

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 3/01 (2006.01)

G06F 3/16 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200810176176.7

[43] 公开日 2009年8月26日

[11] 公开号 CN 101515197A

[22] 申请日 2008.11.14

[21] 申请号 200810176176.7

[30] 优先权

[32] 2008. 2. 19 [33] JP [31] 2008 - 037534

[71] 申请人 株式会社日立制作所

地址 日本东京

[72] 发明人 户上真人 住吉贵志 大渊康成

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 吴丽丽

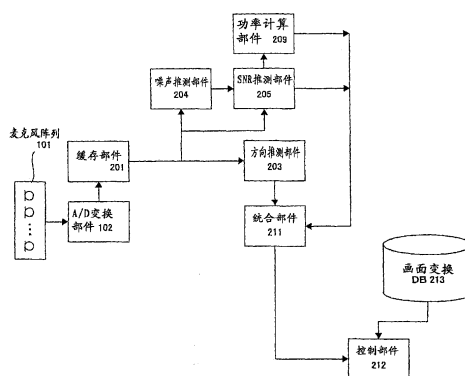
权利要求书 5 页 说明书 20 页 附图 18 页

[54] 发明名称

音响指示设备、音源位置的指示方法和计算机系统

[57] 摘要

本发明提供一种完全不将附属设备设置在桌子上就能够进行指示操作的音响指示设备。具备：具有 2 个以上麦克风元件的麦克风阵列(101)；将模拟音压数据变换为数字音压数据的 A/D 变换部件(102)；存储数字音压数据的缓存部件(201)；从数字音压数据，根据麦克风元件之间的声音的相关性，推测突发性音的音源方向的方向推测部件(203)；推测数字音压数据中的噪声水平的噪声推测部件(204)；根据噪声水平和数字音压数据，推测信号成分的比例的 SNR 推测部件(205)；根据信号成分的比例计算出输出信号并输出的功率计算部件(209)；对音源方向和输出信号进行统合，确定音源位置的统合部件(211)；根据画面变换 DB(213)的数据，将所确定的音源位置变换为显示装置的画面上的点的控制部件(212)。



1. 一种音响指示装置，检测要检测的声音的音源位置，将上述音源位置变换为显示装置的画面上的点，该音响指示装置的特征在于包括：

保持多个麦克风元件的麦克风阵列；

将通过上述麦克风阵列得到的模拟音压数据变换为数字音压数据的 A/D 变换部件；

根据上述麦克风元件之间的上述声音的相关性，从上述数字音压数据推测上述要检测的声音的音源方向的方向推测部件；

从上述数字音压数据推测噪声水平，根据上述噪声水平和上述数字音压数据，计算出上述声音的信号成分，并作为输出信号输出的输出信号计算部件；

将上述音源方向和上述输出信号统合起来，确定上述音源位置的统合部件；

将所确定的上述音源位置变换为上述显示装置的上述画面上的点的控制部件。

2. 根据权利要求 1 所述的音响指示装置，其特征在于：

上述麦克风阵列由多个子麦克风阵列构成，

上述音响指示装置还包括：

通过对每个上述子麦克风阵列根据三角测量将由上述方向推测部件推测出的上述音源方向统合起来，来计算出上述音源方向和距上述音源位置的三角测量部件；

判断上述音源方向和上述距离是否在预先定义的区域内的定位判断部件，其中

上述统合部件将上述输出信号与上述区域内的上述音源方向和上述距离统合起来，确定上述音源位置，

上述控制部件将所确定的上述音源位置变换为上述显示装置的上述画面上的点。

3. 根据权利要求1所述的音响指示装置，其特征在于：

上述麦克风阵列由多个子麦克风阵列构成，

上述音响指示装置还包括：

将上述数字音压数据变换为由时间-频带构成的信号的变换部件；

通过对每个上述子麦克风阵列根据三角测量将由上述方向推测部件使用上述信号推测出的上述音源方向统合起来，来计算出上述音源方向和距上述音源位置的距离的三角测量部件；

判断上述音源方向和上述距离是否在预先定义的区域内的定位判断部件，其中

上述统合部件将上述输出信号与上述区域内的上述音源方向和上述距离统合起来，确定上述音源位置，

上述控制部件将所确定的上述音源位置变换为上述显示装置的上述画面上的一点。

4. 根据权利要求1所述的音响指示装置，其特征在于：

上述麦克风阵列由多个子麦克风阵列构成，

上述音响指示装置还包括：

将上述数字音压数据变换为由时间-频带构成的信号的变换部件；

通过对每个上述子麦克风阵列根据三角测量将由上述方向推测部件使用上述信号推测出的上述音源方向统合起来，来计算出上述音源方向和距上述音源位置的距离的三角测量部件；

判断上述音源方向和上述距离是否在预先定义的区域内的定位判断部件；

判断由上述输出信号计算部件输出的上述输出信号在预先确定的阈值以上的输出信号判断部件；

预先存储了上述要检测的声音的频率特性的音源频率数据库；

存储能够根据上述音源位置确定上述画面上的上述一点的变换表的画面变换数据库，其中

上述统合部件根据上述频率特性对上述阈值以上的上述输出信号进行加权，而与上述区域内的上述音源方向和上述距离统合，确定上述音源位置，

上述控制部件使用上述画面变换数据库的信息，将所确定的上述音源位置变换为上述画面上的一点。

5. 一种音源位置的指示方法，通过处理装置检测要检测的声音的音源位置，将上述音源位置变换为显示装置的画面上的一点，该音源位置的指示方法的特征在于：

在上述处理装置中，

将通过保持多个麦克风元件的麦克风阵列得到的模拟音压数据变换为数字音压数据；

根据上述数字音压数据，进行基于上述麦克风元件之间的上述声音的相关性的上述音源方向的推测；

从上述数字音压数据中推测噪声水平，根据上述噪声水平和上述数字音压数据，计算出上述声音的信号成分，并作为输出信号输出；

通过将上述音源方向和上述输出信号统合起来，而将上述声音的上述音源位置变换为上述显示装置的上述画面上的点。

6. 根据权利要求5所述的音源位置的指示方法，其特征在于：

进而，在上述处理装置中，

上述麦克风阵列由多个子麦克风阵列构成，并且

通过对每个上述子麦克风阵列推测上述音源方向并根据三角测量进行统合，来计算出上述音源方向和距上述音源位置的距离，

通过将上述音源方向与上述输出信号统合起来，将上述声音的音源位置变换为上述显示装置的上述画面上的一点。

7. 根据权利要求5所述的音源位置的指示方法，其特征在于：

进而，在上述处理装置中，

上述麦克风阵列由多个子麦克风阵列构成，并且

抽出上述存储的数字音压数据并变换为由时间-频带构成的信号，

对每个上述子麦克风阵列使用上述信号推测上述音源方向并根据三角测量进行统合，来计算出上述音源方向和距上述音源位置的距离，

判断上述音源方向和上述距离是否在预先定义区域内，

将上述输出信号与上述区域内的上述音源方向和上述距离统合起来，确定上述音源位置，

将所确定的上述音源位置变换为上述显示装置的上述画面上的
一点。

8. 根据权利要求5所述的音源位置的指示方法，其特征在于：

进而，在上述处理装置中，

上述麦克风阵列由多个子麦克风阵列构成，并且

抽出上述存储的数字音压数据并变换为由时间-频带构成的信号，

通过对每个上述子麦克风阵列使用上述信号推测上述音源方向并根据三角测量进行统合，来计算出上述音源方向和距上述音源位置的距离，

判断上述音源方向和上述距离是否在预先定义区域内，

判断根据上述信号和上述信号的噪声水平计算出的输出信号的输出在预先确定的阈值以上，

将上述阈值以上的输出信号与上述区域内的上述音源方向和上述距离统合起来，确定上述音源位置并变换为上述画面上的
一点。

9. 一种计算机系统，至少具备：用于将要检测的声音的音源位置显示在画面上的显示装置；检测上述音源位置，将上述音源位置变换为上述显示装置的画面上的一点的音响指示装置；对使用与上述音响指示装置的上述音源位置有关的信息的程序进行处理的中央计算装置；存储上述程序的存储装置，该计算机系统的特征在于：

上述音响指示装置具备：

保持多个麦克风元件的麦克风阵列；

将通过上述麦克风阵列得到的模拟音压数据变换为数字音压数

据的 A/D 变换部件；

从上述数字音压数据，基于上述麦克风元件之间的上述声音的相关性，推测上述要检测的声音的音源方向的方向推测部件；

从上述数字音压数据中推测噪声水平，根据上述噪声水平和上述数字音压数据，计算出上述声音的信号成分，并作为输出信号输出的输出信号计算部件；

将上述音源方向和上述输出信号统合起来，确定上述音源位置的统合部件；

将所确定的上述音源位置变换为上述显示装置的上述画面上的一点的控制部件。

10. 根据权利要求 9 所述的计算机系统，其特征在于：

进而，在上述音响指示装置中，

上述麦克风阵列由多个子麦克风阵列构成，

将上述数字音压数据变换为由时间 - 频带构成的信号的变换部件；

通过对每个上述子麦克风阵列根据三角测量将由上述方向推测部件使用上述信号推测出的上述音源方向统合起来，来计算出上述音源方向和距上述音源位置的距离的三角测量部件；

判断上述音源方向和上述距离是否在预先定义的区域内的定位判断部件；

判断由上述输出信号计算部件输出的输出信号在预先确定的阈值以上的输出信号判断部件；

预先存储了上述要检测的声音的频率特性的音源频率数据库；

存储能够根据上述音源位置确定上述画面上的上述一点的变换表的画面变换数据库，其中

上述统合部件根据上述频率特性对上述阈值以上的上述输出信号进行加权，而与上述区域内的上述音源方向和上述距离统合，确定上述音源位置，

上述控制部件使用上述画面变换数据库的信息，将所确定的上述音源位置变换为上述画面上的一点。

音响指示设备、音源位置的指示方法和计算机系统

技术领域

本发明涉及用于由用户指定计算机中的显示装置的画面上的 1 点的指示设备 (pointing device)，特别涉及使用了音响信息的指示设备的技术。

背景技术

在计算机的操作中大多使用一般利用了鼠标的指示设备。这时，鼠标的操作与计算机中的显示装置的画面上的光标联合地动作，通过使光标移动到该点上，并在该点上进行点击 (click)，能够选择希望在画面上选择的点。

另外，对于使用了触摸屏的指示设备，也已经作为民用品被广为使用。在触摸屏中，安装了用于检测由用户在显示器上的各点上按压画面的压力的元件，对每个点判定是否进行了按压。

作为使用了音响信息的指示设备，存在以下这样的设备，即在按压了画面时，使用了发出超声波的特殊的笔的设备 (例如参照专利文献 1)。

另外，存在以下这样的装置，即与超声波一起还产生光，根据直到超声波和光到达收音/受光元件为止的时间差，检测出指示位置的装置 (例如参考专利文献 2)。

另外，在显示器上设置振动检测元件，在指尖触摸了显示器时，检测出所产生的振动的方向，根据它检测出点位置的装置 (例如参考专利文献 3)。

专利文献 1: 特开 2002 - 351605 号公报

专利文献 2: 特开 2002 - 132436 号公报

专利文献 3: 特开 2002 - 351614 号公报

但是，在计算机的操作中使用了鼠标的指示设备中，必须将鼠标放置在桌子上，使用并不方便。另外，触摸屏虽然没有附属设备，但需要特殊的显示器，另外必须将按下压力检测装置附加在显示器上的各元件上，需要靠近显示器地进行指示（pointing）。

另外，在专利文献1、2所记载的技术中，用户必须使用特殊笔、座标输入装置。另外，在专利文献3所记载的技术中，必须接触显示面而产生振动并检测出振动。

发明内容

鉴于以上问题点，本发明的课题就是：提供一种在计算机的操作时，完全不需要将附属设备放置在桌子上，即使在脱离以后，也能够使用声音的信息进行指示操作的音响指示设备、音源位置的指示方法和使用了上述音响指示设备的计算机系统。

为了解决上述课题，本发明的音响指示设备是一种检测出要检测的音源位置，将上述音源位置变换为显示装置的画面上的1点的音响指示设备，具备：保持多个麦克风元件的麦克风阵列；将通过上述麦克风阵列得到的模拟音压数据变换为数字音压数据的A/D变换部件；根据上述麦克风元件之间的上述声音的相关性，从上述数字音压数据推测上述要检测的声音的音源方向的方向推测部件；从上述数字音压数据推测噪声水平，根据上述噪声水平和上述数字音压数据，计算出上述声音的信号成分，并作为输出信号输出的输出信号计算部件；将上述音源方向和上述输出信号统合起来，确定上述音源位置的统合部件；将所确定的上述音源位置变换为上述显示装置的上述画面上的一点的控制部件。

进而，在本发明的音响指示装置中，上述麦克风阵列由多个子麦克风阵列构成，上述音响指示装置还包括：通过对每个上述子麦克风阵列根据三角测量将由上述方向推测部件推测出的上述音源方向统合起来，来计算出上述音源方向和距上述音源位置的距离的三角测量部件；判断上述音源方向和上述距离是否在预先定义的区域内的定位

判断部件，其中，上述统合部件将上述输出信号与上述区域内的上述音源方向和上述距离统合起来，确定上述音源位置，上述控制部件将所确定的上述音源位置变换为上述显示装置的上述画面上的一点。

进而，在本发明的音响指示装置中，上述麦克风阵列由多个子麦克风阵列构成，上述音响指示装置还包括：将上述数字音压数据变换为由时间-频带构成的信号的变换部件；通过对每个上述子麦克风阵列根据三角测量将由上述方向推测部件使用上述信号推测出的上述音源方向统合起来，来计算出上述音源方向和距上述音源位置的距离的三角测量部件；判断上述音源方向和上述距离是否在预先定义的区域内的定位判断部件，其中，上述统合部件将上述输出信号与上述区域内的上述音源方向和上述距离统合起来，确定上述音源位置，上述控制部件将所确定的上述音源位置变换为上述显示装置的上述画面上的一点。

