

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公开说明书

[21] 申请号 200610059490.8

[51] Int. Cl.

G01S 3/808 (2006.01)

G01S 5/20 (2006.01)

G10L 21/02 (2006.01)

G06F 17/00 (2006.01)

[43] 公开日 2006 年 9 月 13 日

[11] 公开号 CN 1831554A

[22] 申请日 2006.3.13

[21] 申请号 200610059490.8

[30] 优先权

[32] 2005.3.11 [33] JP [31] 069824/2005

[71] 申请人 株式会社东芝

地址 日本东京都

[72] 发明人 铃木薰 古贺敏之

[74] 专利代理机构 永新专利商标代理有限公司

代理人 王 英

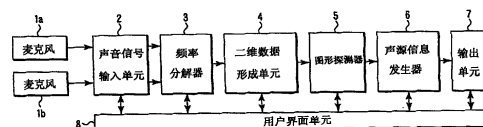
权利要求书 6 页 说明书 39 页 附图 20 页

[54] 发明名称

声音信号处理设备和声音信号处理方法

[57] 摘要

频率分解器分析从麦克风输入到声音信号输入单元的两个振幅数据，且二维数据形成单元对每一频率获得这两个振幅数据之间的相位差。对每一频率的这个相位差给定二维坐标值以形成二维数据。图形探测器分析在 X-Y 平面上所产生的二维数据以探测图形。声源信息发生器处理所探测到的图形的信息以产生声源信息，所述声源信息包含作为声音信号的发生源的声源的数量、每一声源的空间存在范围、由每一声源产生的声音的时间存在期间、每一源声的成分、每一声源的分离声音和每一源声的符号内容。



1、一种声音信号处理设备，包括：

声音信号输入器件，输入在至少两个点处拾取的多个声音信号，该至少两个点在空间中是不同的；

频率分解器，构造成将所述多个声音信号中的每一个分解以获得表示每一频率的相位值的多个频率分解数据组；

相位差计算器，构造成对于所述多个频率分解数据组中的不同频率分解数据组计算每一频率的相位差值；

二维数据形成器件，构造成对于每一对产生表示点群的二维数据，所述点具有在二维坐标系中的坐标值，在该二维坐标系中，频率的函数为第一轴，而通过相位差计算器所计算的相位差值的函数为第二轴；

图形探测器，构造成从该二维数据中探测表现出来自相同声源的频率与相位差之间的比例关系的图形；

声源信息产生器件，构造成根据所述图形来产生声源信息，该声源信息至少含有与声音信号的发生源相应的声源的数量、每一声源的空间存在范围、由每一声源产生的声音的时间存在期间、由每一声源产生的声音的分量构成、对于每一声源分离的分离声音和由每一声源产生的声音的符号内容中的一种，且该声源信息涉及彼此相互区分的各个声源；和

输出器件，输出所述声源信息。

2、根据权利要求1的设备，其中该二维数据形成器件包括构造成在二维坐标系上确定坐标值的坐标值确定器件，在该二维坐标系中，频率的标量倍数为该第一轴，而该相位差值的标量倍数为该第二轴。

3、根据权利要求1的设备，其中该二维数据形成器件包括构造在二维坐标系上确定坐标值的坐标值确定器件，在该二维坐标系中，所述频率的函数为第一轴，而根据由所述相位差计算器所计算的相位差值计算到时差的函数为第二轴。

4、根据权利要求2的设备，其中该图形探测器包括：

投票器件，构造成通过对具有由该坐标值确定器件确定的坐标值的点进行线性霍夫变换，在投票空间中投票来产生投票分布；和

直线探测器件，构造成根据该投票器件产生的所述投票分布来检测投票数量在预定阈值以上的峰值位置，并且根据检测的峰值位置确定直线。

5、根据权利要求3的设备，其中该图形探测器包括：

投票器件，构造成通过在预定方向上投影具有由该坐标值确定器件确定的坐标值的点，在投票空间中投票来产生作为投影投票的外围分布的投票分布；和

直线探测器件，构造成根据该投票器件产生的所述投票分布检测投票数量在预定阈值以上的峰值位置，并且根据检测的峰值位置确定直线。

6、根据权利要求4的设备，其中

该投票器件在该投票空间中投票固定值，和

该直线探测器件探测穿过所述二维坐标系中的每一频率的许多点的直线。

7、根据权利要求4的设备，其中

该频率分解器不仅计算每一频率的相位值还计算每一频率的功率值；

该投票器件投票根据所述功率值的数值，并且

该直线探测器件探测穿过二维坐标系中的每一频率的许多大功率的点的直线。

8、根据权利要求4的设备，其中当从所述投票分布中探测具有不小于预定阈值的投票数量的峰值位置时，该直线探测器件仅获取投票空间中的位置的峰值位置，该位置对应于所述二维坐标系中穿过特定位置的直线。

9、根据权利要求4的设备，其中

当从所述投票分布中探测具有不小于预定阈值的投票数量的峰值位置时，为了探测多个平行直线，该直线探测器件计算相应于构成多个平行直线的各直线的投票合计值，并探测该合计值在预定阈值以上的峰值位置，所述多个平行直线具有相同倾角，且彼此分离开根据所述倾角计算的预定距离。

10、根据权利要求4的设备，其中该声源信息产生器件包括方向估测器件，该方向估测器件构造成根据由该直线探测器探测的所述直线的倾角来计算作为相对于线段的角度声源的空间存在范围，该线段连接拾取所述声音信号的两个点。

11、根据权利要求4的设备，其中该声源信息产生器件包括声源成分估测器件，该声源成分估测器件构造成对于每一频率，计算所述坐标值与由该直线探测器件所探测的直线之间的距离，并根据该距离，估测由相应于该直线的声源所产生的声音的频率成分。

