存储媒体可驻留在ASIC中。ASIC可驻留在用户终端中。在替代方案中,处理器和存储媒体可作为离散组件驻留在用户终端中。

[0198] 应注意,本文揭示的各种方法可由例如处理器等逻辑元件阵列执行,且如本文描述的设备的各种元件可实施为经设计以在此阵列上执行的模块。如本文使用,术语“模块”或“子模块”可指代呈软件、硬件或固件形式的任何方法、设备、装置、单元或包含计算机指令(例如,逻辑表达式)的计算机可读数据存储媒体。应了解,多个模块或系统可组合为一个模块或系统,且一个模块或系统可分离为多个模块或系统以执行相同功能。当以软件或其它计算机可执行指令实施时,过程的元素基本上是用以执行相关任务的代码段,例如例程、程序、对象、组件、数据结构和类似物。术语“软件”应理解为包含源代码、汇编语言代码、机器代码、二进制码、固件、宏码、微码、可由逻辑元件阵列执行的任何一或多个指令集或指令序列,和此些实例的任一组合。程序或代码段可存储在处理器可读媒体中或通过体现于载波中的计算机数据信号在传输媒体或通信链路上发射。

[0199] 本文描述的方法的任务中的每一者可直接以硬件、以由处理器执行的软件模块或以两者的组合来体现。在如本文揭示的方法的实施方案的典型应用中,逻辑元件(例如,逻辑门)阵列经配置以执行所述方法的各种任务中的一者、一者以上或甚至全部。所述任务中的一或多者(可能全部)也可实施为代码(例如,一或多个指令集),体现于计算机程序产品(例如,例如磁盘、快闪或其它非易失性存储器卡、半导体存储器芯片等等一或多个数据存储媒体)中,其可由包含逻辑元件阵列(例如,处理器、微处理器、微控制器或其它有限状态机)的机器(例如,计算机)读取和/或执行。如本文揭示的方法的实施方案的任務也可由一个以上此类阵列或机器执行。在这些或其它实施方案中,所述任务可在例如蜂窝式电话等用于无线通信的装置或具有此通信能力的其它装置内执行。此装置可经配置以与电路交换和/或包交换网络通信(例如,使用例如VoIP等一或多个协议)。举例来说,此装置可包含经配置以接收和/或发射经编码帧的RF电路。

[0200] 明确地揭示了本文揭示的各种方法可由例如手持机、头戴式耳机或便携式数字助理(PDA)等便携式通信装置执行,且本文描述的各种设备可包含在此装置内。

[0201] 在一或多个示范性实施例中,本文描述的操作可以硬件、软件、固件或其任一组合实施。如果以软件实施,那么此些操作可作为一或多个指令或代码存储在计算机可读媒体上或经由计算机可读媒体传输。术语“计算机可读媒体”包含计算机可读存储媒体和通信(例如,传输)媒体两者。举例来说而非限制,计算机可读存储媒体可包括:存储元件阵列,例如半导体存储器(可包含(不限于)动态或静态RAM、ROM、EEPROM和/或快闪RAM)或铁电的、磁阻的、双向的、聚合的或相变存储器;CD-ROM或其它光盘存储装置;和/或磁盘存储装置或其

它磁性存储装置。此存储媒体可以可由计算机存取的指令或数据结构的形式存储信息。通信媒体可包括可用以用指令或数据结构的形式载运所要程序代码且可由计算机存取的任何媒体,包含促进计算机程序从一处转移到另一处的任何媒体。而且,将任何连接恰当地称为计算机可读媒体。举例来说,如果使用同轴电缆、光纤电缆、双绞线、数字订户线(DSL)或例如红外线、无线电和/或微波等无线技术从网站、服务器或其它远程源发射软件,那么同轴电缆、光纤电缆、双绞线、DSL或例如红外线、无线电和/或微波等无线技术包含于媒体的定义中。如本文所使用,磁盘和光盘包含压缩光盘(CD)、激光光盘、光学光盘、数字多功能光盘(DVD)、软磁盘和Blu-ray Disc™(蓝光光盘协会,加利福尼亚环球城),其中磁盘通常以磁性方式再生数据,而光盘用激光以光学方式再生数据。以上各项的组合也应包含在计算机可读媒体的范围内。

[0202] 如本文描述的声信号处理设备可并入到接受语音输入以便控制某些操作或可另外得益于所要噪声与背景噪声的分离的电子装置(例如通信装置)中。许多应用可得益于增强或分离清晰的所要声音与源自多个方向的背景声音。这些应用可包含并入有例如话音辨识和检测、语音增强和分离、话音激活的控制和类似情况的能力的电子或计算装置中的人机接口。可能希望实施此声信号处理设备以适合于仅提供有限处理能力的装置中。

[0203] 如本文描述的设备的实施方案的一或多个元件可用以执行不直接与所述设备的操作相关的任务或其它指令集,例如与其中嵌入所述设备的装置或系统的另一操作相关的任务。此设备的实施方案的一或多个元件也可具有共同的结构(例如,用以在不同时间执行对应于不同元件的代码的部分的处理器,经执行以在不同时间执行对应于不同元件的任务的指令集,或在不同时间执行用于不同元件的操作的电子和/或光学装置的布置)。

[0204] 提供本发明的先前描述以使得所属领域的技术人员能够制作或使用本发明。所属领域的技术人员将容易明了对本发明的各种修改,且在不脱离本发明的范围的情况下可将本文界定的一股原理应用于其它变化。因此,本发明既定不限于本文描述的实例和设计,而是应被赋予与本文揭示的原理和新颖特征一致的最广范围。

[0205] 虽然示范性实施方案可涉及在一或多个独立计算机系统的情形中利用当前揭示标的物的方面,但所述标的物不受此限制,而是可结合例如网络或分布式计算环境等任何计算环境来实施。再者,当前揭示标的物的方面可在多个处理芯片或装置中实施或跨多个处理芯片或装置实施,且存储可类似地跨多个装置实现。这些装置可能包含例如PC、网络服务器和手持式装置。

[0206] 虽然已用特定于结构特征和/或方法动作的语言描述了标的物,但应了解,所附权利要求书中界定的标的物不一定限于上述特定特征或动作。而是,上述特定特征和动作是作为实施权利要求书的实例形式来揭示。