进而，在本发明的音响指示装置中，上述麦克风阵列由多个子麦克风阵列构成，上述音响指示装置还包括：将上述数字音压数据变换为由时间-频带构成的信号的变换部件；通过对每个上述子麦克风阵列根据三角测量将由上述方向推测部件使用上述信号推测出的上述音源方向统合起来，来计算出上述音源方向和距上述音源位置的距离的三角测量部件；判断上述音源方向和上述距离是否在预先定义的区域内的定位判断部件；判断由上述输出信号计算部件输出的上述输出信号在预先确定的阈值以上的输出信号判断部件；预先存储了上述要检测的声音的频率特性的音源频率数据库；存储能够根据上述音源位置确定上述画面上的上述一点的变换表的画面变换数据库，其中，上述统合部件根据上述频率特性对上述阈值以上的上述输出信号进行加权，而与上述区域内的上述音源方向和上述距离统合，确定上述音源位置，上述控制部件使用上述画面变换数据库的信息，将所确定的上述音源位置变换为上述画面上的一点。

进而，在本发明中，提供一种在上述音响指示设备中使用的音源位置的指示方法、以及具备上述音响指示设备的计算机系统。

根据本发明，能够提供一种在计算机的操作时，完全不需要将附属设备放置在桌子上，即使在脱离以后，也能够使用声音的信息进行指示操作的音响指示设备。

进而，能够提供一种在上述音响指示设备中使用的音源位置的指示方法。

进而，能够提供一种使用了上述音响指示设备的计算机系统。

附图说明

图 1 是表示本发明的实施例的一个例子的音响指示设备的概要结构图。

图 2 是只使用时间域的信号的上图音响指示设备的概要结构图。

图 3 是上述音响指示设备和具备它的计算机系统的硬件结构图。

图 4 是表示在上述音响指示设备中使用的子麦克风阵列的直线配置的图。

图 5 是表示上述音响指示设备中的桌子上的用户敲击位置的设定例子的图。

图 6 是表示上述音响指示设备中的敲击位置的检测流程的图。

图 7 是表示上述音响指示设备中的判断和统合的处理流程的图。

图 8 是表示上述音响指示设备中的敲击声音的时间波形的图。

图 9 是上述音响指示设备中的每个时间 - 频率成分的格子的模式图。

图 10 是表示上述音响指示设备中的每个音源方向的功率 (power) 的图。

图 11 是表示上述音响指示设备中的在高度方向上设定敲击区域的例子的图。

图 12 是表示上述音响指示设备中的子麦克风阵列的配置的图。

图 13 是表示将上述音源指示设备应用于敲击音检测装置的应用例子的图。

图 14 是表示将上述音源指示设备应用于敲击音检测装置的其他

应用例子的图。

图 15 是表示将上述音源指示设备应用于敲击音检测装置的其他应用例子的图。

图 16 是表示将上述音源指示设备应用于敲击音检测装置的其他应用例子的图。

图 17 是表示将上述音源指示设备应用于敲击音检测装置的其他应用例子的图。

图 18 是表示将上述音源指示设备应用于敲击音检测装置的其他应用例子的图。

101: 麦克风阵列; 102: A/D 变换部件; 103: 中央计算装置;
104: 易失性存储器; 105: 存储介质; 106: 显示装置; 201: 缓存部件;
202: STFT 部件; 203: 方向推测部件; 204: 噪声推测部件; 205:
SNR 推测部件; 206: 三角测量部件; 207: 定位判定部件; 208: SNR
判断部件; 209: 功率计算部件; 210: 功率判断部件; 211: 统合部件;
212: 控制部件; 213: 画面变换 DB; 214: 音源频率 DB; 301:
桌子; 302: 显示器; 303、304、305: 子麦克风阵列

具体实施方式

以下, 参考附图, 详细说明本发明的实施例。

图 1 是表示本发明的实施例的一个例子的音响指示设备的概要结构图。音响指示设备例如是代替个人计算机(以下称为“PC”)的鼠标装置而使用的指示设备, 是通过由用户敲击桌子而能够指定在显示部件上表示的特定位置的指示设备。另外, 将通过敲击桌子的声音等而要检测为音响指示设备的音源的声音称为“突发性音”。图 1 所示的音响指示设备具备: 至少由 2 个以上的麦克风元件(以下也称为“麦克风”)构成的麦克风阵列 101; 将由麦克风阵列 101 得到的各麦克风元件的多信道突发性音的模拟音压数据变换为数字音压数据的 A/D (Analogue to Digital) 变换部件 102; 存储特定量的数字音压数据的缓存部件 201; 将数字音压数据变换为时间-频率域的信号 STFT

(Short Term Fourier Transform) 部件 202; 将麦克风阵列分割为多个子麦克风阵列 (以下也称为“子阵列”), 根据方位角和仰角, 推测出基于同一子麦克风阵列内的麦克风元件之间的声音的相关性计算出的突发性音的方向的方向推测部件 203; 对针对每个子麦克风阵列求出的音源方向进行统合, 测量音源的方位角、仰角和距离的三角测量部件 206; 判断由三角测量部件 206 求出的音源的位置是否在预定的范围内的定位判断部件 207; 从数字音压数据推测背景的噪声功率的噪声推测部件 204; 从数字音压数据和噪声功率推测 SNR(Signal to Noise Ratio) 的 SNR 推测部件 205; 输出 SNR 推测部件 205 所输出的 SNR 推测值为预定的阈值以上的 SNR 的 SNR 判断部件 208; 从数字音压数据和 SNR 计算信号功率的功率计算部件 209; 输出信号功率为预定的阈值以上的信号功率的功率判断部件 210; 将 SNR 判断部件和功率判断部件同时确定了的时间 - 频率成分作为定位判断部件预先定义了的区域内的音源位置座标而输出的统合部件 211; 将该音源位置座标变换为显示画面上的特定的点的控制部件 212。

进而, 还具备: 预先存储作为对象的声音的频率特性的音源频率数据库 (以下称为“DB”) 208; 使音源座标与显示画面上的特定的点对应的画面变换 DB213。

另外, 在将上述数字音压数据只用作由时间域构成的信号的情况下, 也可以省略 STFT 部件 202、功率判断部件 210、SNR 判断部件和音源频率 DB208 而确定音源位置。在图 2 中, 表示只使用时间域的信号在上述音响指示设备的概要结构图。图 2 表示了用于确定音源位置的最小限的结构。在此, 输出信号计算部件是指噪声推测部件 204、SNR 推测部件 205 和功率计算部件 209。进而, 为了更正确地确定音源位置, 还需要包含三角测量部件 206 和定位判断部件 207。

图 3 是上述音响指示设备和具备它的计算机系统的硬件结构图。图 3(a) 表示上述音响指示设备的硬件的结构图, 具备: 将上述模拟音压数据变换为数字音压数据的 A/D 变换部件 102; 进行与上述音响指示设备有关的处理的中央计算装置 103; 存储器 104; 存储与上述

音响指示设备相关的程序、麦克风阵列的各麦克风元件的物理座标的存储介质 105。在中央计算装置 103 上，一边使用易失性存储器 104，一边执行上述程序，由此来实现图 1 所示的上述音响指示设备中的除了麦克风阵列 101 和 A/D 变换部件 102 以外的各结构部件。

图 3 (b) 是具备上述音响指示设备的计算机系统的硬件结构图。上述计算机系统具备：音响指示设备 10；对使用与音响指示设备 10 的音源位置有关的信息的程序进行处理的中央计算装置 20；在上述程序和计算处理时使用的存储装置 30；用于将音源位置显示在画面上的显示装置。

接着，更详细地说明图 1 所示的各结构部件。

另外，还具备用于将音源位置显示为画面上的点的显示装置 106。

对每个信道，将特定量的由 A/D 变换部件 102 进行了变换的多信道的数字音压数据储存在缓存部件 201 中。时间 - 频率域中的处理并不是在每次得到 1 个采样就进行处理，而是在取得了多个采样后，统一地进行处理。直到储存了特定量的音压数据为止完全不进行处理，而是在储存了特定量的数字音压数据后，才开始进行处理。

缓存部件 201 具有储存该特定量的数字音压数据的功能。根据从 0 开始的索引 i ，对每个麦克风元件区别地对由各麦克风元件得到的数字音压数据进行处理。设 n 为整数，将开始数字变换后第 n 次采样的第 i 个麦克风元件的数字音压数据标记为 $x_i(n)$ 。

在 STFT (Short Term Fourier Transform) 部件 202 中，依照以下的公式 (1)，将每个麦克风元件的数字音压数据变换为时间 - 频率域的信号。

公式 (1)

$$X_i(f, \tau) = \sum_{n=0}^{N-1} w(n) x_i(s\tau + n) e^{-j\frac{2\pi f}{N}n}$$

在此，用公式 (2) 定义 j 。

公式 (2)

$$j = \sqrt{-1}$$

另外， $X_i(f, \tau)$ 为第 i 个元件的第 f 个频率成分。 f 从 0 开始，到 $N/2$ 结束。 N 为变换为时间 - 频率域的信号的数字音压数据的数据长度。通常称为帧大小。 S 通常被称为帧偏移 (frame shift)，是变换为时间 - 频率域的信号时的数字音压数据的错开量。缓存部件 201 针对每个麦克风元件，直到取得新的 S 采样为止，持续地储存数字音压数据，在取得了 S 采样后，由 STFT 部件 202 变换为时间 - 频率域的信号。

τ 被称为帧索引，相当于变换为时间 - 频率域的信号次数。 τ 从 0 开始。 $w(n)$ 被称为窗口函数，通常使用 blackman 窗口、hanning 窗口、hamming 窗口这样的函数。通过使用窗口函数，能够进行高精度的时间 - 频率分解。

变换为时间 - 频率域的信号的数字音压数据被发送到方向推测部件 203。

在方向推测部件 203 中，首先将形成麦克风阵列的麦克风元件分割为多个子麦克风阵列。然后，针对每个子麦克风阵列，在各个坐标系中推测音源方向。在分割为子麦克风阵列时，例如在分割为 R 个子麦克风阵列的情况下，形成麦克风阵列的 M 个麦克风元件被分配到 R 个子麦克风阵列的任意至少一个。也可以分配到 2 个以上的子麦克风阵列，在该情况下，多个子麦克风阵列具有同一麦克风元件。

图 4 是表示子麦克风阵列的图。图 4 (a) 是表示子麦克风阵列的直线配置的图。在直线配置的情况下，将与麦克风元件排列的阵列方向垂直的方向确定为角度 0 度，可以只推测将音源和子麦克风阵列连接起来的直线、从该方向旋转半时针与上述直线所成的角度 θ 。另外， d 表示麦克风间隔。图 4 (b) 是表示将上述 M 个麦克风元件分配到 R 个子麦克风阵列的状态的图，构成为向 1 个子麦克风阵列分配 3 个麦克风元件。

在将子麦克风阵列的 2 个麦克风元件平行地配置在桌子上面时，将上述角度 θ 推测为水平方向的方位角。另一方面，在将上述 2 个麦克风元件垂直地配置在桌子上面时，将上述角度 θ 推测为垂直方向的仰角。这样，推测方位角和仰角。

假设子麦克风阵列具有至少 2 个麦克风元件，在所具有的麦克风元件有 2 个的情况下，根据公式 (3) 推测 θ 。

公式 (3)

$$\theta(f, \tau) = \arcsin \frac{\rho(f, \tau)}{2\pi F d c^{-1}}$$

在此， ρ 为 2 个麦克风元件的输入信号的帧 τ 、频率索引 f 的相位差。 F 为频率索引 f 的频率 (Hz)。 F 为 $(f + 0.5) / N \times F_s / 2$ 。 F_s 是 A/D 变换部件 102 的采样速率。 d 为 2 个麦克风元件的物理间隔 (m)。 c 为音速 (m/s)。严谨地说音速依存于温度、介质的密度而变化，但通常固定为 340m/s 等的一个值而使用。

方向推测部件 203 的内部的处理由于是对每个时间 - 频率都一样的处理，所以以后省略时间 - 频率的索引 (f, τ) 而表示。方向推测部件 203 的处理对每个时间 - 频率进行同样的处理。在子麦克风阵列使用 3 个以上的麦克风元件，并且将它们配置在直线上的情况下，根据直线配置的 SPIRE 算法，能够高精度地计算出其方向。在 M.Togami, T.Sumiyoshi, and A.Amamo, "Stepwise phase difference restoration method for sound source localization using multiple microphone pairs", ICASSP2007, vol.I, pp.117 ~ 120, 2007 中，记载了 SPIRE 算法的详细。

在 SPIRE 算法中，使用相邻的麦克风元件的间隔 (以下成为“麦克风间隔”) 不同的多个麦克风对，因此理想的是将构成子麦克风阵列的麦克风元件配置为各个麦克风间隔不同。按照麦克风间隔从小到大的顺序，对各麦克风对进行排序。按照用于确定一个麦克风对的标记对 p 进行索引， $p = 1$ 为麦克风间隔最短的麦克风对， $p = P$ 为麦克风间隔最长的麦克风对。从 $p = 1$ 到 $p = P$ ，顺序地执行以下的处理。

首先，找到满足下式（4）的整数 n_p 。

公式（4）

$$\hat{\rho}_{p-1} \frac{d_p}{d_{p-1}} - \pi \leq \rho_p + 2\pi n_p \leq \hat{\rho}_{p-1} \frac{d_p}{d_{p-1}} + \pi$$