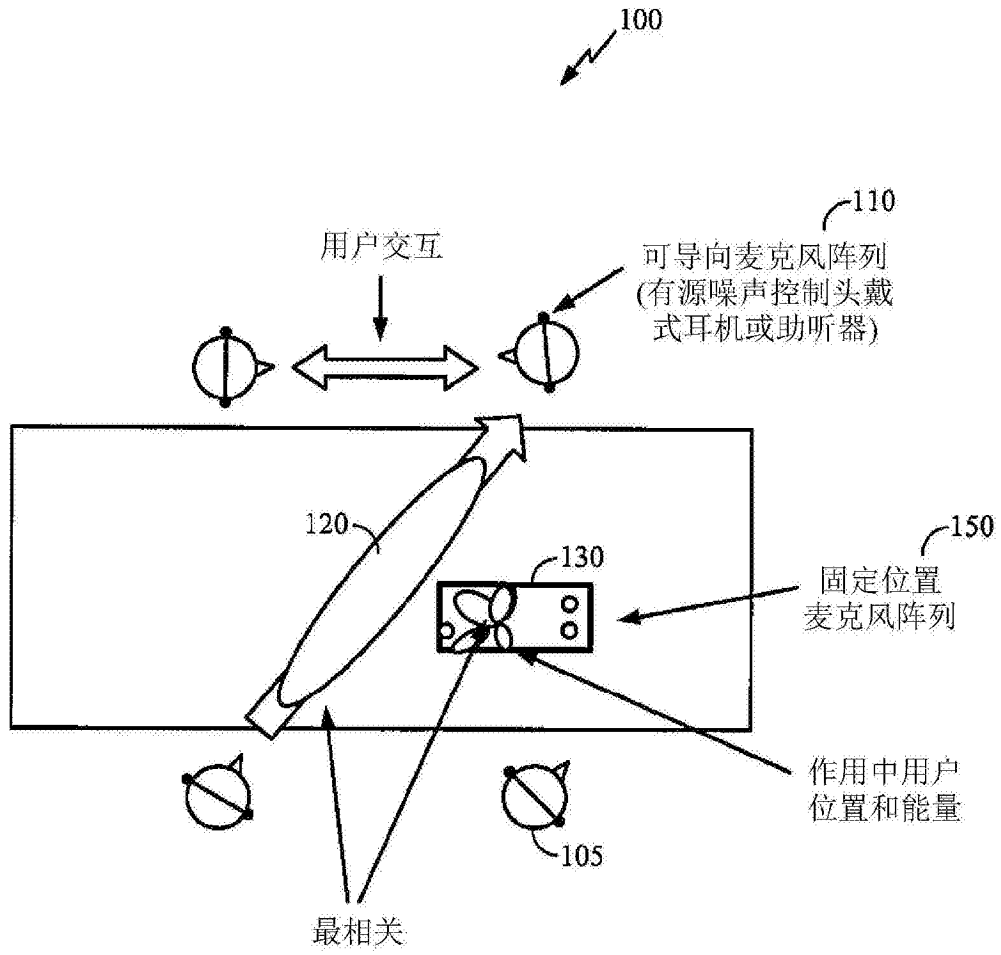


图1

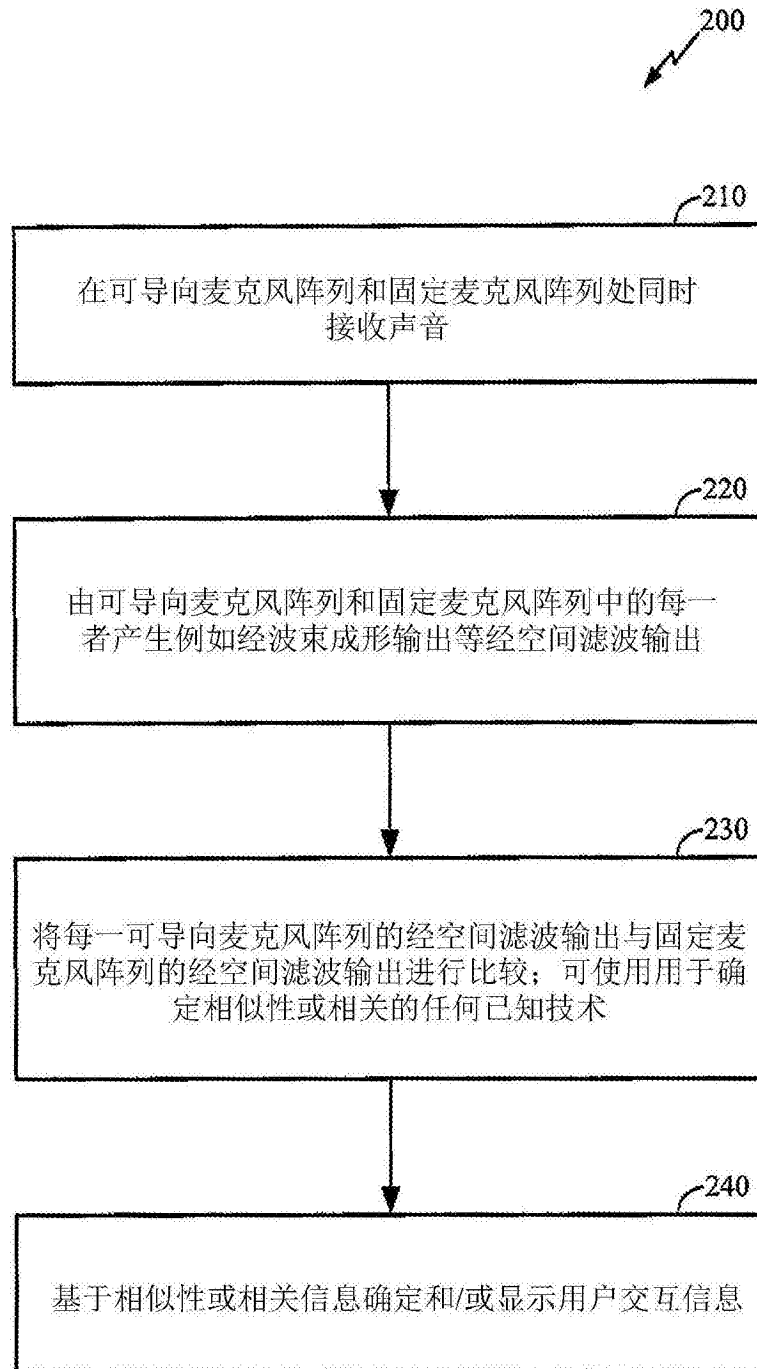


图2

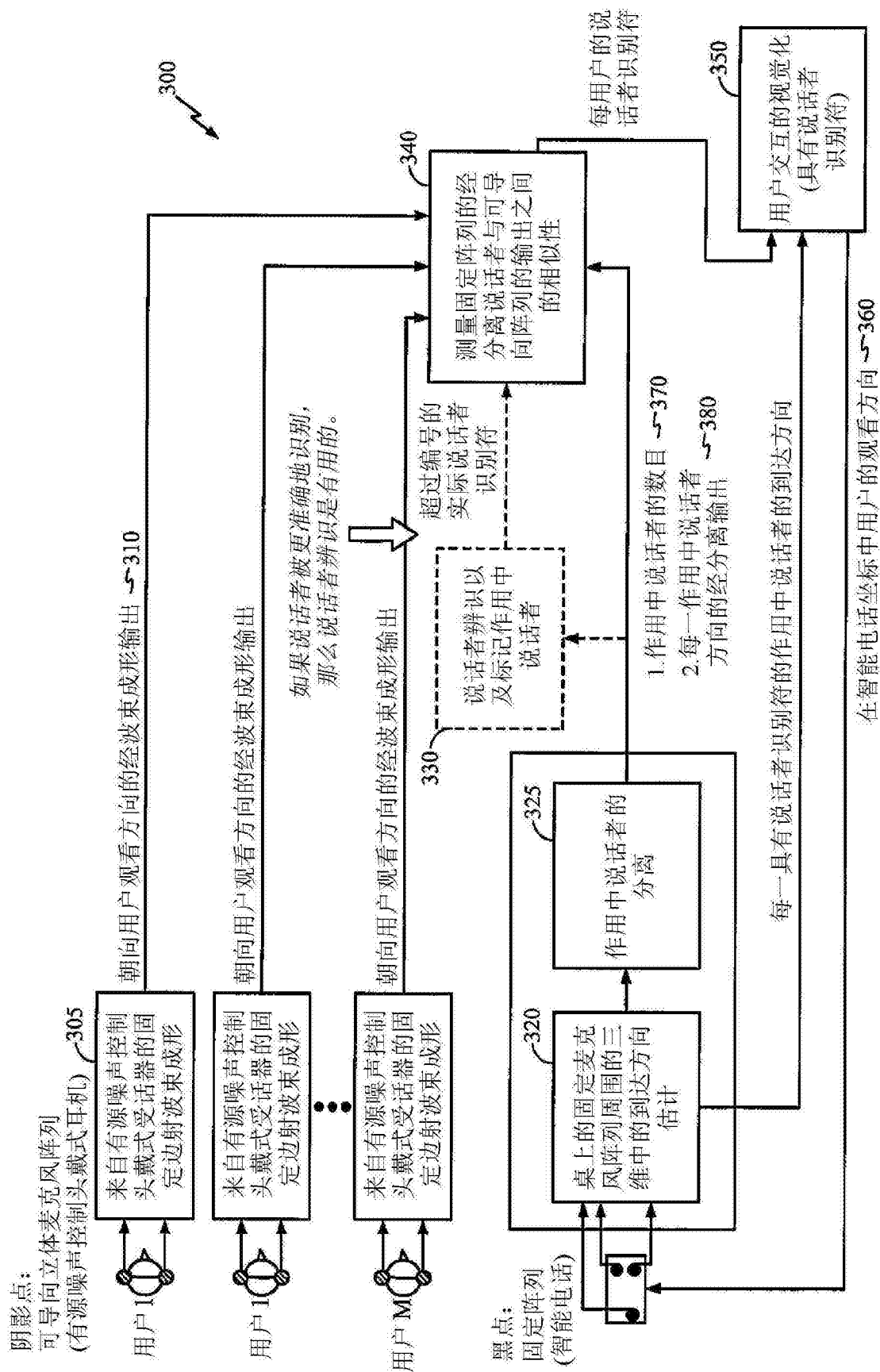


图3

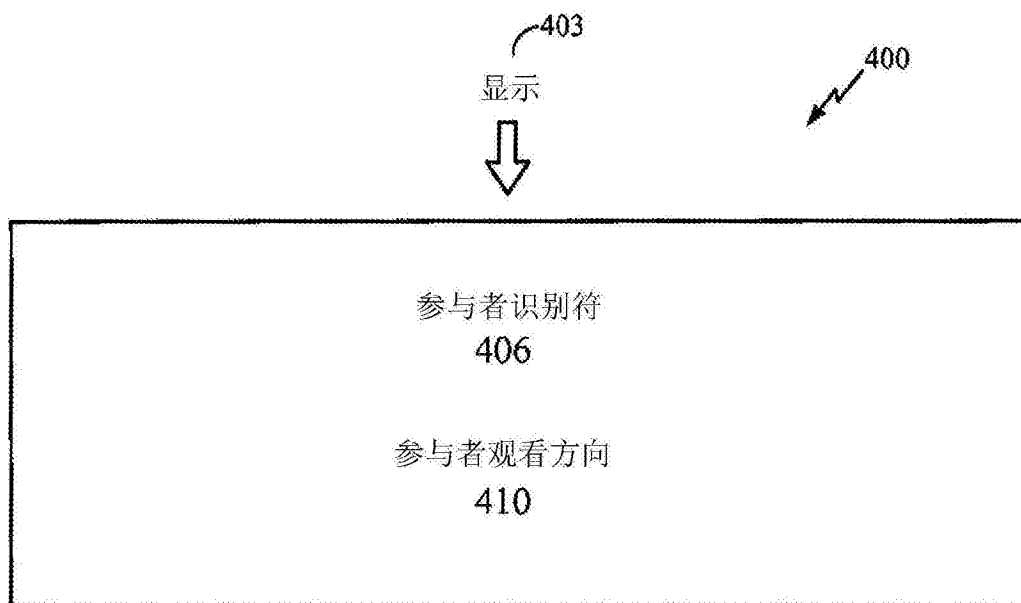


图4

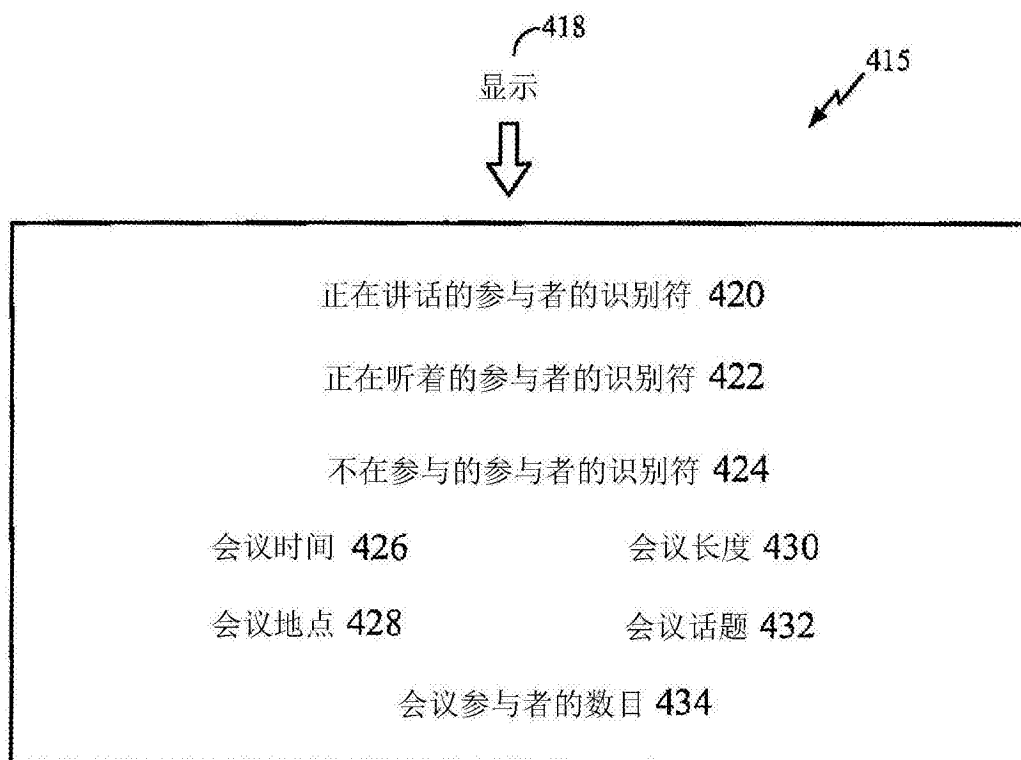


图5

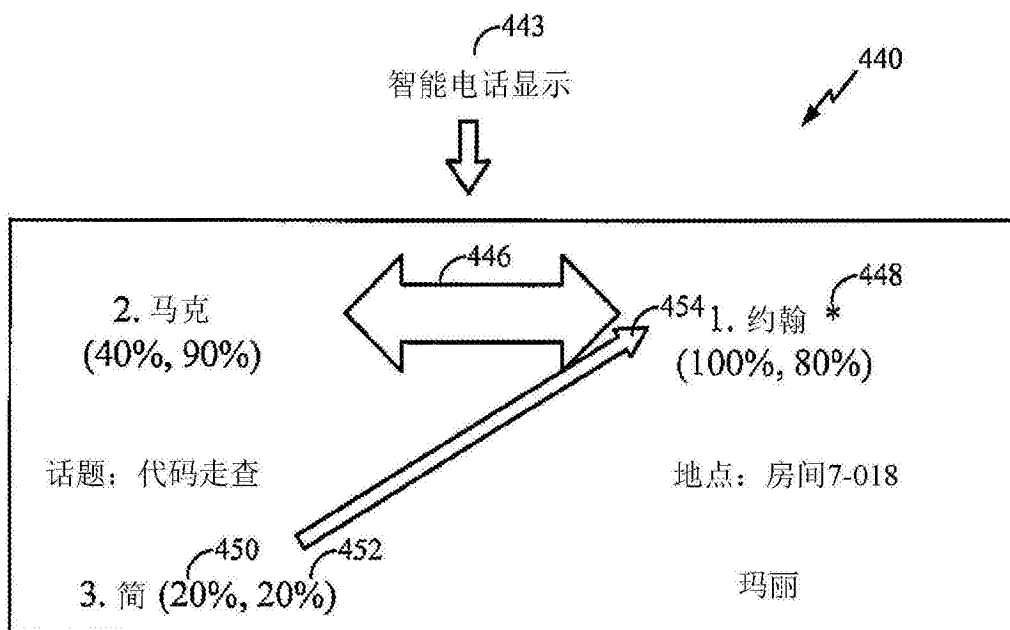


图6

显示

462

460

话题:	设计	代码走查	文档化
识别符:		参与比率:	
简	20%	40% (L)	10%
约翰	60% (L)	30%	30%
马克	15%	15%	15%
玛丽	5%	15%	45% (L)

图7

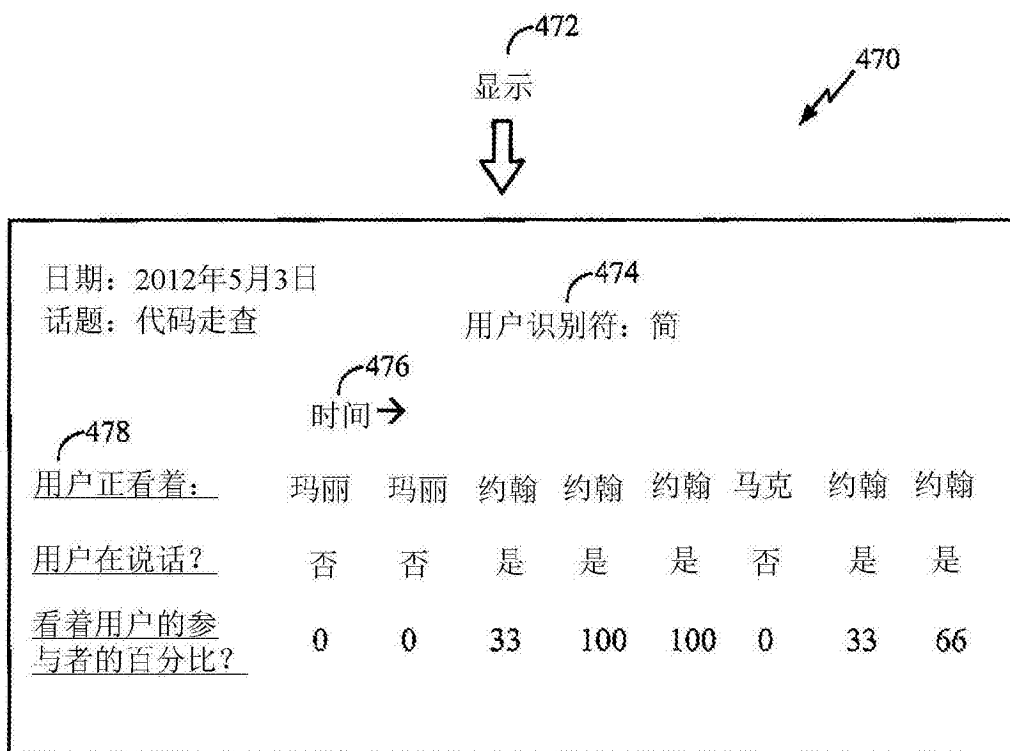


图8

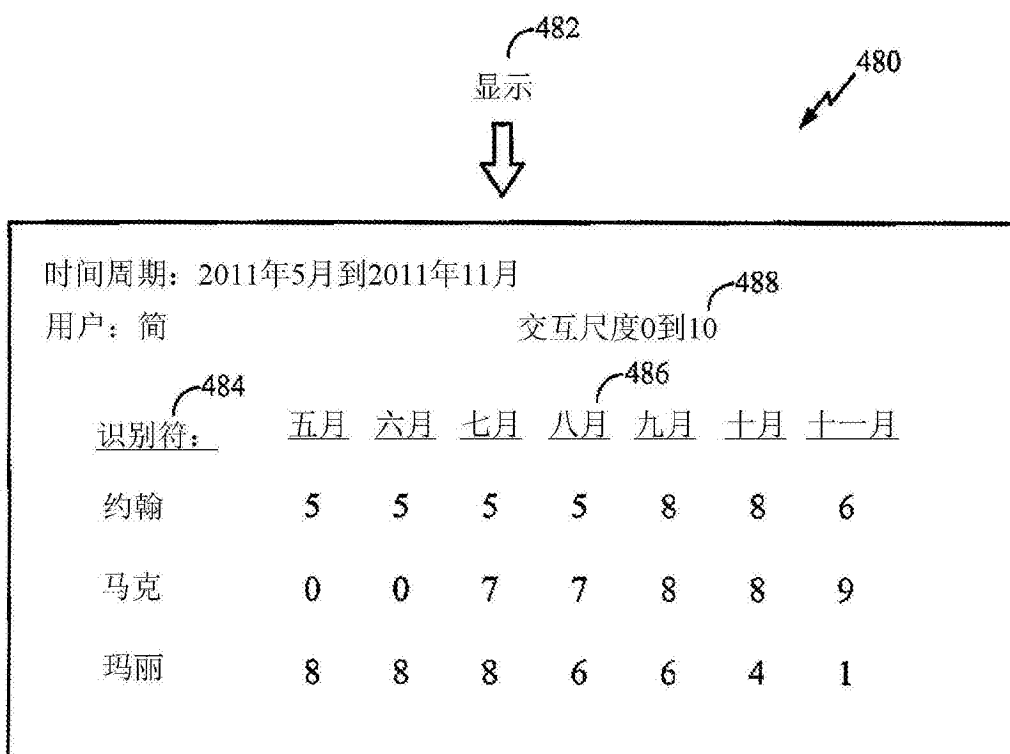


图9

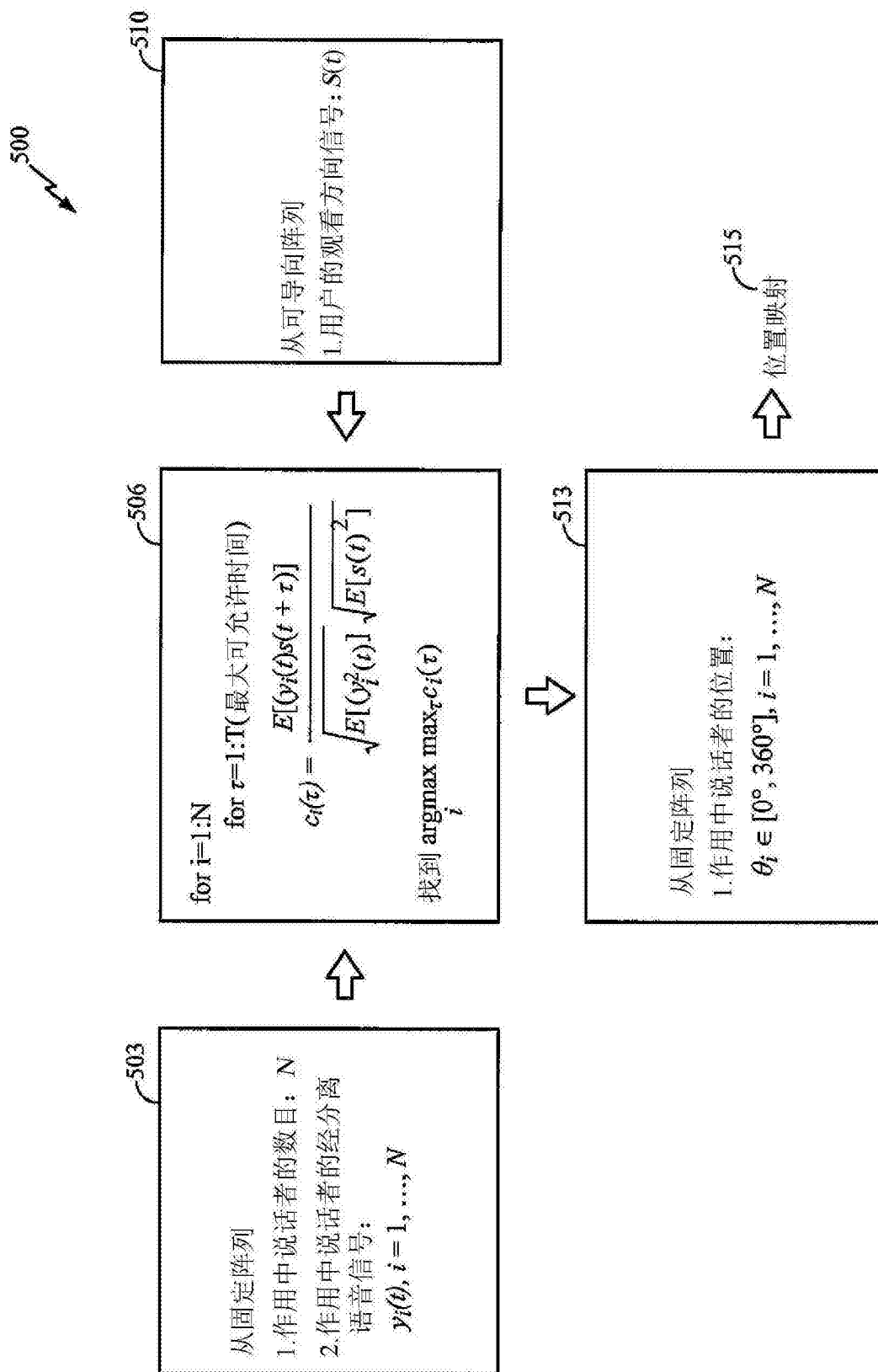


图10

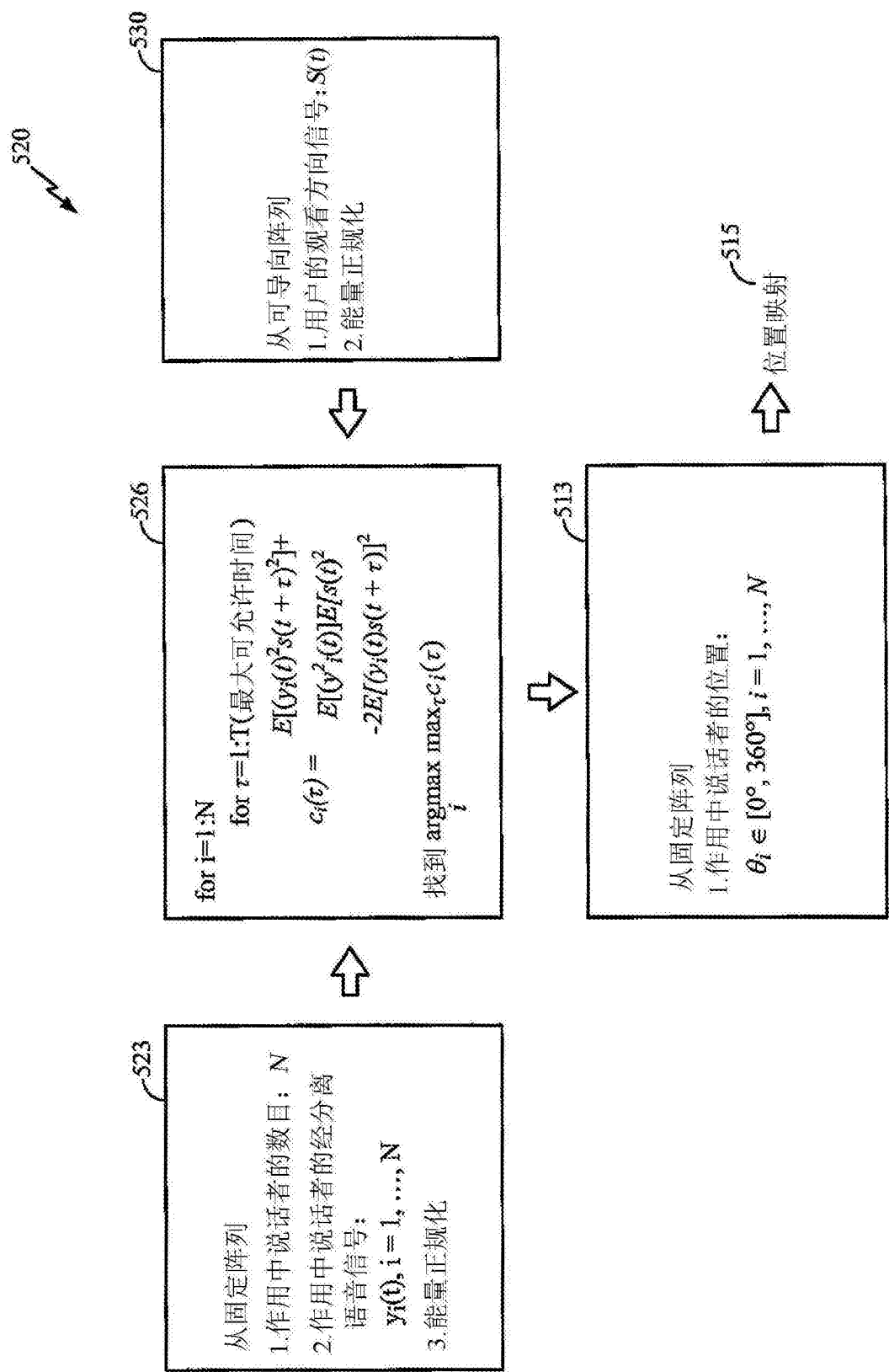


图11