由于不等式所概括的范围相当于 2π ，所以一定只找到一个解。

然后，执行下式（5）。

公式（5）

$$\hat{\rho}_{p-1} = \rho_p + 2\pi n_p$$

另外，在对 $p = 1$ 进行上述处理之前，作为初始值，设定下述的公式（6）。

公式（6）

$$\hat{\rho}_0 = 0$$

另外， d_p 为第 p 个麦克风对的麦克风元件之间的间隔。在直到 $p = P$ 为止执行了上述处理之后，用公式（7）推测音源方向。

公式（7）

$$\theta(f, \tau) = \arcsin \frac{\hat{\rho}_p(f, \tau)}{2\pi F d_p c^{-1}}$$

对于音源方向推测的推测精度，已知麦克风间隔长度越长则越高，但已知如果麦克风间隔的长度为推测方向的信号的半波长以上，则无法从麦克风之间的相位差确定一个方向，而存在具有相同相位差的 2 个以上的方向（空间分区）。在 SPIRE 算法中，具备以下这样的机构：在按照长的麦克风间隔产生的 2 个以上推测方向中，选择与根据短麦克风间隔求出的音源方向接近的方向。因此，具有根据产生空间分区那样的长的麦克风间隔也能够高精度地推测音源方向的优点。在麦克风对是非直线配置的情况下，根据非直线配置用的 SPIRE 算法，根据方位角和情况，也能够对仰角进行计算。

另外，在上述数字音压数据不是时间 - 频率域的信号的情况下，即在只是时间域的情况下，无法使用上述 SPIRE 算法。在只是时间域的情况下，可以使用 GCC - PHAT (Generalized Cross Correlation PHase Transform) 法，来推测方向。

在噪声推测部件 204 中，根据 STFT 部件 202 的输出信号，推测背景的噪声水平。在推测噪声水平时，使用 MCRA (Minimum Controlled Recursive Averaging) 等。MCRA 的噪声推测处理以 minimum statistics 法为基础。minimum statistics 法针对每个频率，将数帧之间最小的功率作为噪声功率的推测值。一般，声音和桌子的敲击音对于每个频率大多突发地具有大的功率，保持长时间的大功率的情况少。因此，在数帧之间取最小功率的成分可以近似为只包含噪声的成分，即使在声音发出区间，也可以高精度地推测噪声功率。将推测出的每个麦克风元件/频率的噪声功率表示为 $N_i(f, \tau)$ 。 i 是麦克风元件的索引，对每个麦克风元件推测噪声功率。另外，噪声功率由于对每个帧都更新，所以为依存于 τ 的值。噪声推测部件 204 输出推测出的每个麦克风元件/频率的噪声功率 $N_i(f, \tau)$ 。

在只是时间域的情况下，噪声具有以下特征：与突发性音相比，功率的输出小，但持续时间长，因此能够推测噪声功率。

SNR 推测部件 205 根据推测出的噪声功率和麦克风阵列的输入信号 $X_i(f, \tau)$ ，按照以下的公式(8)推测 SNR(Signal To Noise Ratio)。

公式(8)

$$SNR_i(f, \tau) = 10 \log_{10} \frac{|X_i(f, \tau)|^2}{N_i(f, \tau)} - 1$$

$SNR_i(f, \tau)$ 是麦克风索引 i 的帧 τ 、频率索引 f 的 SNR。SNR 推测部件 205 输出推测出的 SNR。SNR 推测部件 205 也可以在时间方向上对输入功率进行平滑化等。通过进行平滑化，能够推测出对噪声具有强稳定性的 SNR。

三角测量部件 206 将对每个子麦克风阵列求出的音源方向进行统合，测量方位角、仰角以及到音源位置的距离。针对对各子麦克风

阵列在各个坐标系中求出的音源方向，用以下的公式(9)表示对第*i*个子麦克风阵列求出的音源方向。

公式(9)

$$\theta_i(f, \tau)$$

例如，如图4所示那样，将与阵列方向垂直的方向确定为角度0度，将从与阵列方向平行的方向旋转半时针定义为音源方向。在此，一般音源方向由方位角和仰角的2个要素构成，但在子麦克风阵列是直线配置的情况等下，在只能推测出任意一方的情况下，也可以只用该一方要素构成。在该情况下，将在要素数为1的第*i*个子麦克风阵列的座标系统中求出的音源方向变换为绝对坐标系中的音源方向。设变换后的绝对坐标系中的音源方向为 P_i 。根据第*i*个子麦克风阵列的结果，可以将音源推测为存在于音源方向 P_i 上。由此，可以认为推测为在全部子麦克风阵列中求出的音源方向 P_i 的交点是音源位置是妥当的。由此，在三角测量部件206中，将音源方向 P_i 的交点输出为音源位置。

一般，有无法将音源方向 P_i 的交点确定为一个的情况。在这样的情况下，对于全部子麦克风阵列的组，求出2个音源方向的交点，将这些交点的平均值输出为音源位置。通过平均化，对交点位置的离散增加抵抗性。

根据情况，有2个音源方向不具有交点的情况。在该情况下，在相应的时间-频率域的音源位置推测时不使用在不具有该交点的子麦克风阵列的组中求出的解，或者在相应的时间-频率域中，不进行音源位置推测。不具有交点时可以考虑为存在观测对象的音源以外的音源等，而相位差的信息包含噪声的情况。因此，通过不使用在这样的时间-频率域中推测出的音源位置，能够更高精度地推测音源位置。

另外，在子麦克风阵列是直线配置的情况下，无法推测方位角和仰角的双方，只能够推测子麦克风阵列的阵列方向与音源所成的角。在该情况下，音源存在于子麦克风阵列的阵列方向与音源所成的角成

为推测出的值那样的平面上。将在各子麦克风阵列中求出的这样的平面的交点输出为音源位置或音源方向。在全部的子麦克风阵列是直线配置的情况下，将对全部子麦克风阵列的组合求出的平面的交点的平均值输出为音源位置。通过平均化，对许多交点位置的离散增加抵抗性。

另外，在一部分子麦克风阵列是直线配置，其他子麦克风阵列是非直线配置的情况下，通过将直线配置的子麦克风阵列的1个和非直线配置的子麦克风阵列的1个组合起来，能够得到一个音源位置的推测值。在组合直线配置和非直线配置而使用的情况下，将交点被确定为一个那样的子麦克风阵列的最小数作为1个单位，将按照全部子麦克风阵列的组合而求出的交点的平均值作为最终的音源位置的推测值输出。

定位判断部件207判断由三角测量部件206求出的音源位置是否在桌子上，或者该音源位置是否在预定的敲击区域内。在同时满足根据由三角测量部件206求出的音源位置的信息计算的音源相对于桌子的高度的绝对值是否为预定的阈值以下、以及根据音源位置的信息计算的音源在桌子上的平面座标是否位于敲击区域内的2个观点的情况下，从定位判断部件207输出音源方向和相对于音源的距离作为音源位置的信息。另外，也可以将音源方向和相对于音源的距离作为方位角和仰角输出。另外，在定位判断部件中，也可以在同时满足上述2个观点的情况下，输出正的判断结果，在不同步满足的情况下，输出负的判断结果，在后述的统合部件211中，将正的判断结果与从三角测量部件输出的音源方向和距离统合。将在后面说明敲击区域的定义。

SNR判断部件208输出由SNR推测部件205输出的每个时间-频率的SNR的推测值为预定的阈值以上的时间-频率成分。功率计算部件209根据由SNR推测部件205输出的每个时间-频率的SNR，用以下的公式(10)计算信号功率 P_s 。

公式 (10)

$$P_s = \frac{SNR}{SNR + 1} P_x$$

在此， P_x 为输入信号的功率。

在功率判断部件 210 中，输出由功率计算部件 209 输出的每个时间 - 频率的信号功率为预定的阈值以上的时间 - 频率成分。在统合部件 211 中，针对由功率判断部件 210 和 SNR 判断部件 208 同时确定的时间 - 频率成分，用保存在音源频率 DB208 中的每个频率的加权对由相应成分的功率计算部件 209 输出的功率进行加权。即，在能够预先测量敲击桌子的声音等作为对象的声音的频率特性的情况下，将该频率特性存储在音源频率 DB208 中。然后，通过用存储在音源频率 DB208 中的功率进行加权，而能够更高精度地进行位置推测。

对于没有由功率判断部件 210 和 SNR 判断部件 208 同时确定的时间 - 频率成分，将加权设为 0。另外，对于由定位判断部件 207 判断为没有在敲击区域内的时间 - 频率成分，也将加权设为 0。

另外，在本实施例中，输出信号判断部件被称为 SNR 判断部件 208 和功率判断部件 210。

将敲击区域切割为一边为数 cm 左右的格子状，在对于每个时间 - 频率，在第 i 个格子内包含相应成分的音源位置的推测结果的情况下，将附加了相应的加权的功率加到该格子的功率 P_i 上。对全部的每个时间 - 频率，这样进行格子的功率的加法处理。将进行了加法处理后的功率为最大的格子作为最终的音源位置而输出。预先定义格子的大小和数目。

另外，也可以构成为也对进行格子的功率的加法处理的时间长度预先进行定义，或者使用 VAD (Voice Activity Detection)，只对判断为声音区间的时间段进行上述加法处理。通过缩短进行加法处理的时间，在响起敲击音后，能够进一步缩短直到判断该音源位置为止的反应时间。但是，如果进行缩短，则有对噪声抗力变弱的缺点。

另外，通过延长进行加法处理的时间，具有以下优点：在产生了敲击音后，直到判断该音源位置为止的反应时间变长，但对噪声的抗力强。需要考虑到这样的折中的关系，来决定进行加法处理的时间，但一般敲击音在 100ms 左右的短时间内发音停止，因此理想的是也将进行加法处理的时间设定为该程度的时间。另外，在格子的最大功率比预定的阈值小的情况下，作为没有敲击音而将结果废弃。另一方面，在格子的最大功率大于预定的阈值的情况下，输出该音源位置，结束统合部件 211 的处理。

在控制部件 212 中，根据画面变换 DB213 的信息，将由统合部件 211 输出的敲击音的音源位置的座标变换为画面上的特定的点。

画面变换 DB213 将音源位置的座标作为输入，保存用于对画面上的特定的点进行变换的表。在变换时，也可以是 2 行 2 列的矩阵的线性变换等，能够根据敲击音的音源位置确定画面上 1 个点的变换的任意一个。例如无视在音源位置推测时得到的与音源的高度有关的信息，对 PC 进行控制，使得将音源的水平面上的位置信息和画面上的 1 个点对应起来，恰好用鼠标点击变换后的画面上的 1 点、或者拖动。另外，也可以利用高度信息，例如在从某一定以上的高度发出了声音的情况下，看作是双击了画面上的 1 点，另外在从某一定以下的高度发出了声音的情况下，看作是单击了画面上的 1 点等，与高度信息对应地改变解释。这样，能够进行更多样的用户操作。

图 5 是表示桌子上的用户的敲击位置的设定例子的图。在被敲击对象的桌子 301 上，事前指定桌子上的某平面作为敲击区域。在推测出的敲击音的音源位置在该敲击区域内的情况下，接受该声音。麦克风阵列可以如子麦克风阵列 303 ~ 305 那样设置在显示器 302 上，也可以另外设置在桌子上。在此，子麦克风阵列 303 推测仰角，子麦克风阵列 304、305 推测方位角。通过将子麦克风阵列设置在显示器上，能够使麦克风阵列的座标轴的中心与显示器的中心对齐，能够更直观地指定显示器的假想空间上的一点。

图 6 是表示利用上述桌子的敲击位置检测来判别用户按下了的

画面上的按键的装置的处理流程的图。

在系统启动后，在结束判断 501 中，判断计算机关机 (shutdown) 了的情况，或判断是否通过用户按下桌子的敲击位置检测程序的结束按键等的任意方法，应该结束程序。

在结束判断 501 中判断为结束的情况下，进行结束，而处理结束。在没有判断为结束的情况下，处理转移到数字变换 502，将通过麦克风阵列取得的模拟音压数据变换为数字音压数据。由 A/D 变换部件进行变换。将变换后的数字化了的数字音压数据取入到计算机上。数字变换可以对每 1 个采样进行，也可以与桌子敲击音的最小处理长度一致等地，同时将多个采样取入到计算机。在时间 - 频率变换 503 中，利用短时间傅立叶变换，将取得的数字数据分解为每个时间 - 频率的成分。通过使用短时间傅立叶变换，能够对每个频率成分推测声音的到来方向。

在使用桌子敲击音程序的环境中，除了桌子敲击音以外，作为噪声还存在人的声音的情况很多。已知人的声音在时间 - 频率域中是稀疏的信号，成分偏于一部分频带中。因此，通过这样在时间 - 频率域中推测音源方向，能够容易地废弃人声所偏在的频率成分，提高敲击音检测精度。

在检测结果废弃判断 505 中，判断检测出的敲击音是否确实是桌子上的敲击区域内的敲击音。在判断为不是敲击音的情况下，处理转移到结束判断 501。在判断为是敲击音的情况下，预先定义敲击区域的各点与画面上的 1 点的映射，通过判别按下了敲击位置的按下位置判别 506，依照该映射。根据敲击位置的信息，确定画面上的 1 点。通过按键存在判断 507 来判断在敲击区域的位置是否存在按键，在判断为不存在的情况下，处理返回到结束判断 501。在判断为存在按键的情况下，作为按键动作 508，执行与在画面上使用鼠标等其他指示设备点击该按键的情况一样的处理。

图 7 表示定位判断部件、功率判断部件、SNR 判断部件和统合部件的具体处理流程。在定位判断部件 207 中，对每个时间 - 频率成

分，根据使用多个子麦克风阵列由三角测量部件计算出的音源方向和距离，即方位角和仰角的信息，判断上述方位角和仰角是否在预先定义的敲击区域内（定位判断 601）。预先定义的敲击区域可以如图 5 记载的敲击区域那样是桌子上的长方形的区域，也可以在空间上具有厚度。也可以是能够根据仰角和方位角的信息判断是否在敲击区域中的空间。