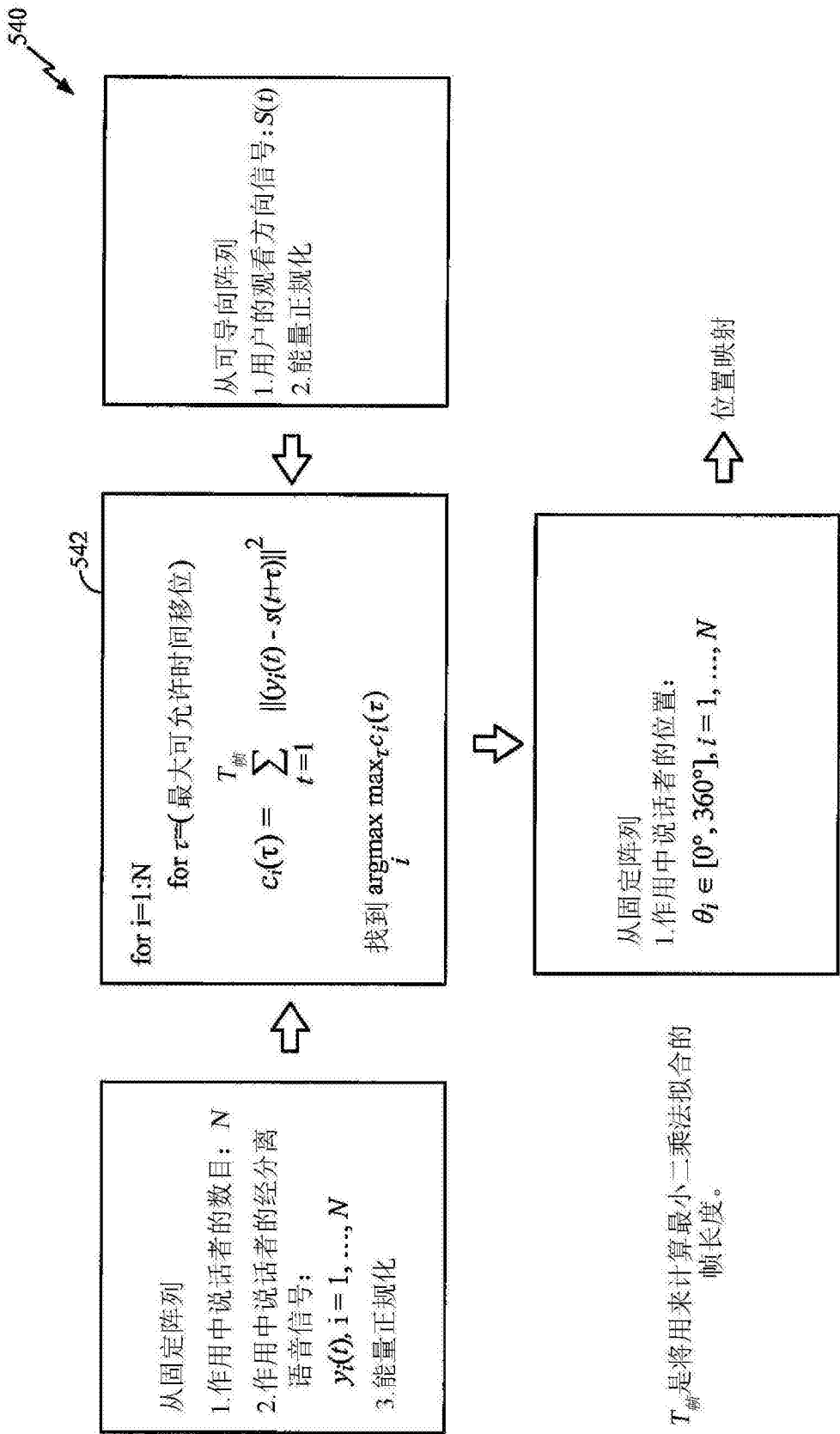


图12

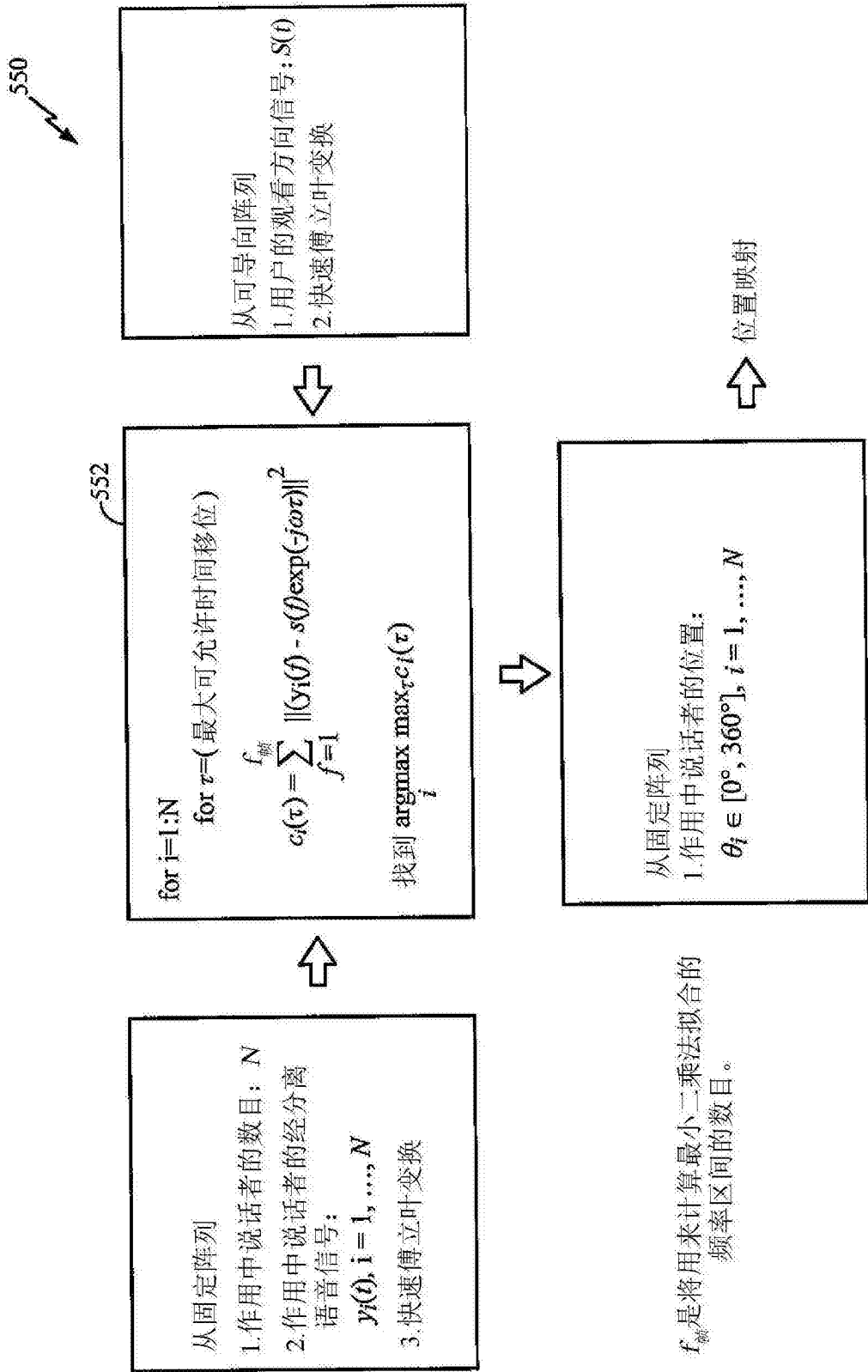


图13

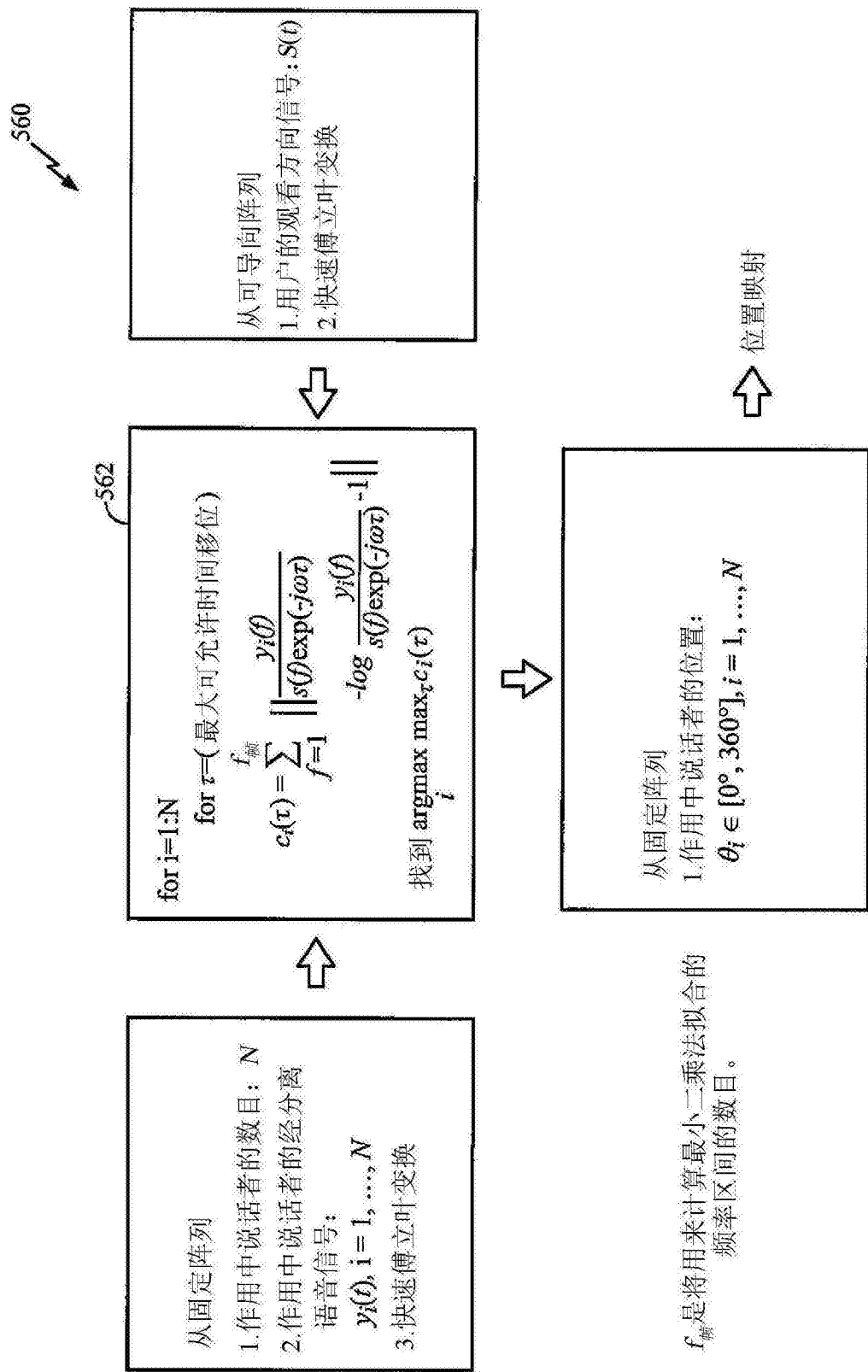


图14

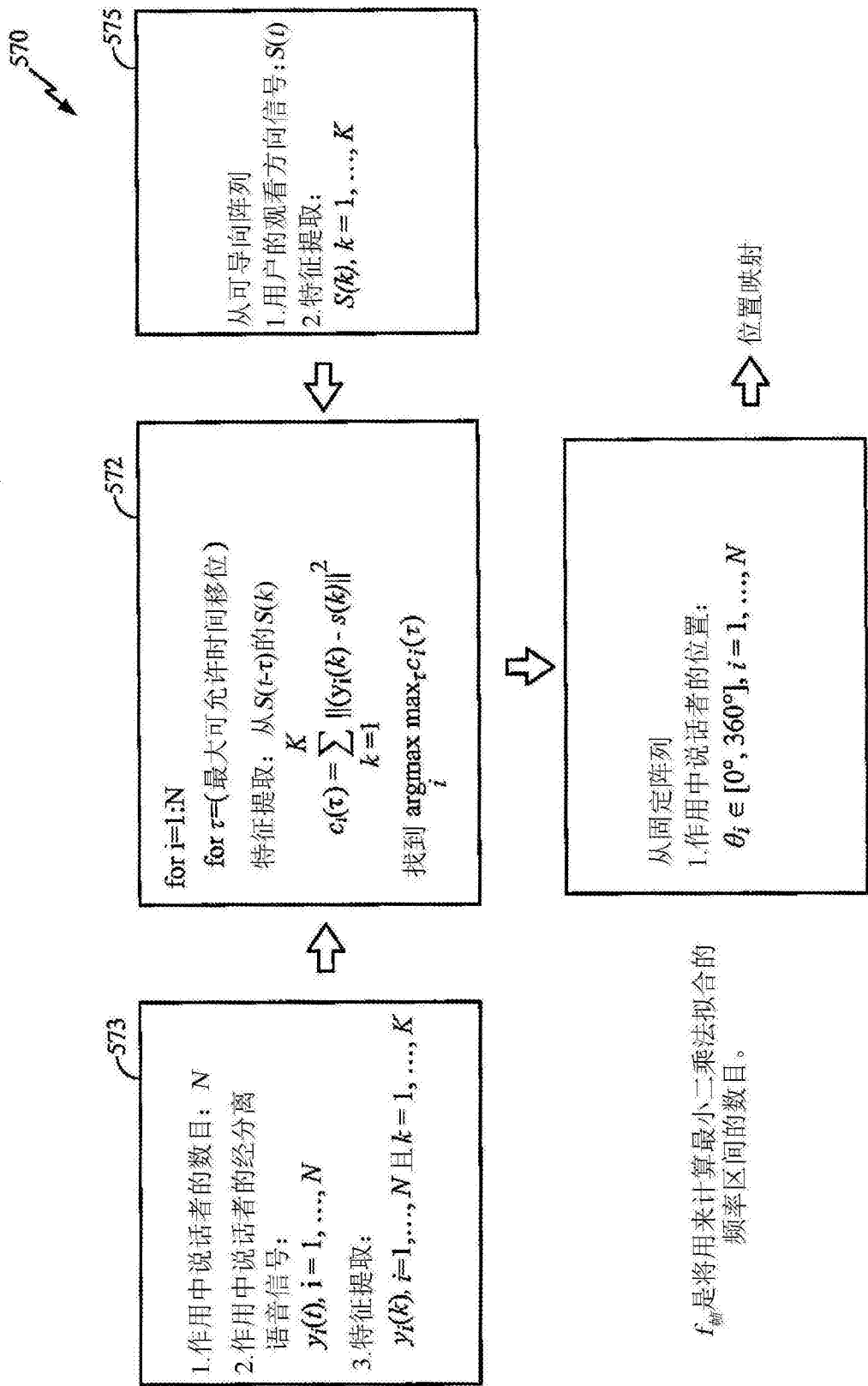


图15