功率判断部件 210 判断敲击音的大小是否比利用上述 MCRA 法那样的方式推测出的噪声功率大（噪声功率比较 602）。MCRA 法是从声音和背景噪声混杂的声音中，推测背景噪声的功率的方式。MCRA 法是基于最小统计量的方法。最小统计量是指在将声音假设为突发地具有大音量的前提下，将在过去数帧内最小的功率看作为背景噪声的功率的方法。但是，根据最小统计量推测出的背景噪声功率具有比实际的背景噪声的功率小的倾向。MCRA 法是对根据最小统计量推测出的背景噪声功率在时间方向上进行平滑化等而进行修正，计算出接近实际的背景噪声的功率的值的的方法。敲击音不是声音，但在突发地取大功率的观点上，显示出与声音一样的统计上的性质，因此能够适用 MCRA 法等背景噪声功率的推测法。

在敲击音的功率比噪声功率大的情况下，接着，计算背景噪声与敲击音的功率的 SNR。SNR 判断部件 208 判断敲击音的功率是否比计算出的 SNR 大（SNR 判断 603），在大的情况下，将该时间 - 频率成分判断为敲击音成分。

在统合部件 211 中，预先将敲击区域分割为格子状。根据该成分的方位角和仰角的推测值，将被判断为敲击音成分的时间 - 频率成分分配到相应的格子内。在分配时，在相应的格子中，将施加了依存于频率的加权的值与该敲击音成分的功率相加。只在预先定义的频带和预先定义的时间长度内，进行这样的处理。然后，检测出功率最大的格子（格子检测 604），将该格子的方位角和仰角作为敲击音的方位角和仰角输出，确定音源位置。在此，在功率最大的格子的功率小于预先定义的阈值的情况下，判断为不存在敲击音。

另外，定位判断部件 207、功率判断部件 210 和 SNR 判断部件 208 各自的处理顺序并不限于图 7 所示的顺序。也可以在统合部件 211 的处理之前，结束定位判断部件 207、功率判断部件 210 和 SNR 判断部件 208 各自的处理。

图 8 是表示典型的敲击音的时间波形的图。敲击音突发地取大的值（敲击音的直接音）。然后，敲击音的回音、混响成分到来。该回音、混响成分可以看作是从各种方向到来的声音。因此，难以将回音、混响成分与直接音比较而推测方向，因此在敲击音的方向推测中使用并不理想。回音、混响成分一般比直接音的功率小，因此可以用对于突发的大声音之后的功率比该突发的大声音小的成分不看作是敲击音的方法进行判别。根据这样的观点，在由频率判断部件将每个时间-频率的敲击音成分分配到各格子中时，也可以与前一帧相比添加以下这样的处理，即不将功率小的成分分配到格子中。通过添加，能够对回音、混响抗力强地检测敲击音。

图 9 是与将每个时间-频率成分分配到格子的情况有关的模式图。设想了敲击音检测装置被作为鼠标等 PC 操作设备的代替品而使用的情况。为此，设想了在使用敲击音检测装置的环境中，存在人的说话声音等许多声音音源的情况。因此，需要一种即使在存在声音音源的环境中，也强有力地动作的敲击音检测装置。一般，可以认为声音在时间-频率域中是稀疏的信号。即，声音在时间-频率域中，其功率偏在于一部分成分中。因此，通过除去该偏在的一部分成分，即使在存在声音音源的环境中，也能够使敲击音检测装置强有力地动作。

在统合部件 211 中，判断仰角和方位角是否在敲击区域内，只在位于敲击区域内的情况下，看作是敲击音。通过进行这样的判断，能够废弃声音成分偏在的一部分时间-频率域。

也可以在统合部件 211 中，进行动作使得输出功率最大的格子，但在对各个子麦克风阵列求出了功率最大的方向后，对该最大方向进行统合，通过三角测量而推测敲击音的音源方向。

图 10 是表示子麦克风阵列中的每个方向的频度的例子的图。例如，如图 10 所示那样，对每个子麦克风阵列，将从各个子麦克风阵列看到的每个方向的功率相加。在将时间 - 频率成分分配到 2 维平面或 3 维空间上这样的方式下，每个格子的被分配数极少的情况也很多。在这样的情况下，对每个子麦克风阵列分别计算直方图，在分别计算出产生各个直方图的最大值的方向后，通过三角测量统合，能够进行强有力的推测。

图 11 是使敲击区域在高度方向上具有厚度而设定的例子。如该例子那样，通过使敲击区域在高度方向上具有厚度，除了对一些仰角方向的推测误差成为强有力以外，还能够检测出手指在空中发音等产生的声音。

图 12 是表示子麦克风阵列的配置的一个例子的图。是配置多个子麦克风阵列 1101 ~ 1104 使得围住敲击区域的例子。如图 12 所示那样，通过配置为围住敲击区域，与图 5、图 11 所示的子麦克风阵列 303 ~ 305 的配置相比，能够更高精度地检测出敲击音位置。

图 13 是将上述音源指示设备应用于敲击音检测装置中的应用例子的图。将显示器 1204 配置在桌子上，使得显示面与桌子面平行。敲击音区域为显示器画面全体。通过这样的设定，在用户敲击了显示器上的显示面的一点的情况下，能够知道该敲击点在哪里。即，可以使用图 13 所示那样的敲击检测装置代替触摸屏。另外，在触摸屏中，如其名字所示那样，只能够检测出“是否触摸了”，但如果使用本发明的敲击音检测装置，则通过使敲击区域在高度方向上具有厚度，在空间上用手指发声的情况下，也能够进行检测。

图 14 是将敲击音检测装置应用于棒球的“击打判断装置”的应用例子。在所谓的击打判断装置中，从投球区域 1301 向图 14 所示那样的靶子 1305 投球。然后，判断判断击中了靶子 1305 上的从 1 到 9 的块中的哪个块。球击中了靶子时产生的音是突发地具有大功率的突发性音，因此能够由本发明的敲击音检测装置检测。在该靶子上配置图 14 所示那样的多个子麦克风阵列 1302 ~ 1304，并适用敲击音检测装

置，由此在球击中靶子时，能够判断击中了 1~9 的哪个块、或者是否击中了框。由于球击中了框时的金属音和球击中了块时的音的频率特性不同，所以通过参照被判断为敲击音的频率成分的特性，能够区别球击中了框的情况和球击中了块的情况。

图 15 是将敲击音检测装置应用于足球的“射门位置判断装置”的应用例子。作为结构，与图 14 的击打判断装置是相同的结构。由使用了子麦克风阵列 1402~1404 的敲击音检测装置判断从踢球区域 1401 踢出的球击中了靶子 1405 内的 1 到 9 的块中的哪个块。

图 16 是将敲击音检测装置应用于乒乓球的“弹落位置判断装置”的应用例子。能够知道乒乓球的球弹落了的位置。作为结构，与击打判断装置和射门位置判断装置一样。由使用了子麦克风阵列 1502~1507 的敲击音检测装置判断球弹落在球台 1501 上的位置。由于乒乓球的球弹落在球台 1501 上时的声音是突发的音，所以能够由本敲击音检测装置进行判断。由此，在乒乓球的实况转播等中，能够得到视听者以前无法得到的球的轨迹相关的信息。

图 17 是将敲击音检测装置应用于“网球的壁打”，检测出击中墙壁的球的位置的应用例子。在经常用网球进行的壁打练习中，以前没有能够知道击中了墙壁的哪里的装置，因此无法判断击中的方向是好还是坏。通过使用了配置在墙壁 1601 上的多个子麦克风阵列 1602~1606 的敲击音检测装置，能够检测出球的击中位置。例如，通过预先存储球的击中位置，以后显示在计算机的显示器上，能够看到击中位置的离散是否大等这样的结果。

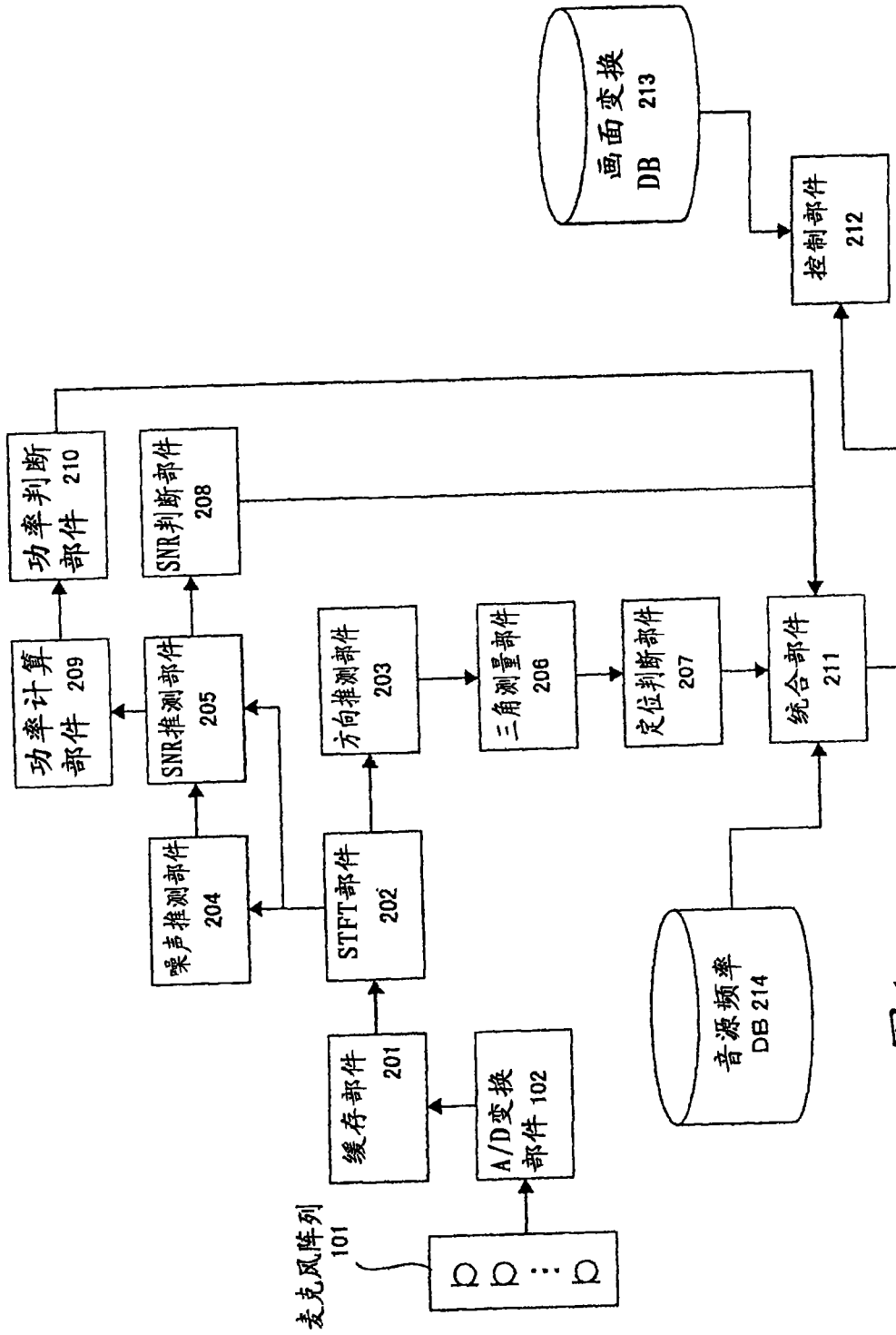


图1

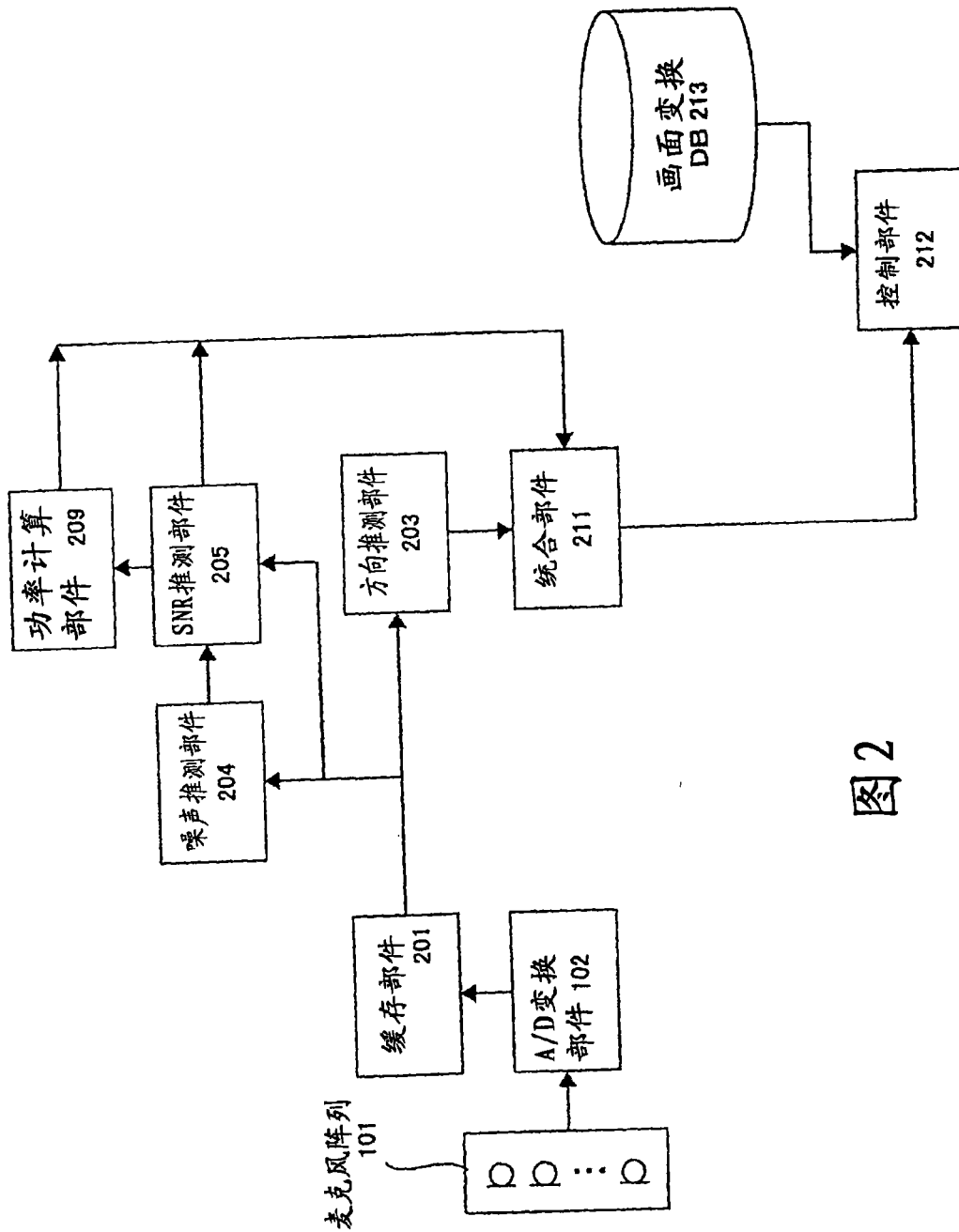


图2

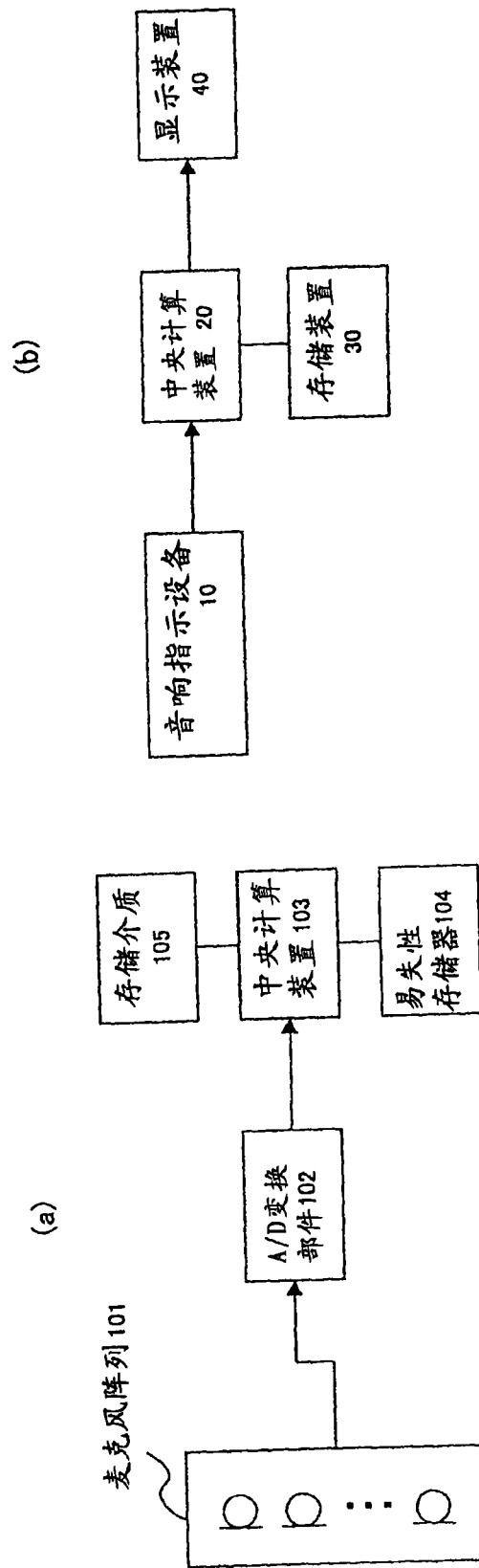
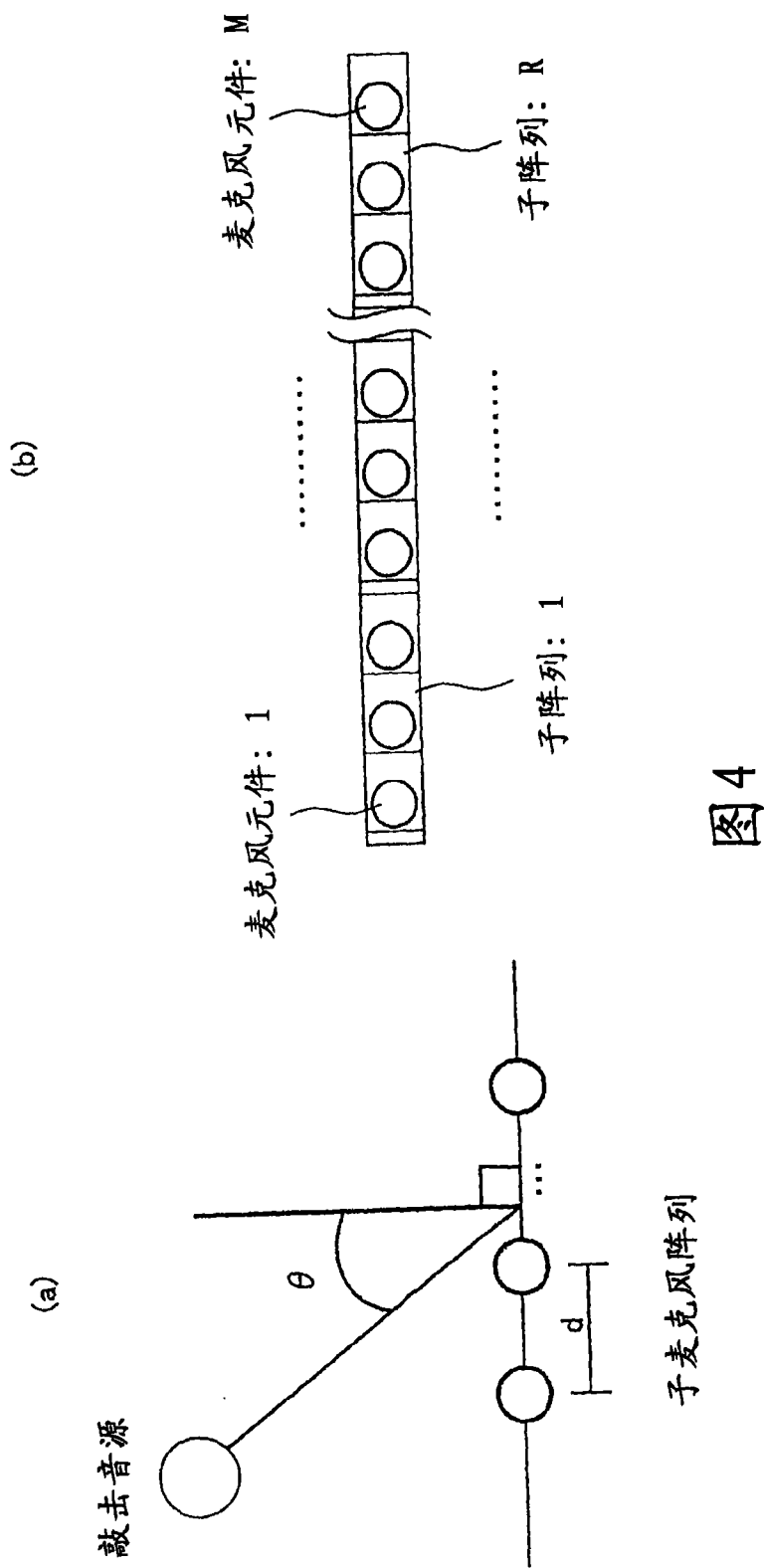