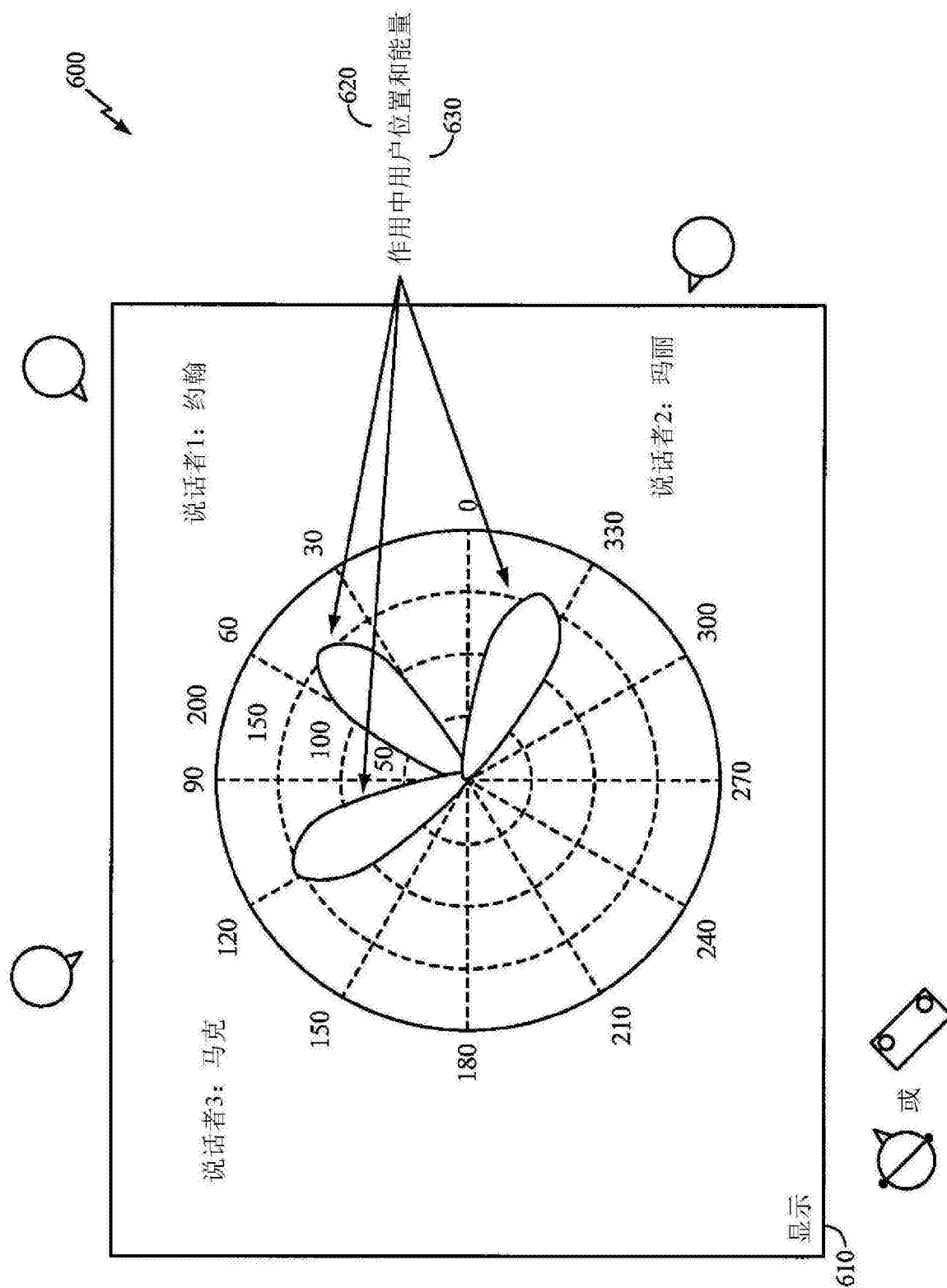


图16

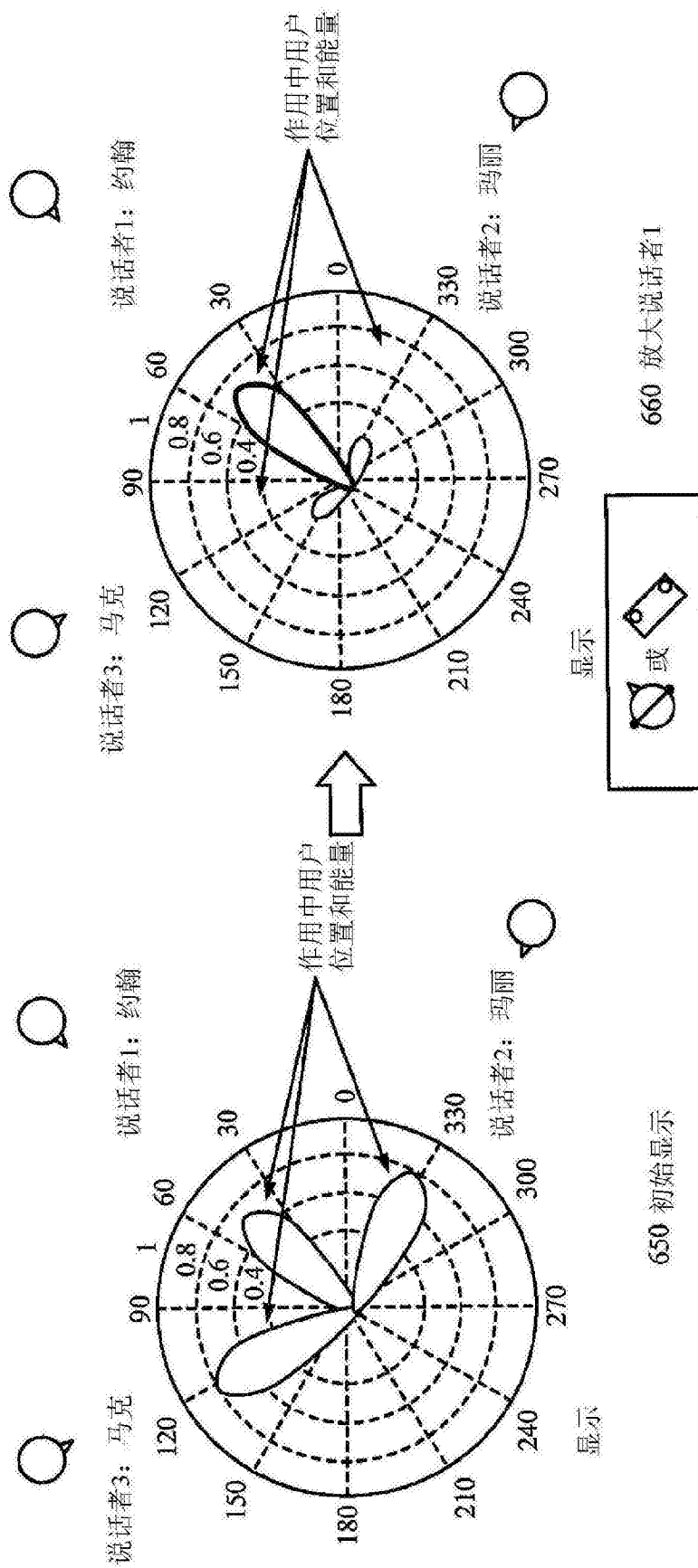


图17

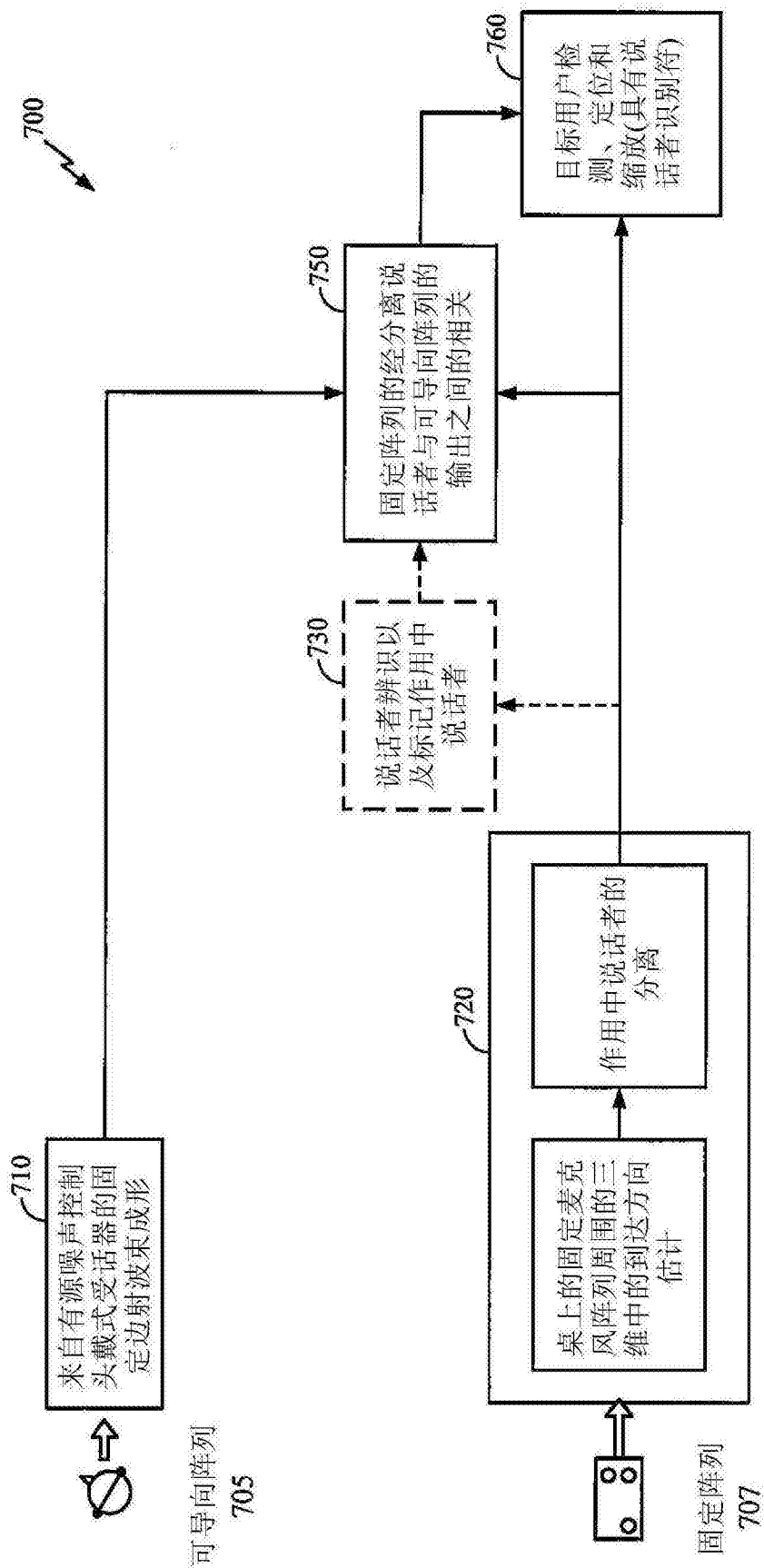


图18

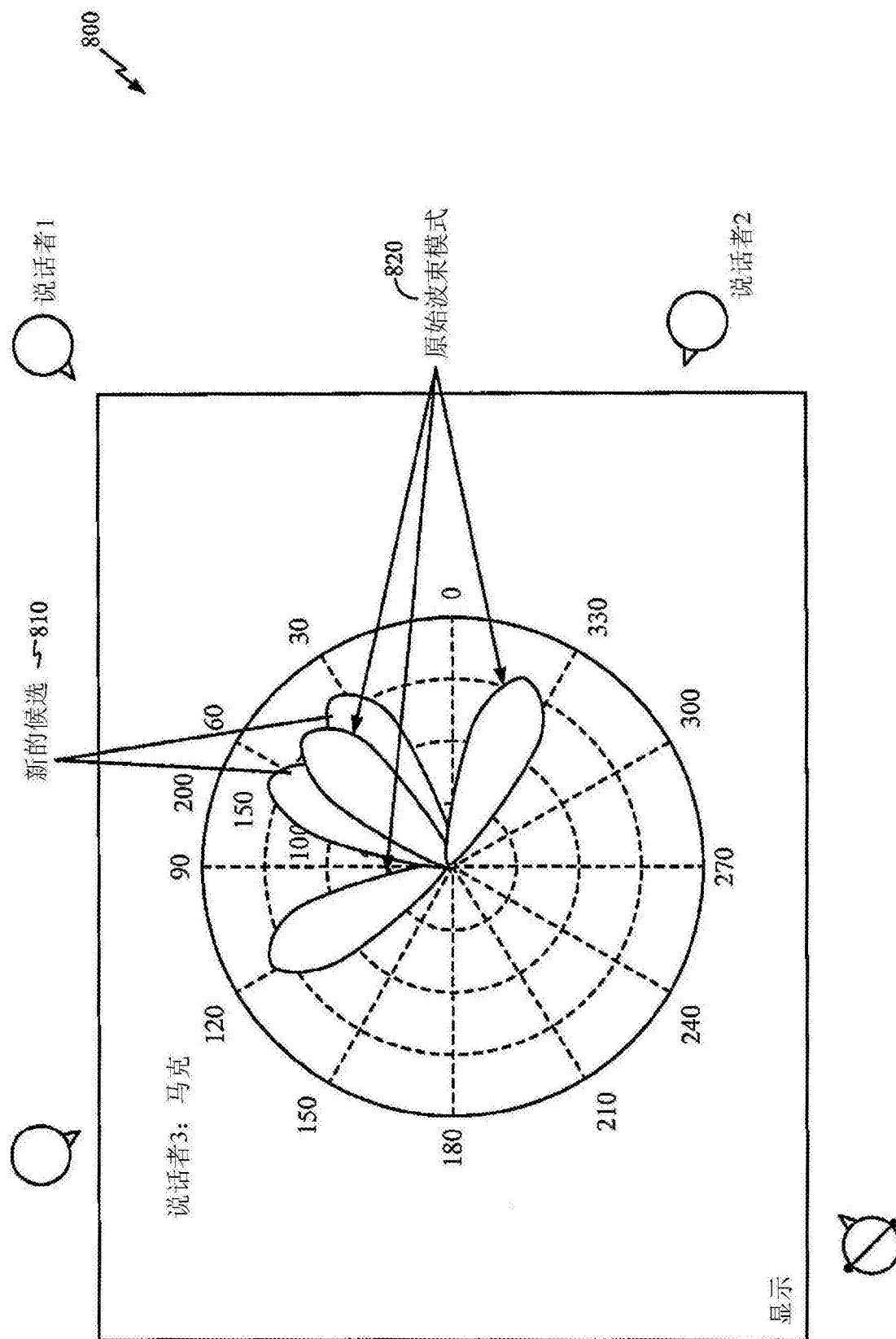


图19

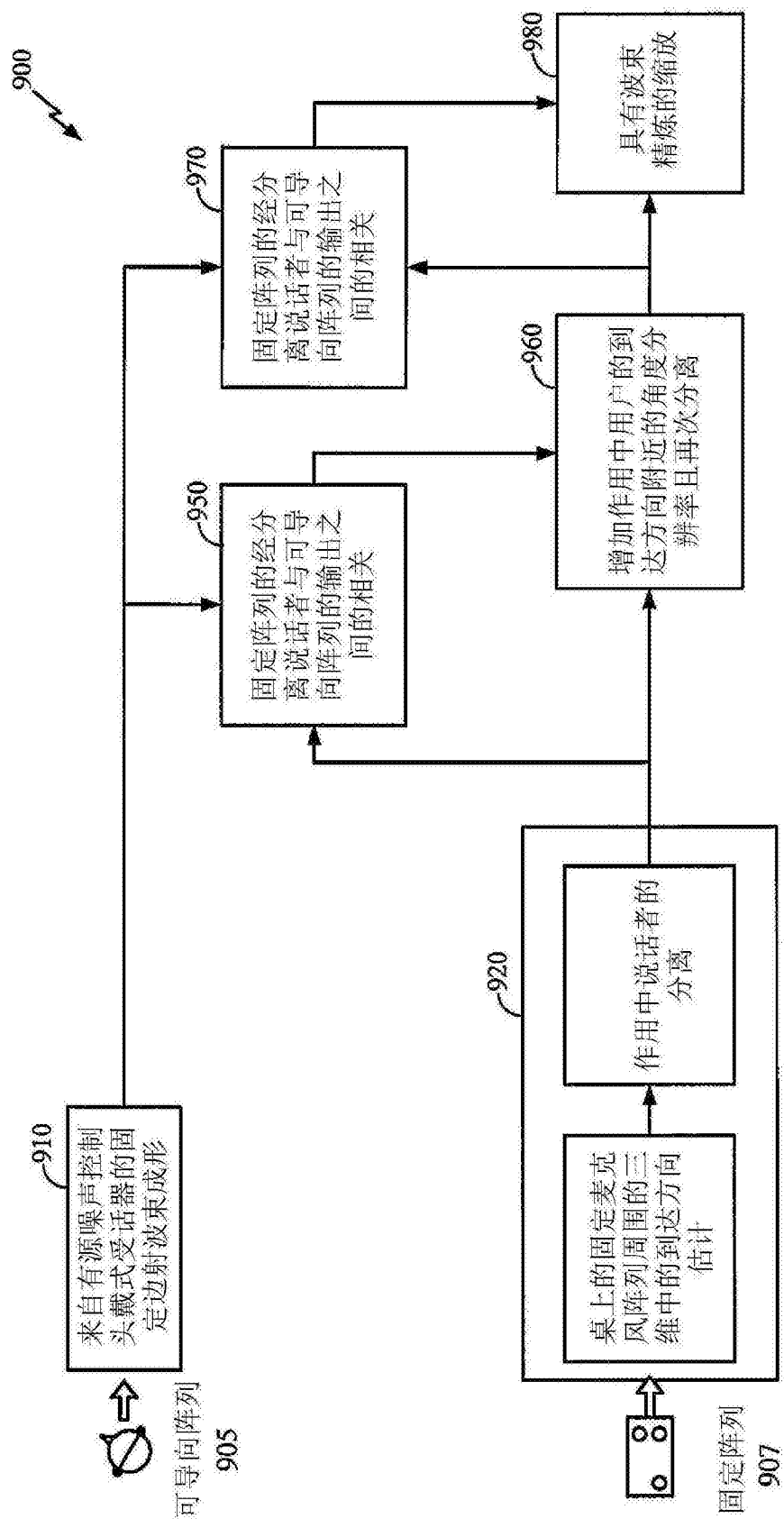


图20

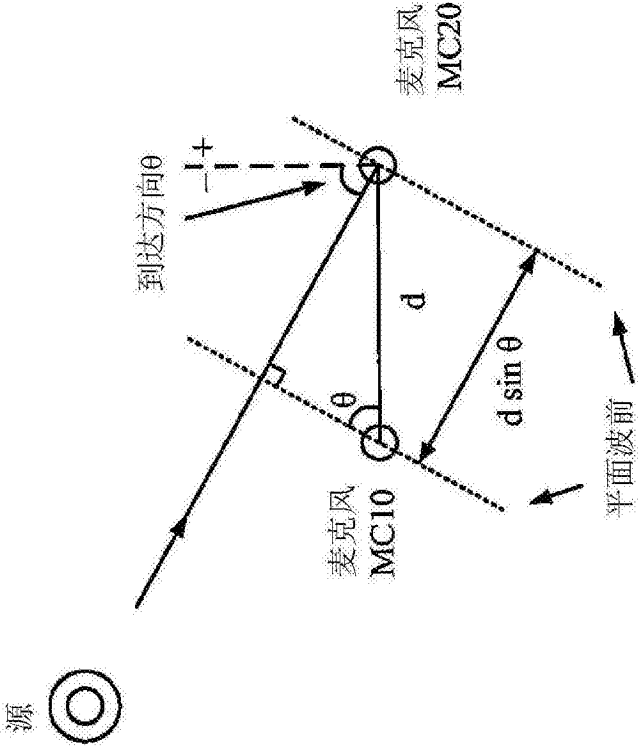


图21

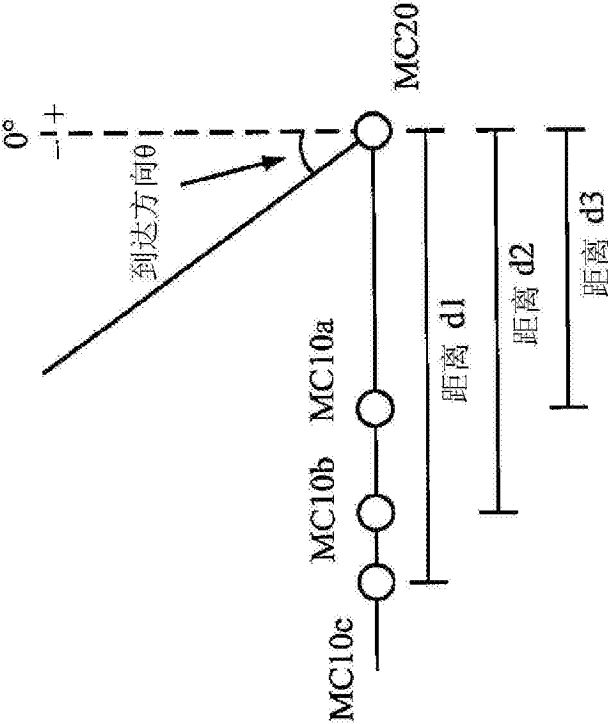
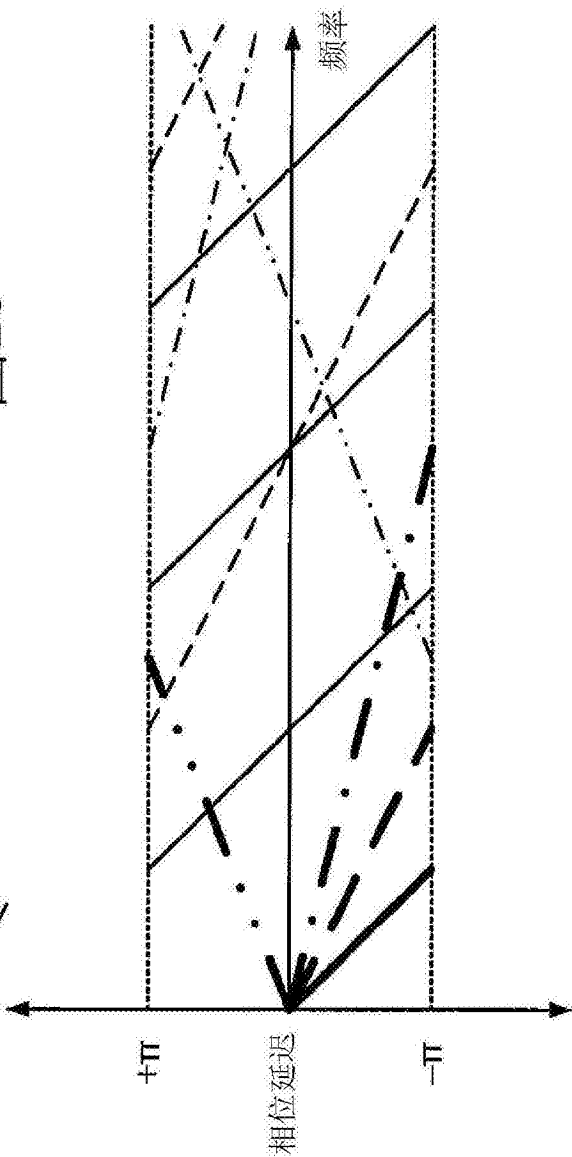
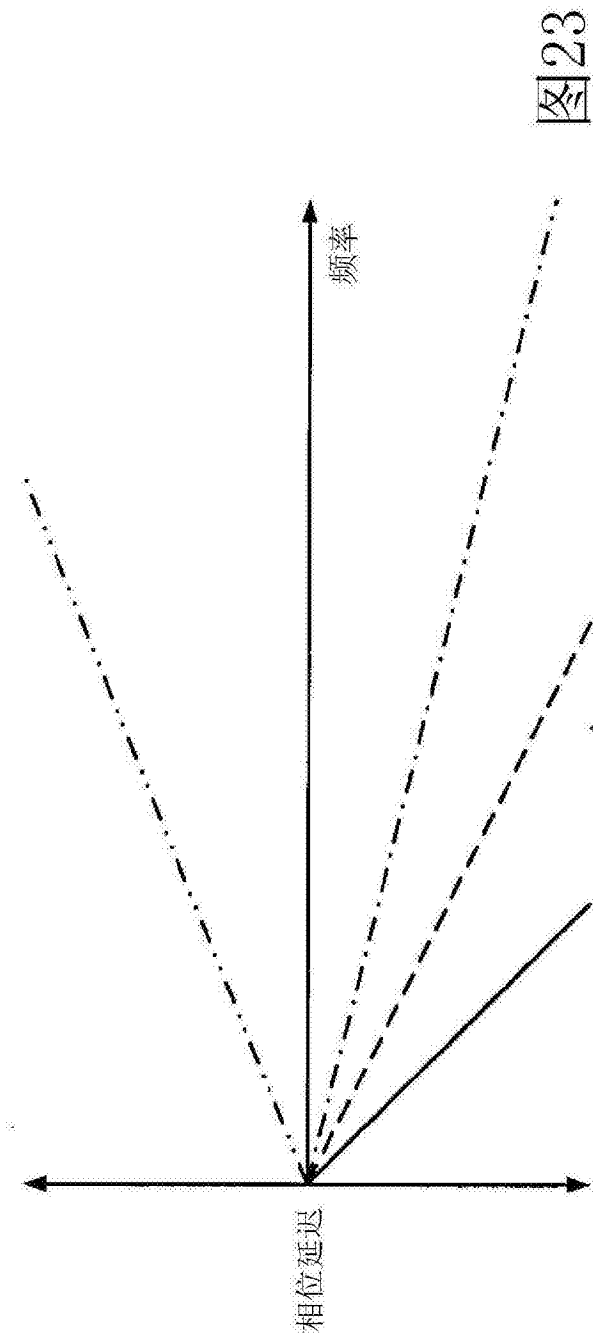