图 3



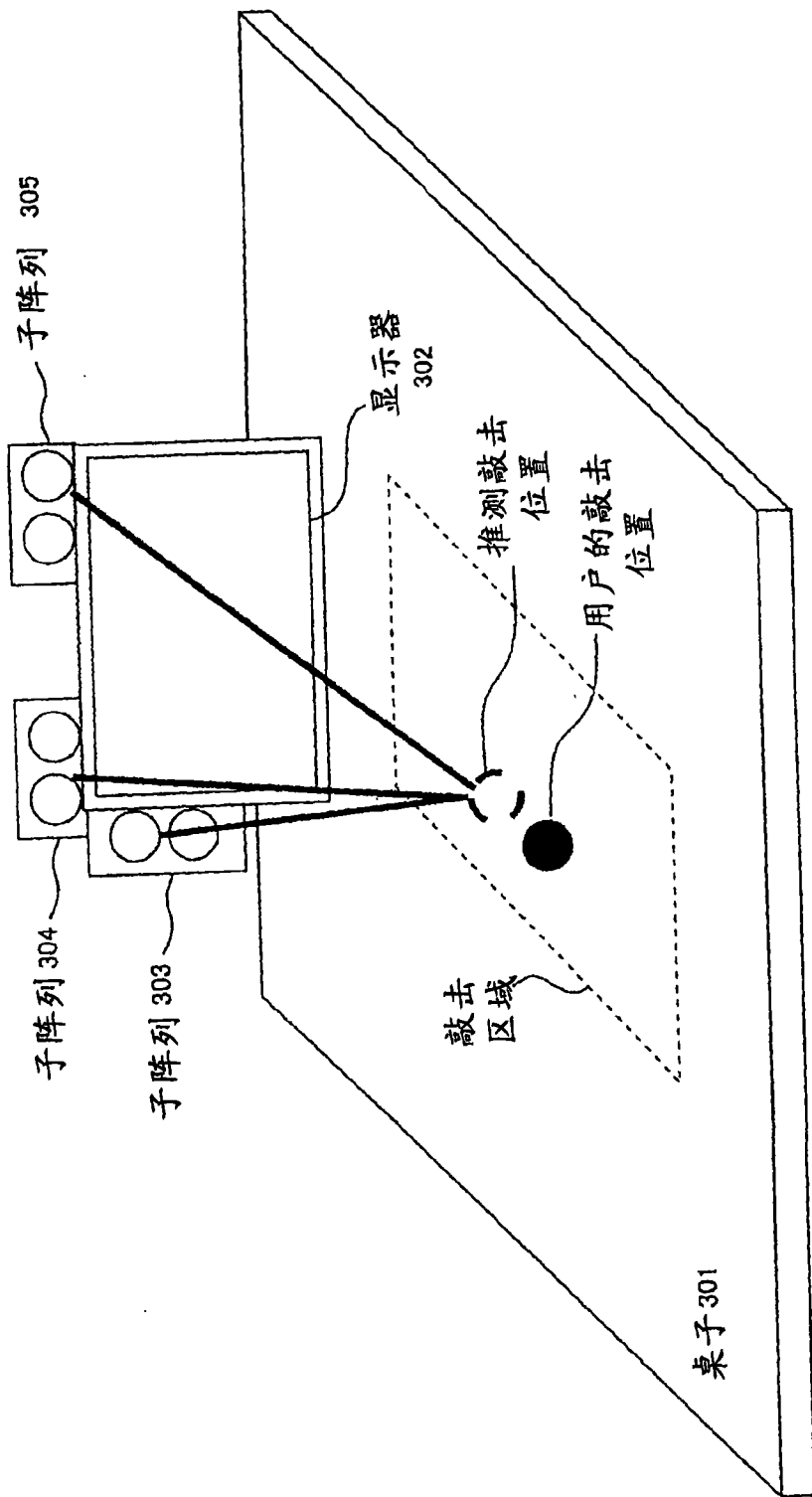


图5

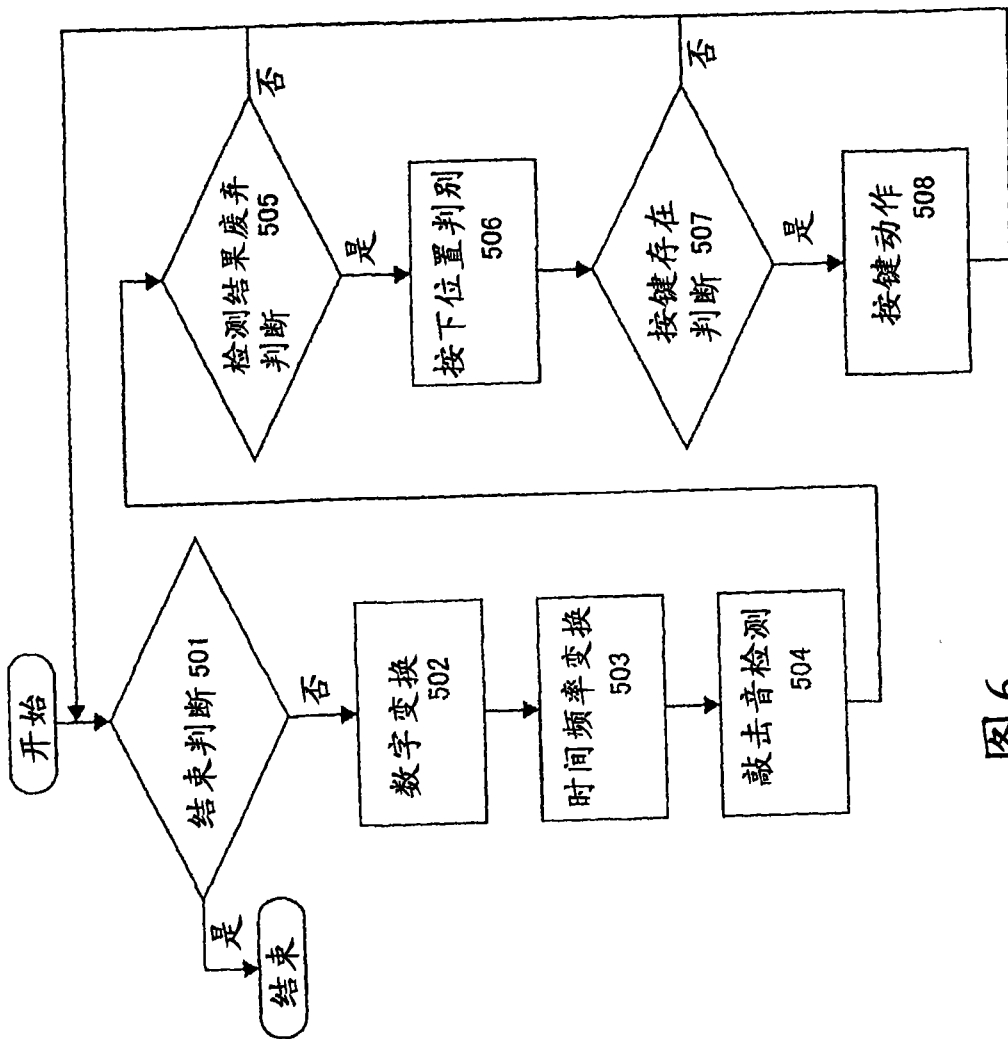


图6

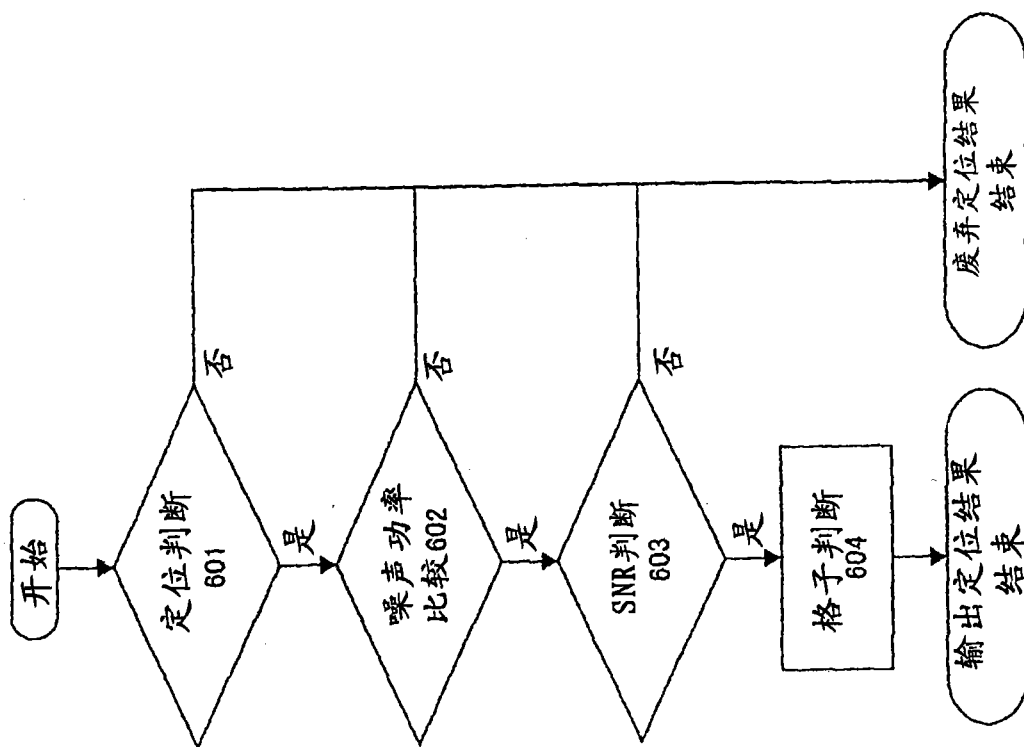


图7

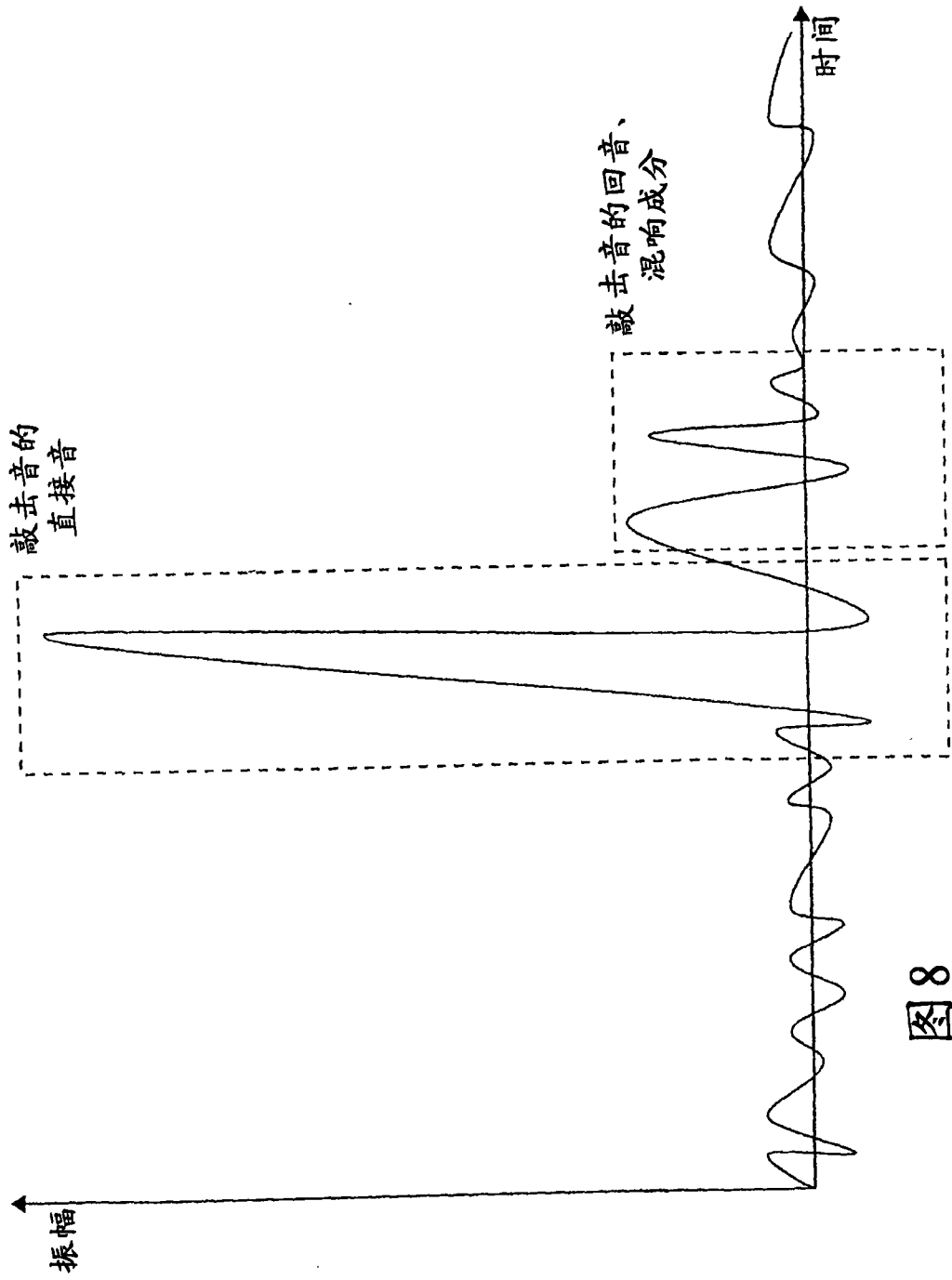


图8

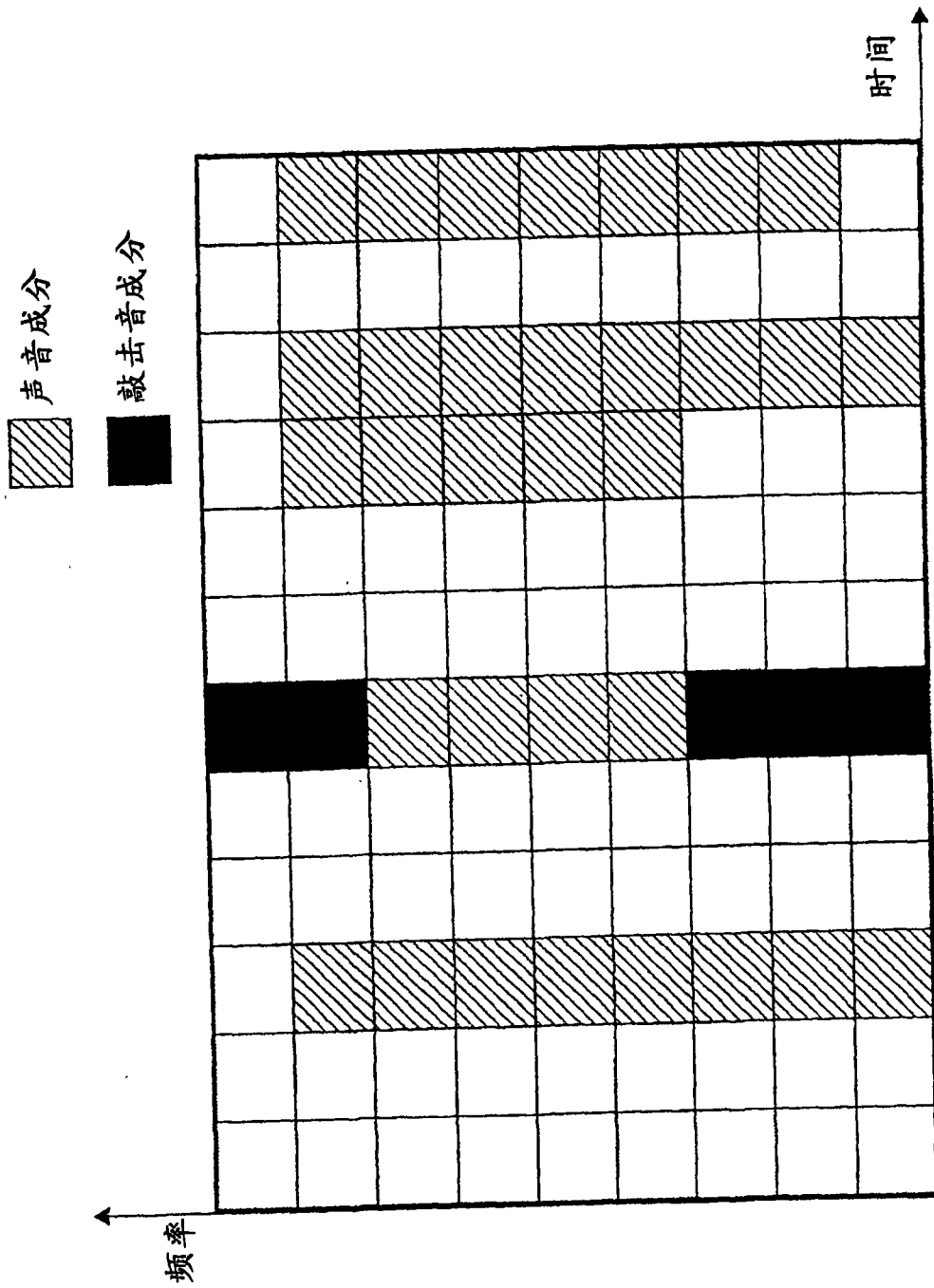


图9

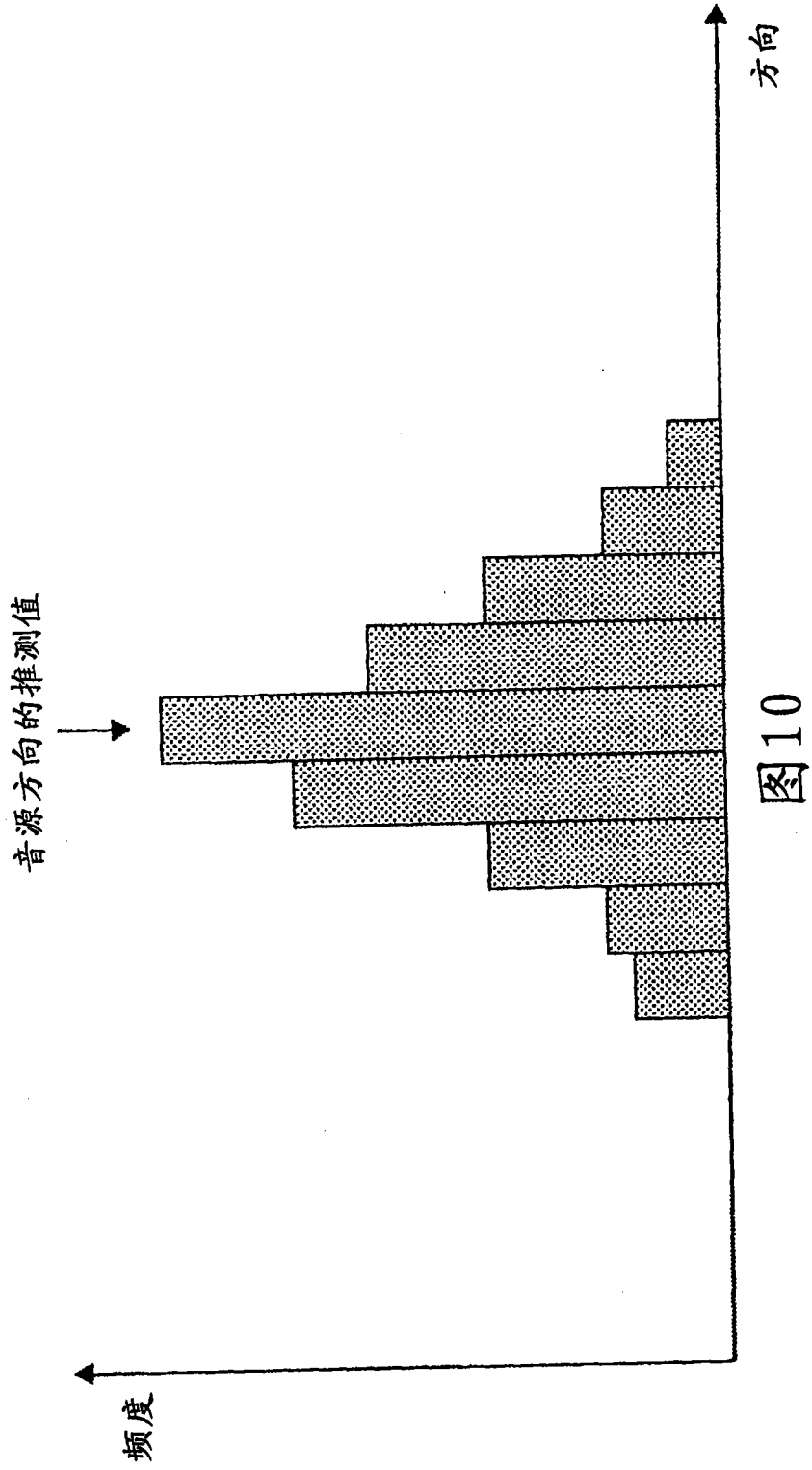


图10

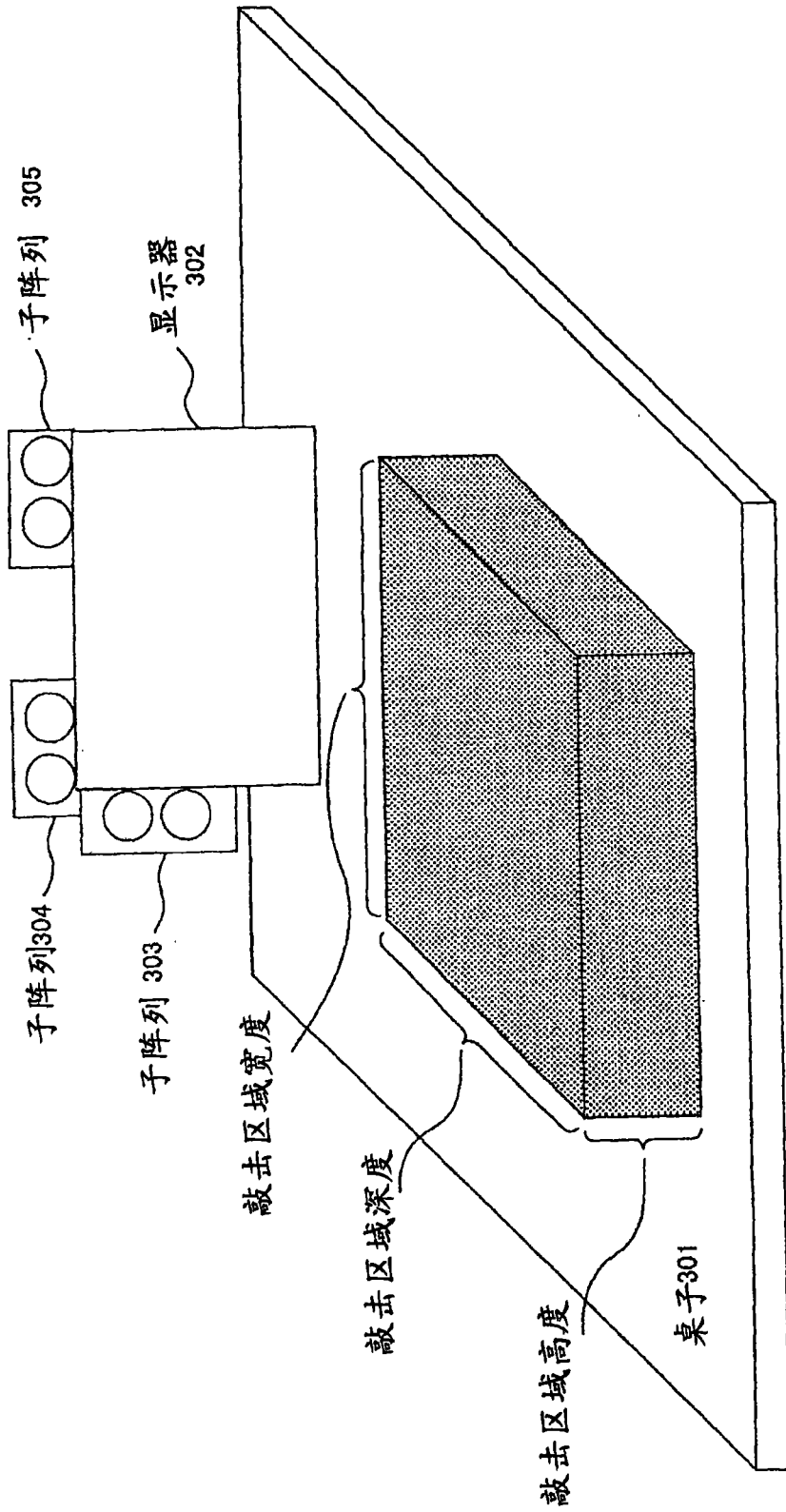


图11

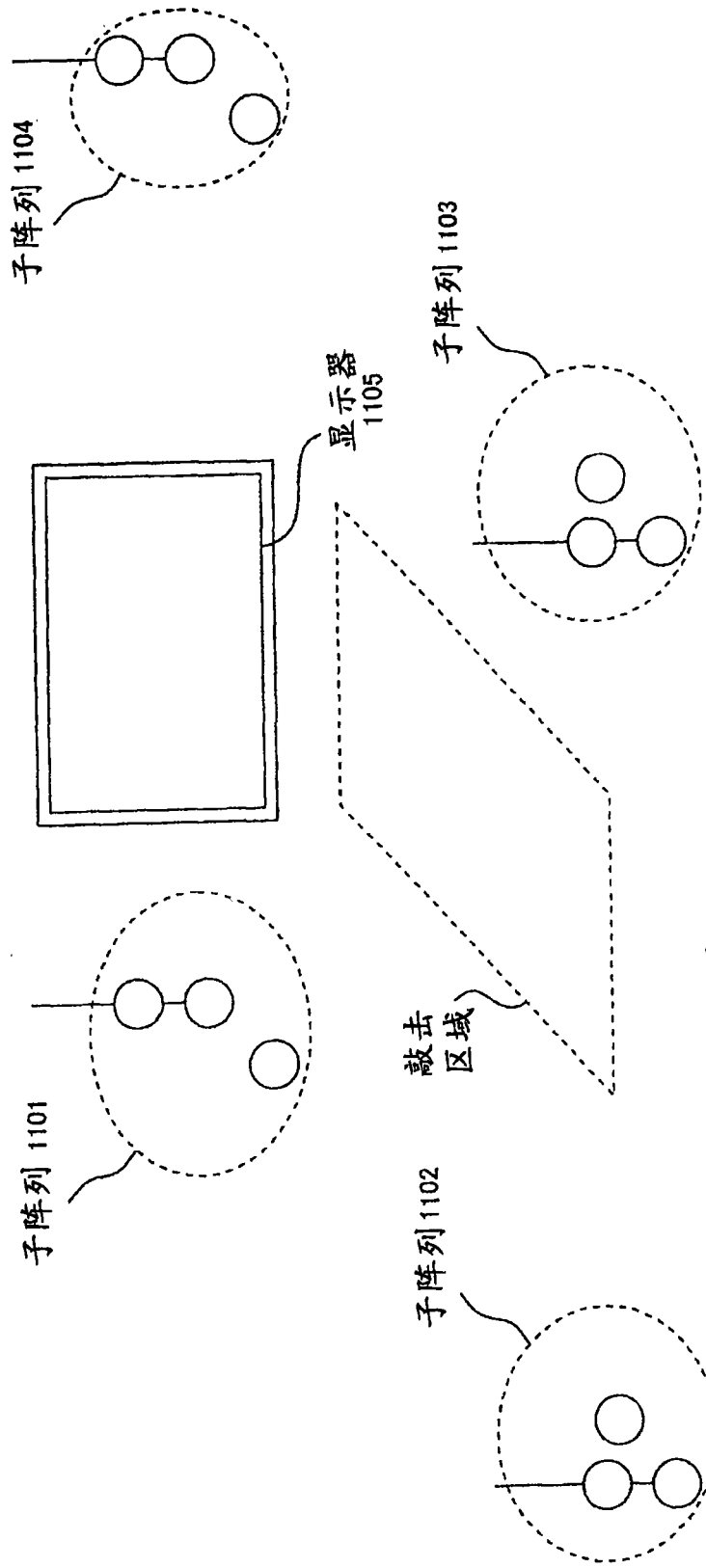


图12

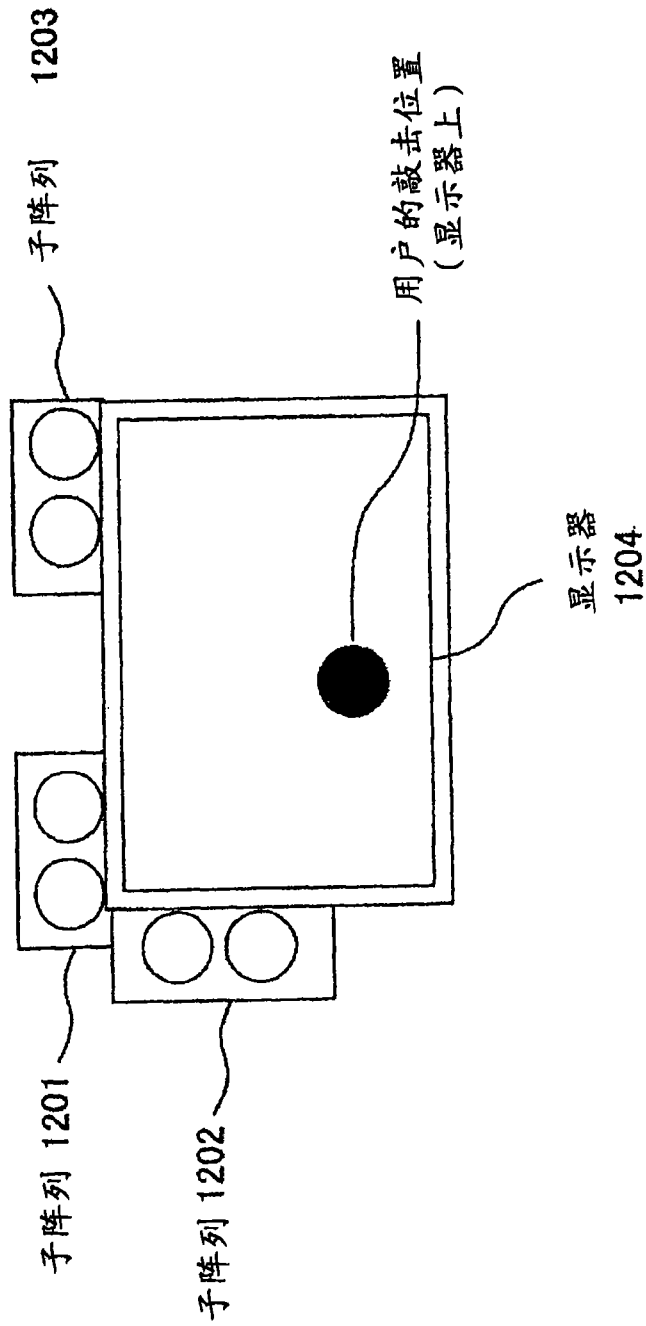


图13

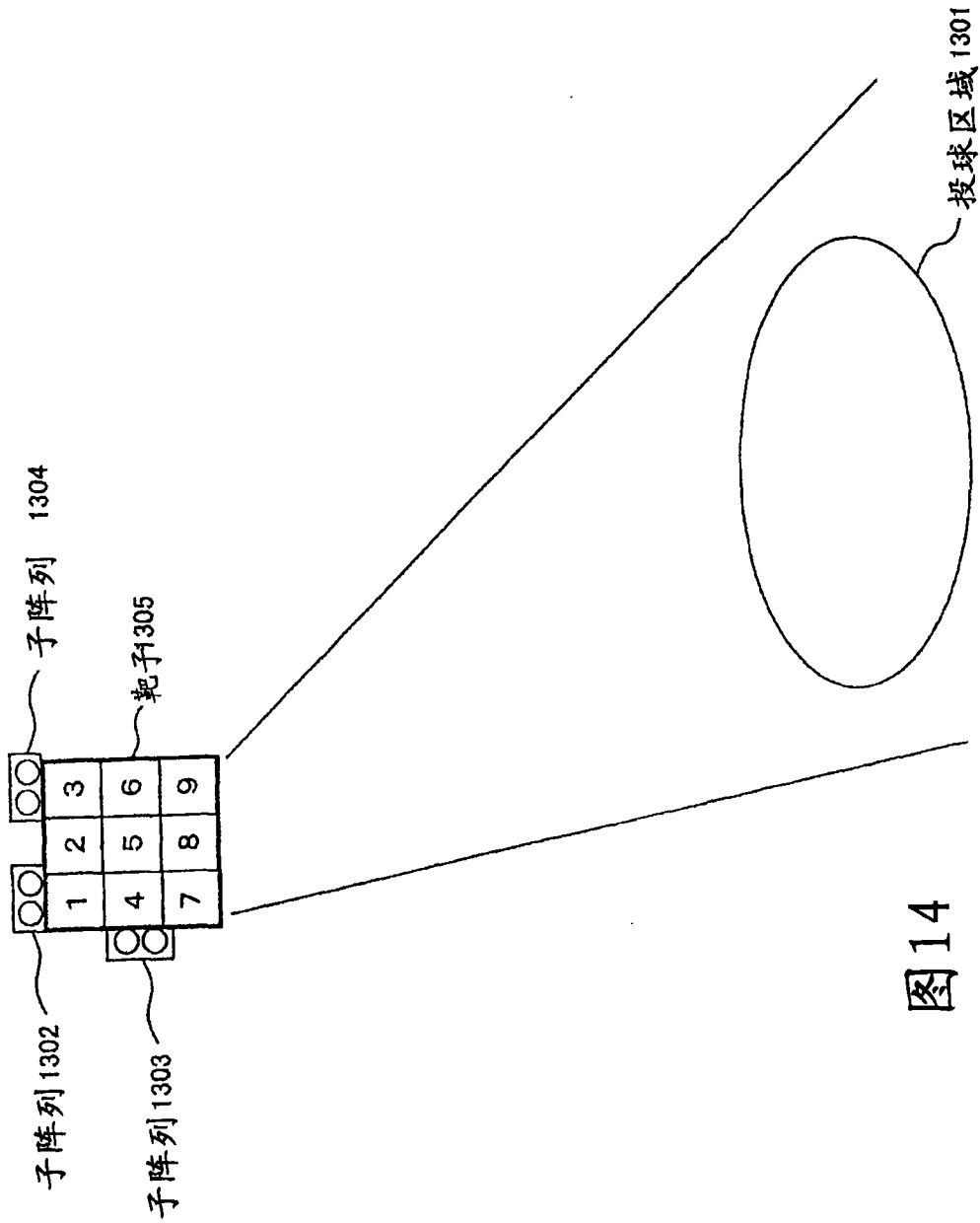