图22



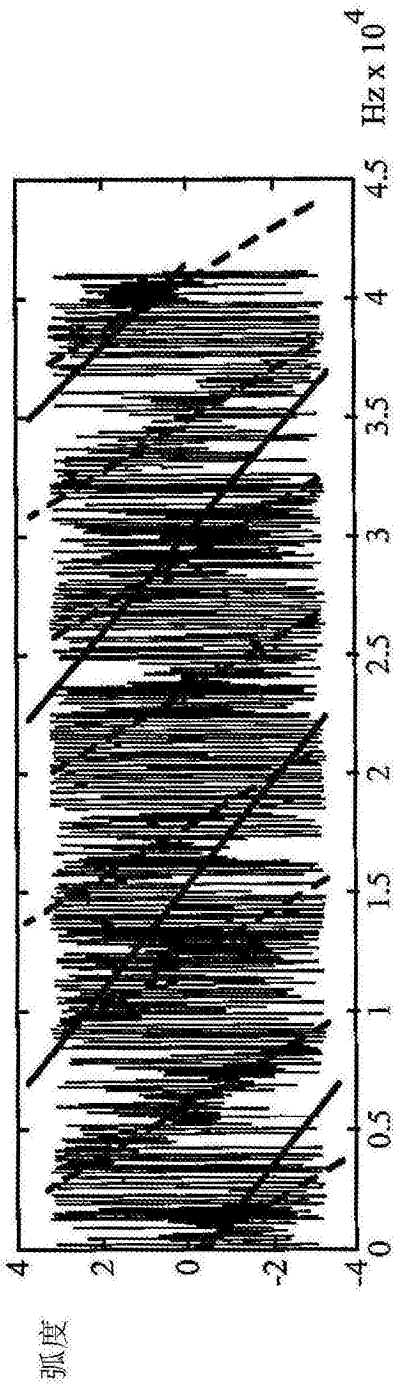


图25

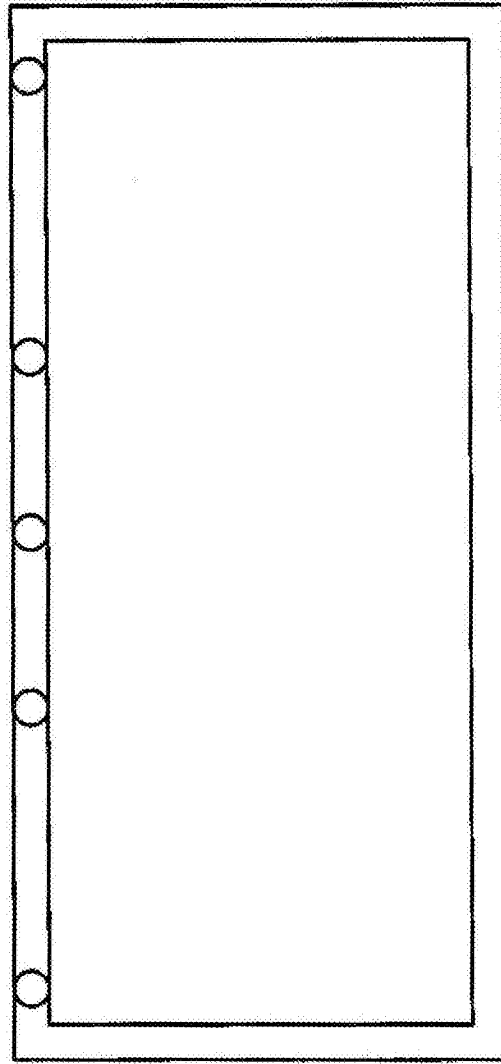


图26

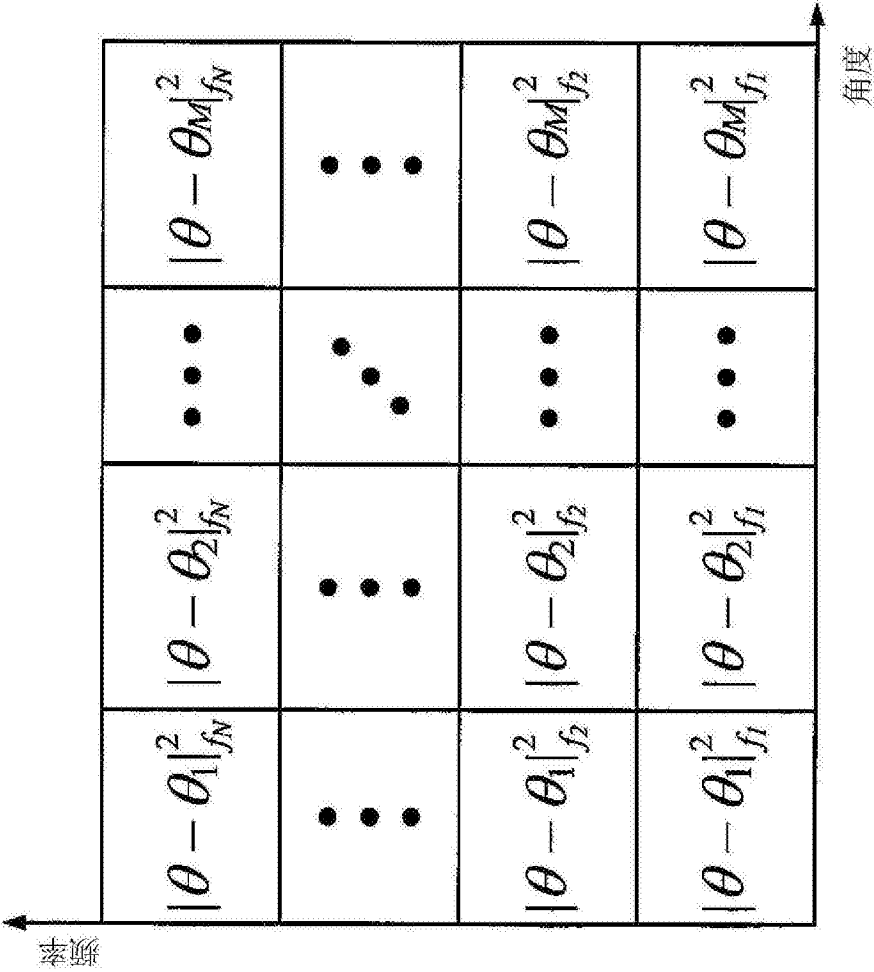


图27

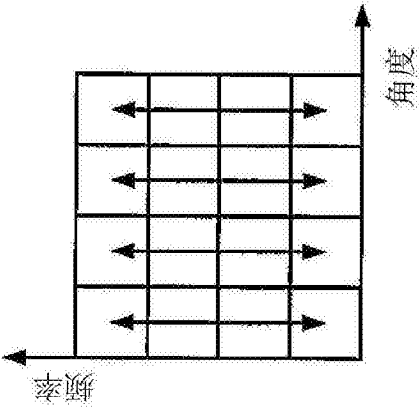


图28

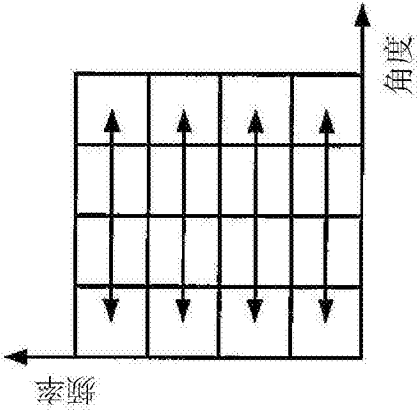


图29

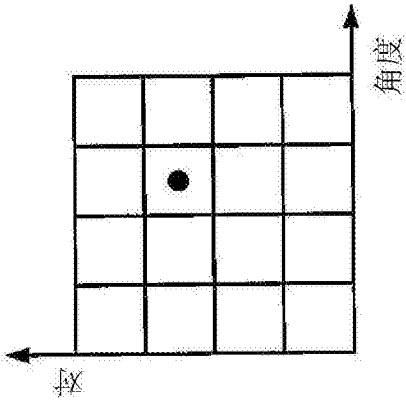


图30

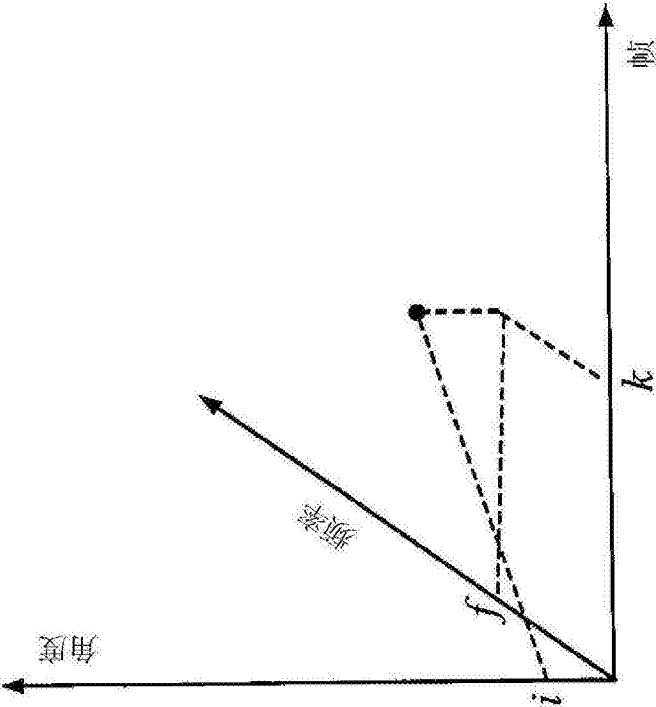


图31

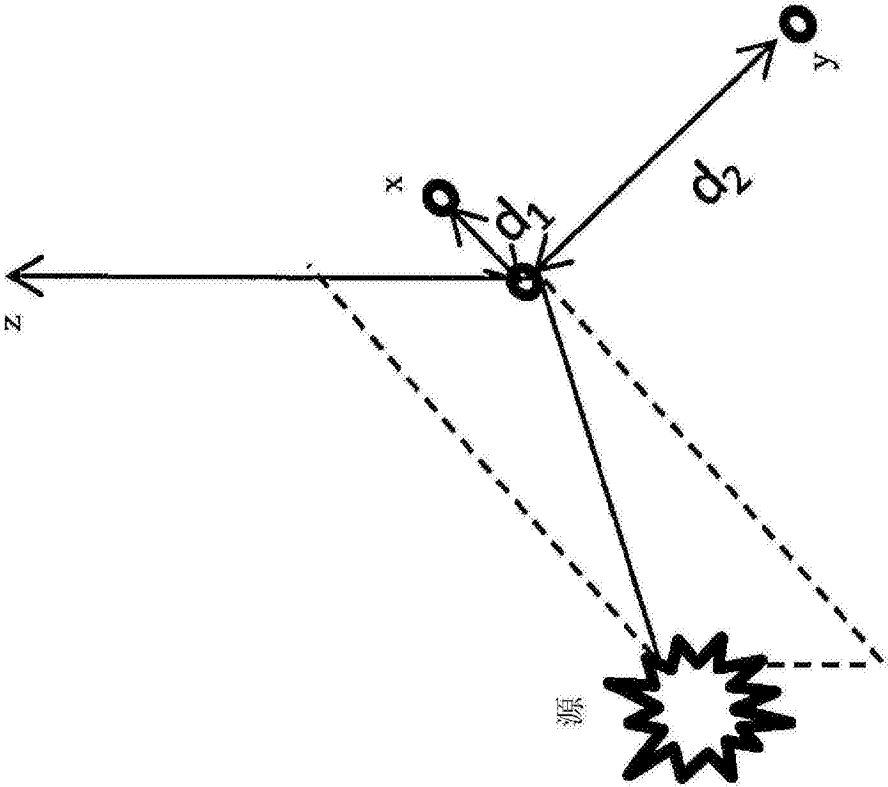


图32

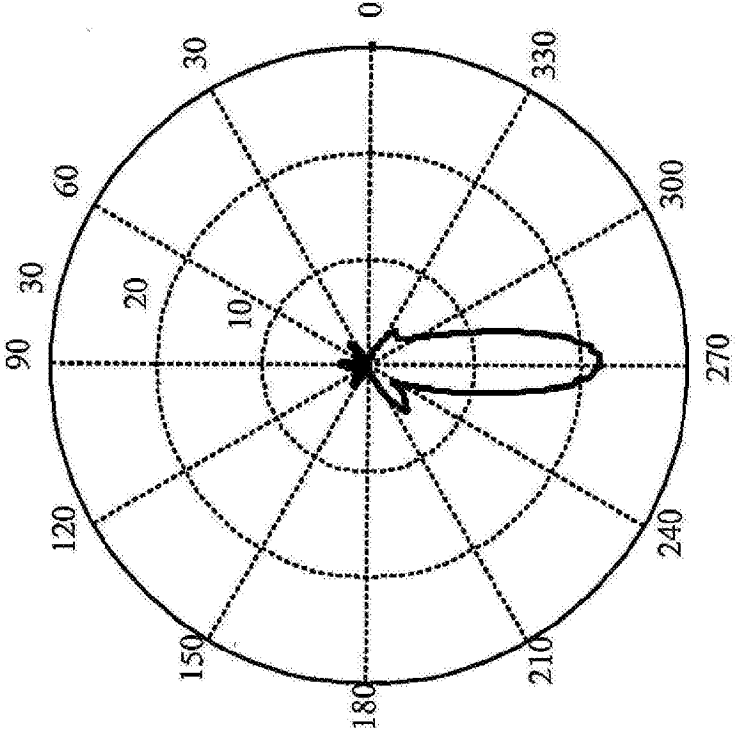


图33

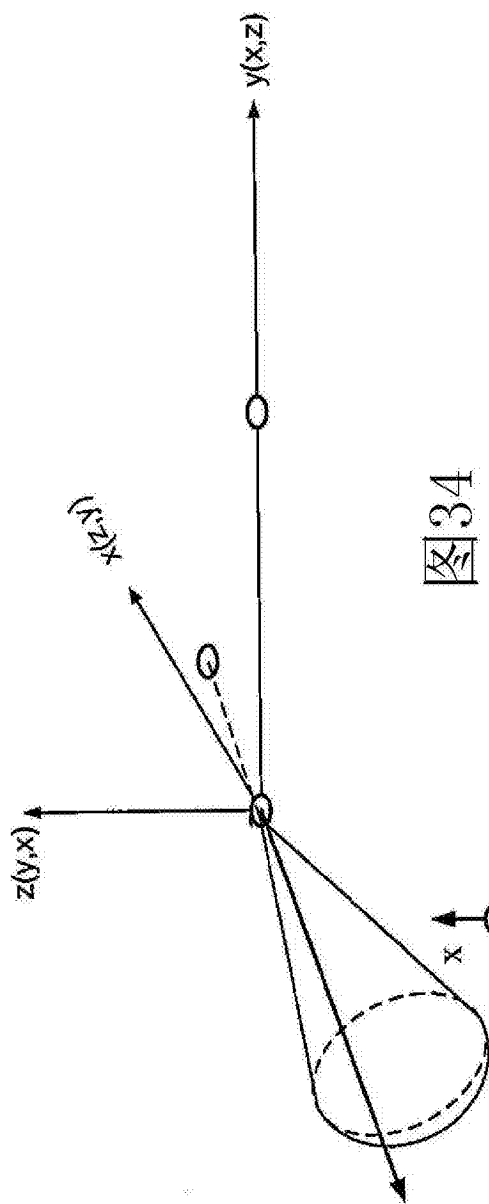


图34

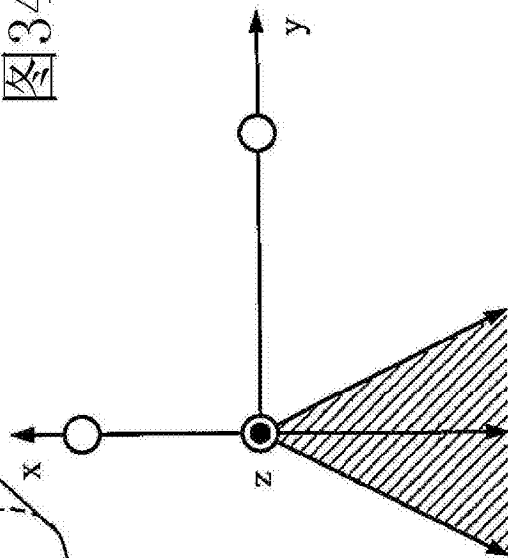


图35

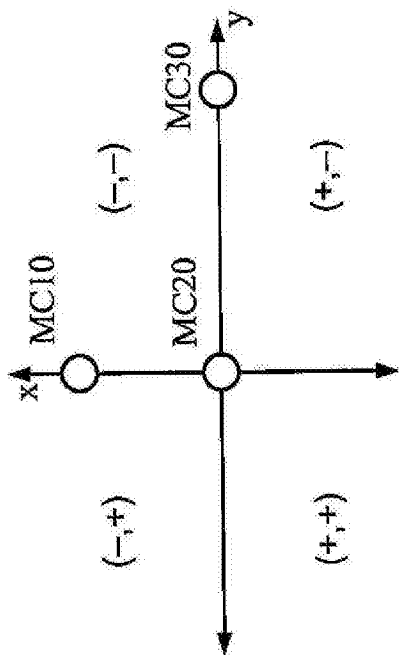


图36

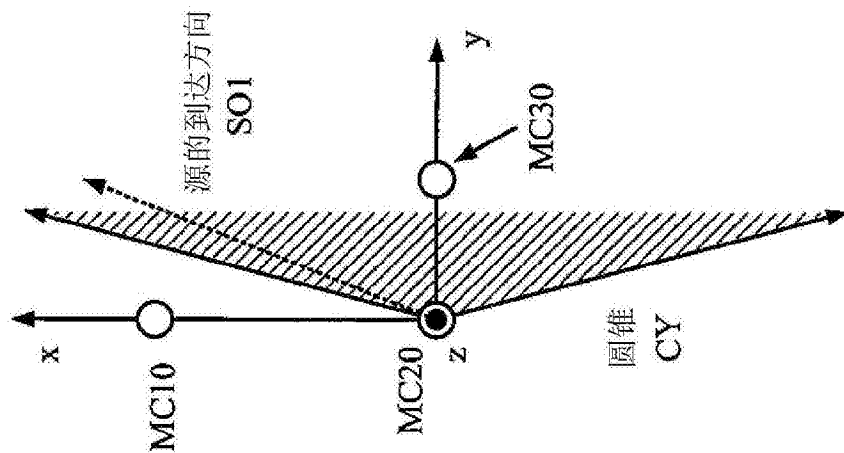


图37

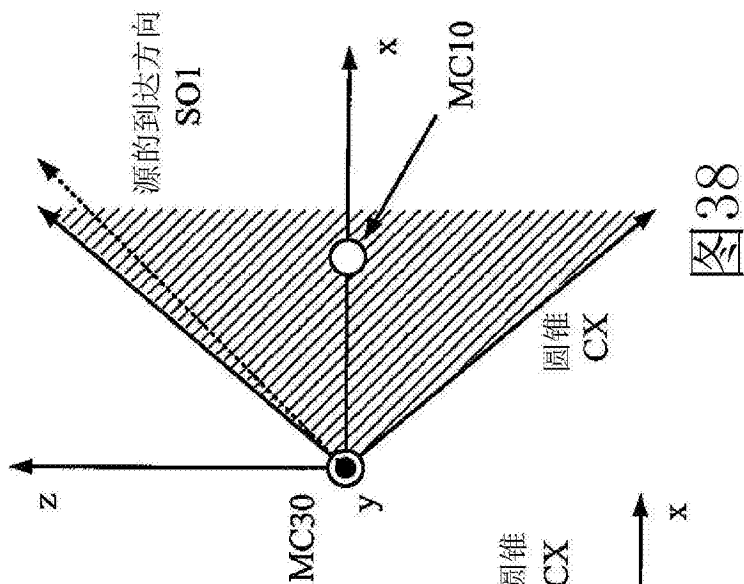


图38

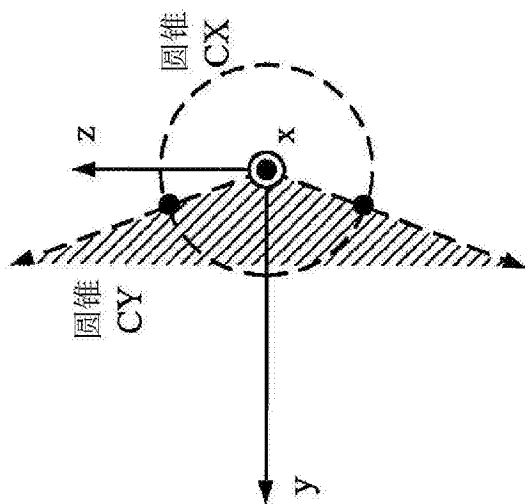


图39

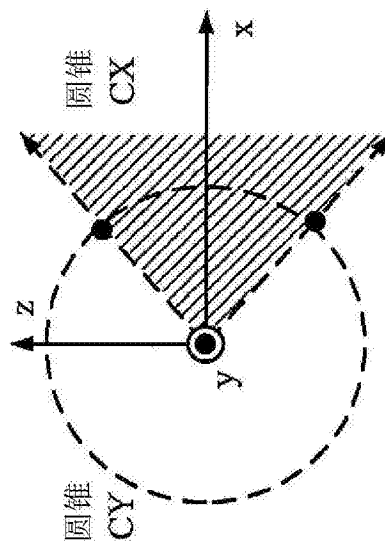


图40

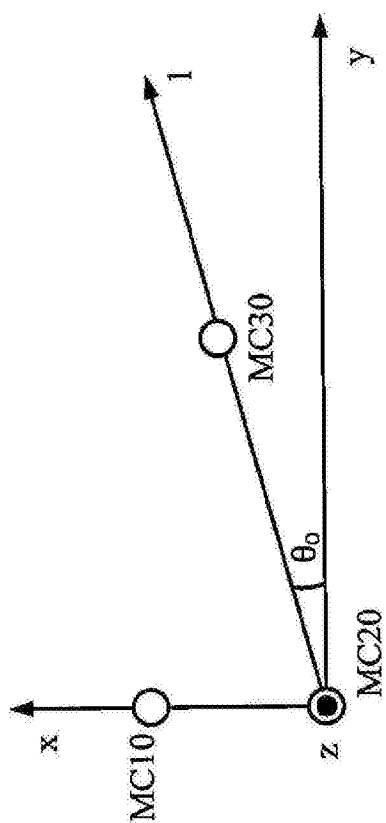


图41

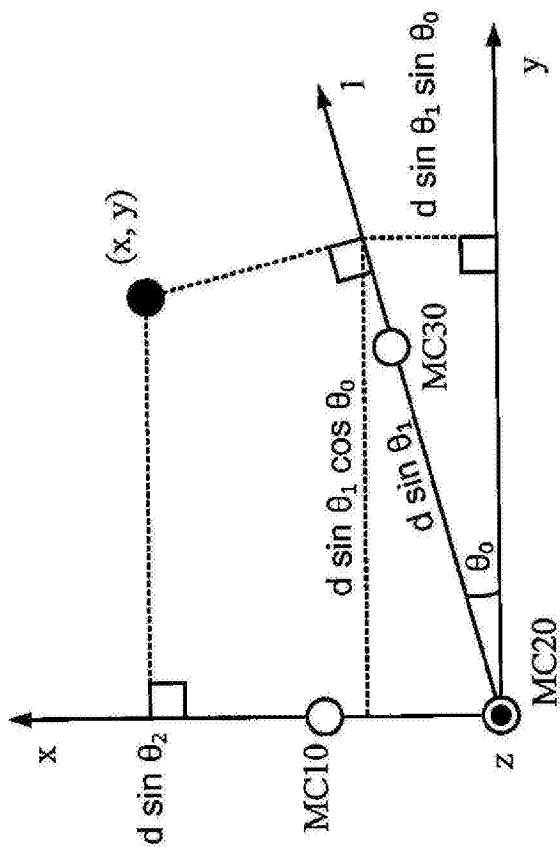


图42

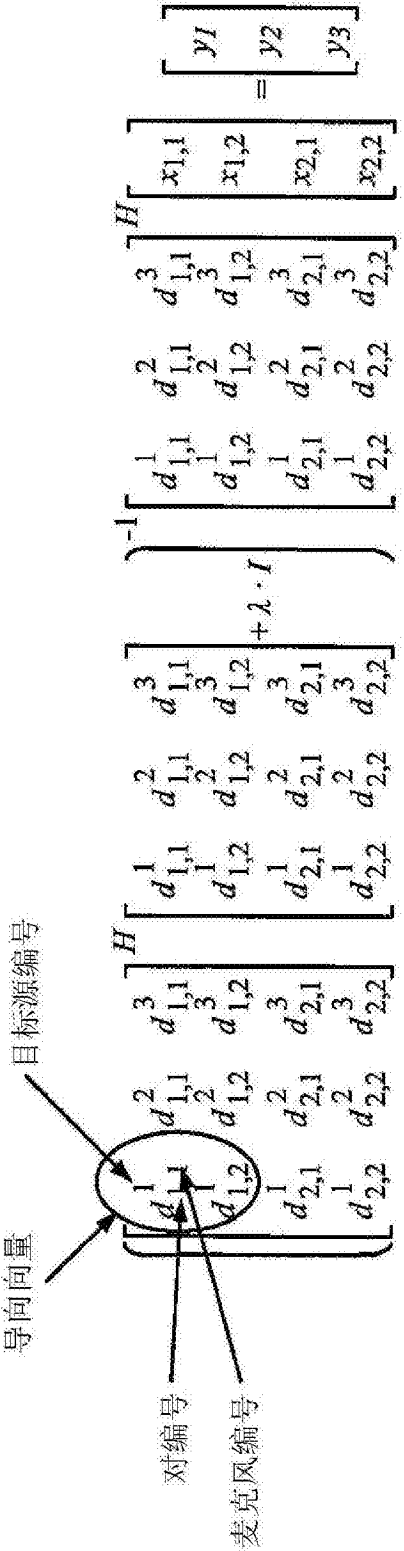


图43

[illegible]

图44



图45

$$\begin{array}{c} \left[\begin{array}{c} d_{1,1}^1 \\ d_{1,1}^2 \\ d_{1,1}^3 \\ d_{1,2}^1 \\ d_{1,2}^2 \\ d_{1,2}^3 \\ d_{2,2}^1 \\ d_{2,2}^2 \\ d_{2,2}^3 \end{array} \right]^H \quad (\Gamma + \lambda \cdot D)^{-1} \quad \left[\begin{array}{c} d_{1,1}^1 \\ d_{1,1}^2 \\ d_{1,1}^3 \\ d_{1,2}^1 \\ d_{1,2}^2 \\ d_{1,2}^3 \\ d_{2,2}^1 \\ d_{2,2}^2 \\ d_{2,2}^3 \end{array} \right]^{-1} \quad (\Gamma + \lambda \cdot D)^{-1} \quad \left[\begin{array}{c} x_{1,1} \\ x_{1,2} \\ x_{2,2} \end{array} \right] \end{array}$$

图46

目标源编号 1, 2, ..., N, 其中 N(源的数目) ≤ M(对的数目)

$$\begin{pmatrix}
 \begin{matrix} \vdots \\ d_{1,1}^1 \\ d_{1,2}^1 \\ \vdots \\ d_{M,1}^1 \\ d_{M,2}^1 \end{matrix} &
 \begin{matrix} \vdots \\ d_{1,1}^1 \\ d_{1,2}^1 \\ \vdots \\ d_{M,1}^1 \\ d_{M,2}^1 \end{matrix} &
 \begin{matrix} \vdots \\ d_{1,1}^N \\ d_{1,2}^N \\ \vdots \\ d_{M,1}^N \\ d_{M,2}^N \end{matrix} &
 \begin{matrix} \vdots \\ d_{1,1}^N \\ d_{1,2}^N \\ \vdots \\ d_{M,1}^N \\ d_{M,2}^N \end{matrix} &
 \begin{matrix} \vdots \\ d_{1,1}^N \\ d_{1,2}^N \\ \vdots \\ d_{M,1}^N \\ d_{M,2}^N \end{matrix} &
 \begin{matrix} \vdots \\ d_{1,1}^N \\ d_{1,2}^N \\ \vdots \\ d_{M,1}^N \\ d_{M,2}^N \end{matrix} \\
 \end{pmatrix}
 + \lambda \cdot I
 \begin{pmatrix}
 \vdots \\
 d_{1,1}^1 \\
 d_{1,2}^1 \\
 \vdots \\
 d_{M,1}^1 \\
 d_{M,2}^1 \\
 \vdots \\
 d_{1,1}^N \\
 d_{1,2}^N \\
 \vdots \\
 d_{M,1}^N \\
 d_{M,2}^N \\
 \vdots \\
 d_{1,1}^N \\
 d_{1,2}^N \\
 \vdots \\
 d_{M,1}^N \\
 d_{M,2}^N
 \end{pmatrix}
 \begin{matrix}
 H \\
 H \\
 H \\
 H \\
 H \\
 H
 \end{matrix}
 \begin{bmatrix}
 x_{1,1} \\
 x_{1,2} \\
 x_{2,1} \\
 x_{2,2} \\
 \vdots \\
 x_{M,1} \\
 x_{M,2}
 \end{bmatrix}
 \parallel
 \begin{bmatrix}
 y_1 \\
 y_2 \\
 y_3
 \end{bmatrix}$$

每对的导向向量

对编号

麦克风编号

图47

$$\begin{aligned} (d_{1,1}^1, d_{1,2}^1) &= \left(\exp\left(-\frac{j2\pi f 0.073 \sin(\theta_1)}{c}\right), 1 \right) \\ (d_{2,1}^1, d_{2,2}^1) &= \left(\exp\left(-\frac{j2\pi f 0.036 \sin(\theta_2)}{c}\right), 1 \right) \end{aligned}$$

图48

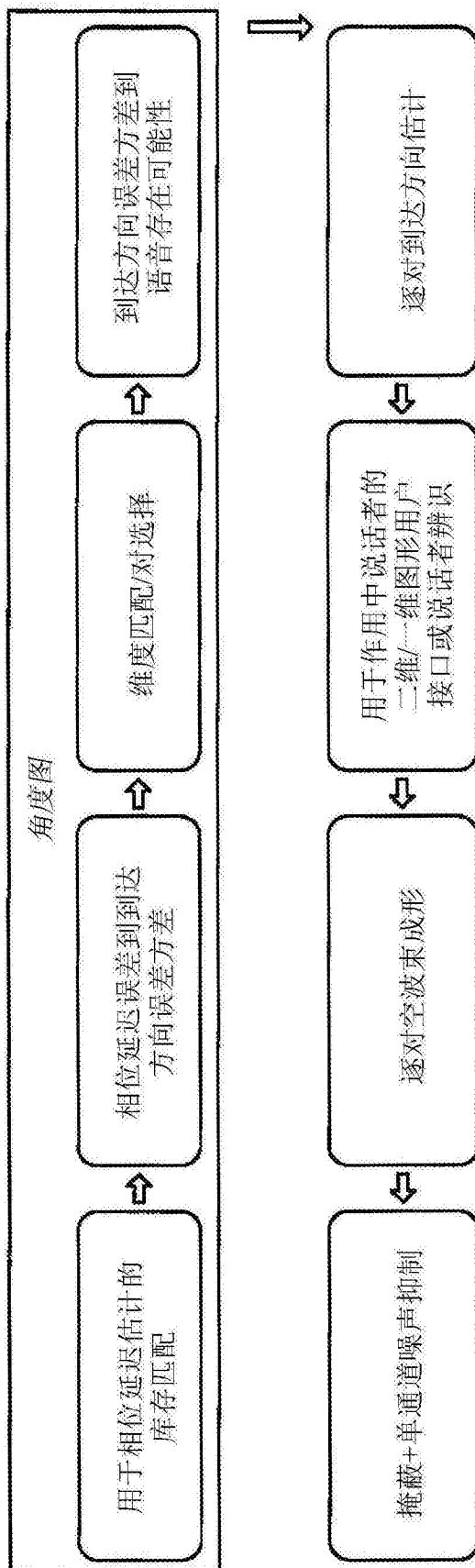


图49