图14

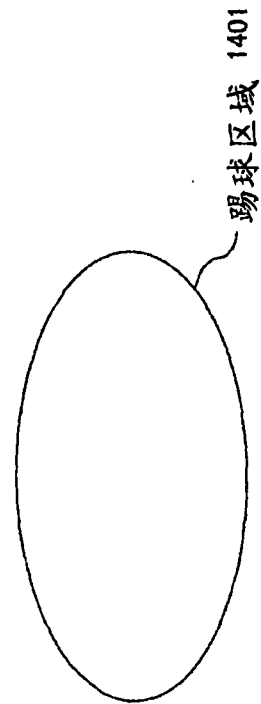
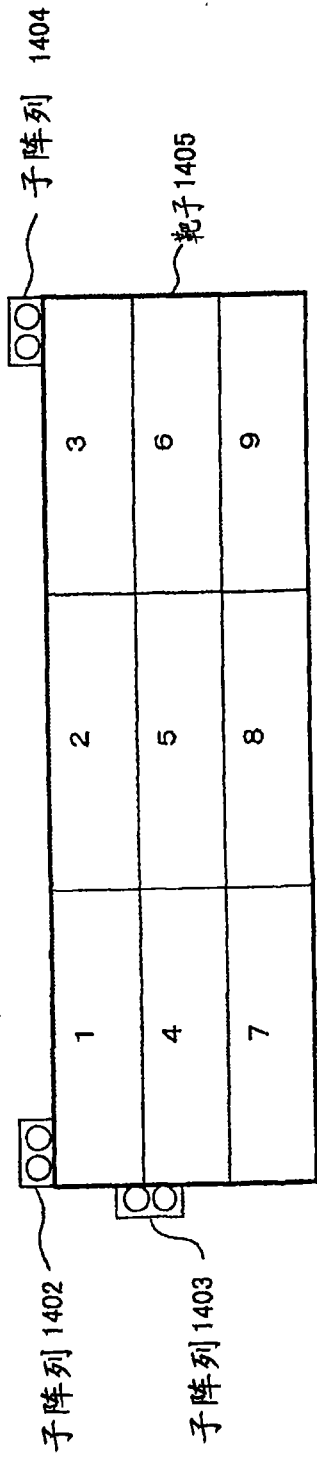


图15

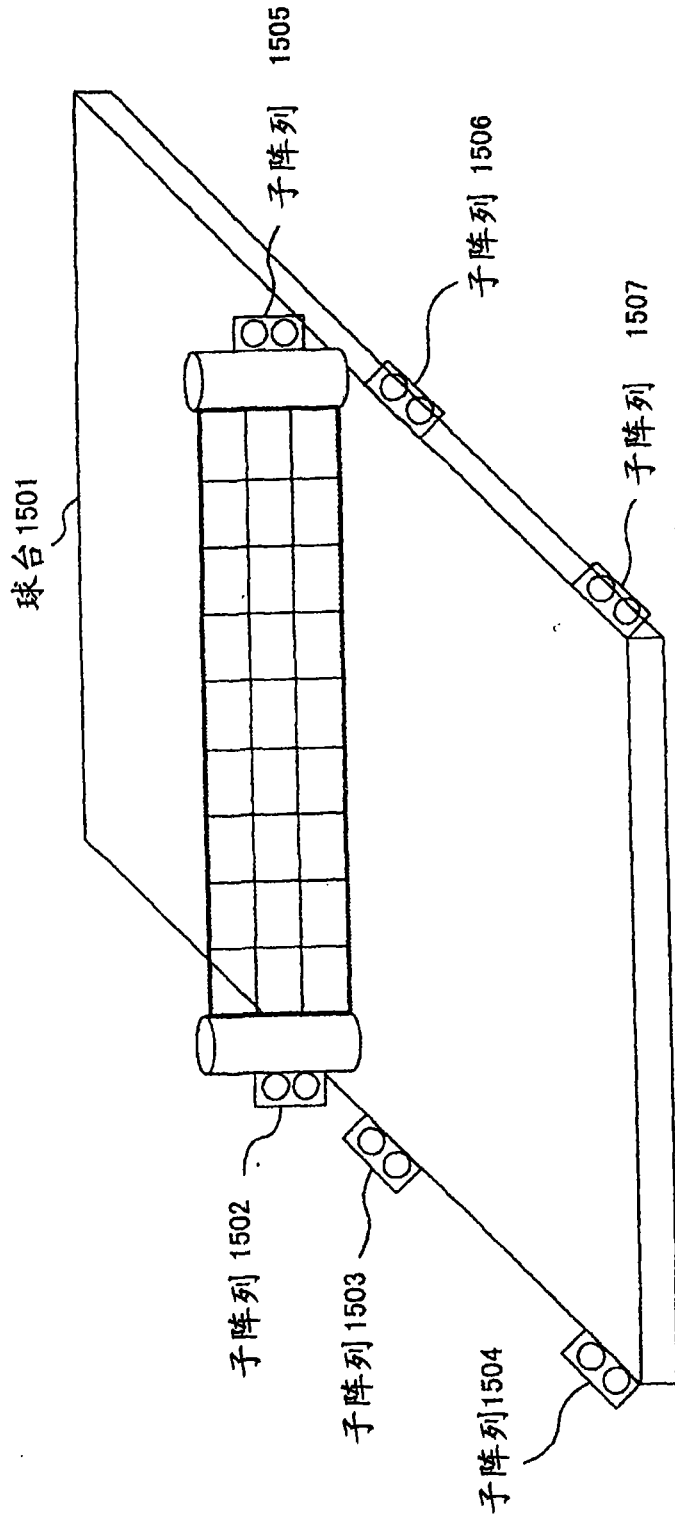


图16

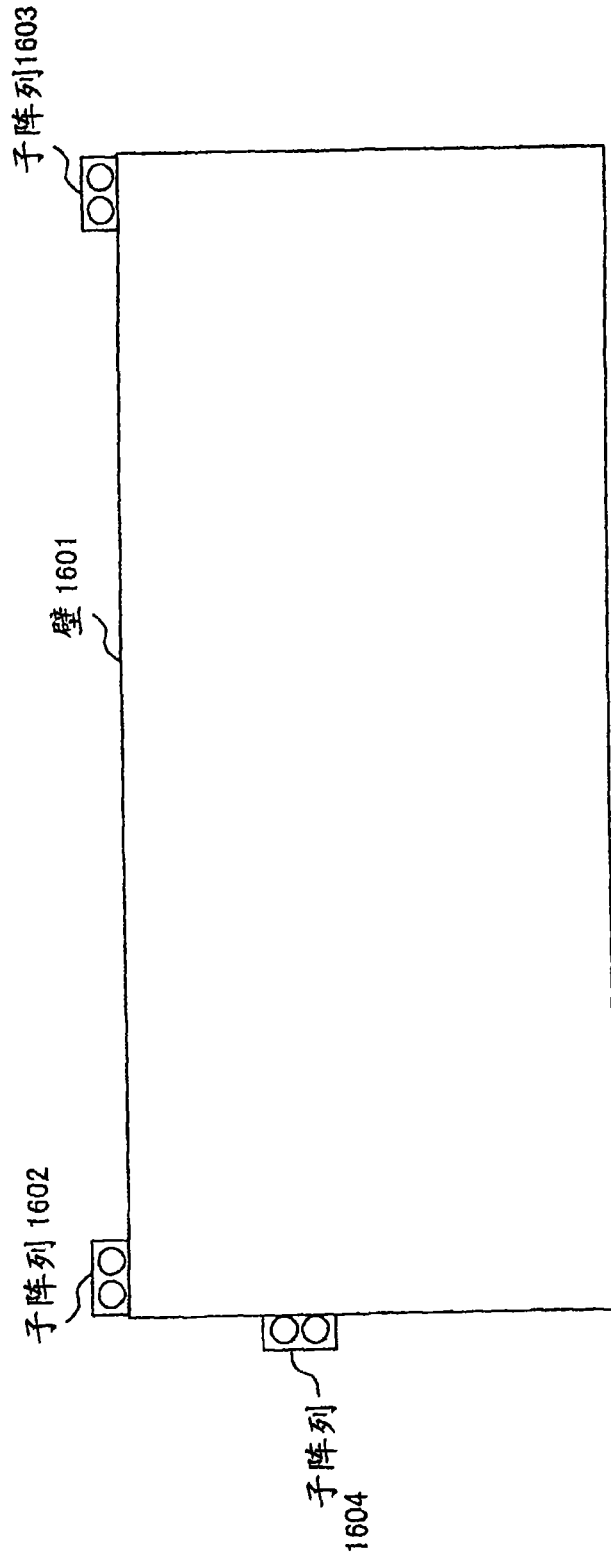


图17

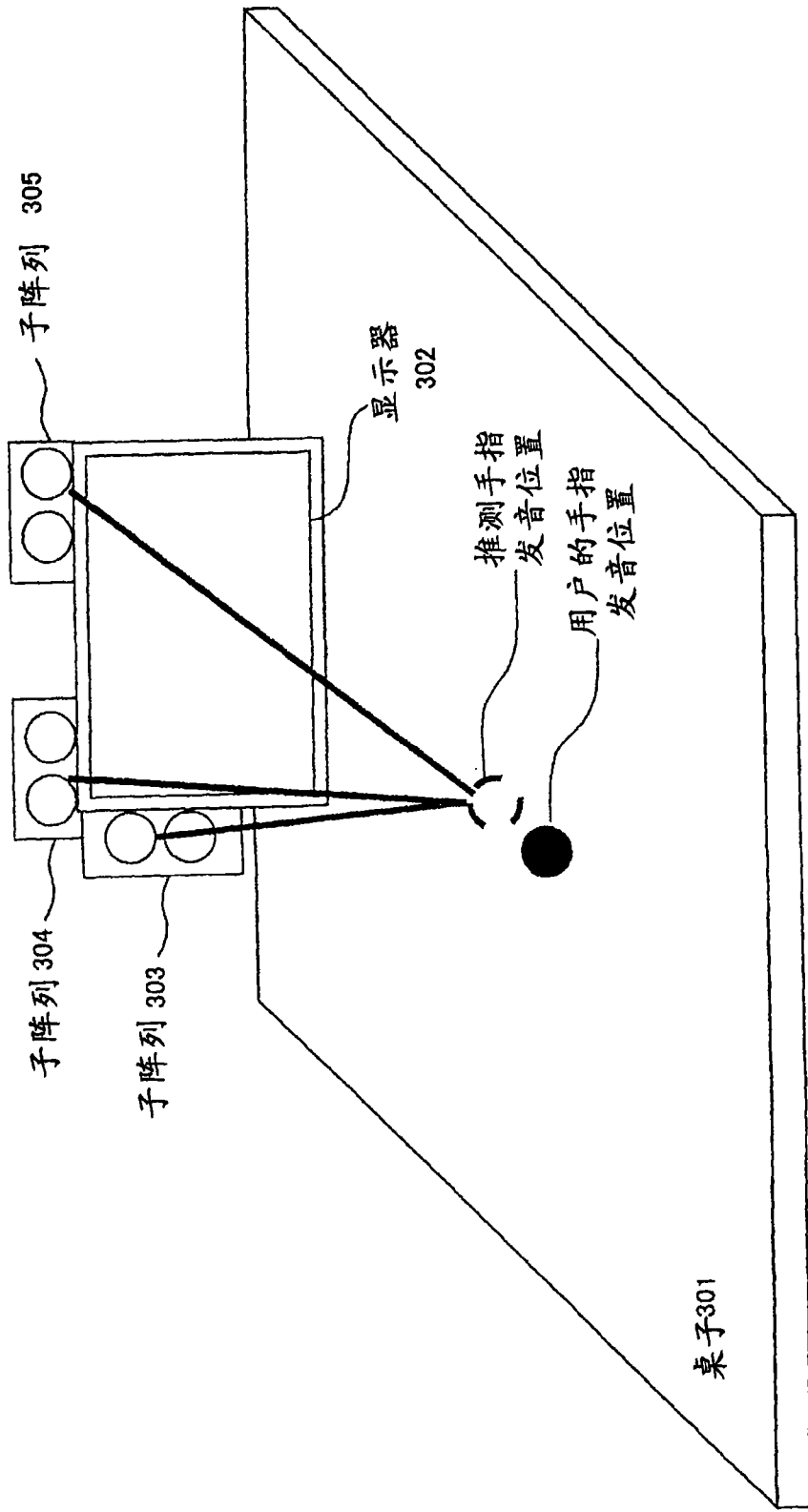


图18