12、根据权利要求4的设备，其中所述声源信息产生器件包括：
声源成分估测器件，构造成对于每一频率计算所述坐标值与由所述直线探测器件所探测的直线之间的距离，并根据该距离，估测由相应于该直线的声源所产生的声音的频率成分；和
分离声音提取器件，构造成从所述声音的所述估测频率成分中合成由所述声源产生的声音信号数据。

13、根据权利要求11的设备，其中该声源成分估测器件确定一个频率为由相应于所述直线的声源产生的声音的频率成分，根据所述一个频率的所述坐标值到所述直线的距离不大于预定阈值。

14、根据权利要求11的设备，其中所述声源成分估测器件将所述坐标值到所述直线的距离在预定阈值内的频率作为由相应于所述直线的声源产生的声音的频率成分的候选项，并使该频率成分归属于最近的直线。

15、根据权利要求11的设备，其中
所述频率分解器不仅计算每一频率的相位值还计算每一频率的功率值，并且

所述声源成分估测器件计算随着所述坐标值到所述直线的距离中的增加而单调下降的非负系数，并确定通过将频率的功率乘以该非负系数所获得的值为由相应于所述直线的声源产生的声音的频率成分的功率值。

16、根据权利要求4的设备，其中所述声源信息产生器件包括：
方向估测器件，构造成根据由所述直线探测器件探测的所述直线

的倾角来计算作为相对于线段的角度声源的空间存在范围，该线段连接拾取所述声音信号的两个点；和

自适应阵列处理器件，构造成根据所述角度设定属于声源方向的跟踪范围，并且仅允许来自存在于该跟踪范围内的声源的声音穿过，由此提取由该声源产生的声音的声音信号的数据。

17、根据权利要求1的设备，还包括用户界面器件，该用户界面器件构造成使用户来检验和改变关于该设备操作的设定信息。

18、根据权利要求1的设备，还包括用户界面器件，该用户界面器件构造成使用户来保存和读取关于该设备操作的设定信息。

19、根据权利要求1的设备，还包括用户界面器件，该用户界面器件构造成向用户呈现所述二维数据或图形。

20、根据权利要求1的设备，还包括用户界面器件，该用户界面器件构造成向用户呈现所述声源信息。

21、根据权利要求1的设备，其中所述图形探测器从作为所述二维数据组的时间序列的三维数据组中探测图形。

22、一种声音信号处理方法，包括：

输入在至少两个点处拾取的多个声音信号，所述至少两个点在空间中是不同的；

将所述多个声音信号中的每一个分解以获得表示每一频率的相位值的多个频率分解数据组；

对于所述多个频率分解数据组中的不同频率分解数据组计算每

一频率的相位差值；

对于每一对，产生表示点群的二维数据，所述点群具有在二维坐标系中的坐标值，在该二维坐标系中，频率的函数为第一轴，而所计算的相位差值的函数为第二轴；

从该二维数据中探测表现出来自相同声源的频率与相位差之间的比例关系的图形；

根据所述图形来产生声源信息，该声源信息至少含有与所述声音信号的发生源相应的声源的数量、每一声源的空间存在范围、由每一声源产生的声音的时间存在期间、由每一声源产生的声音的分量构成、对于每一声源分离的分离声音和由每一声源产生的声音的符号内容中的一种，且该声源信息涉及彼此相互区分的各个声源；和

输出所述声源信息。

声音信号处理设备和声音信号处理方法

技术领域

本发明涉及声音信号处理，且特别地，涉及对例如在介质中传播的声波的发射源数量、每一个发射源的方向以及来自每一发射源的声波的频率成分的估测。

背景技术

近些年来，在机器人听觉研究的领域中，提出一种在噪音环境下估测多个目标声源的数量和方向（声源定位）并且分离和提取每一源声（声源分离）的方法。

例如，Futoshi Asano 在“Separating Sounds(声音分离)”(Measurement and Control, Vol.43, No. 4, pp. 325-330, April 2004)中描述了一种方法，该方法在具有背景噪音的环境下通过 M 个麦克风来测量 N 个声源、从利用 FFT（快速傅立叶变换）处理每一麦克风的输出而获得的数据中产生空间相关矩阵、将矩阵分解成本征值以获得大的主本征值（main eigenvalue）、并将声源的数量 N 估测为主本征值的数量。该方法利用了以下特性，即，具有方向性的信号诸如源声（source sound）映射到主本征值，而没有方向性的背景噪音映射到所有本征值。相应于主本征值的本征向量为由来自声源的信号展布（spread）的信号部分空间（partial space）的基向量，而相应于剩余本征值的本征向量为由背景噪音信号展布的噪音部分空间的基向量。可以通过利用噪音部分空间的基向量实施多信号分类（MUSIC）方法来搜索每一个声源的位置向量。可以通过给定了在通过搜索获得的方向上的方向性的光束成形器提取来自发现的声源的声音。然而，如果声源的数量 N 与麦克风的数量 M 相同，则不能限定噪音部分空间。

同样，如果声源的数量 N 超过 M ，则会存在不可探测的声源。因此，能够估测的声源数量小于麦克风的数量 M 。该方法不特别地对声源施加任何大的限制，且还在数学上很完美。然而，为了处理大数量的声源，需要比声源多的麦克风。

同样，Kazuhiro Nakadai 等在 “Real-time Active Person Tracking by Hierarchical Integration of Audiovisual Information” (Artificial Intelligence Society AI Challenge Research Meeting, SIG-Challenge-0113-5, pp. 35-42, June 2001) 描述了一种利用一个麦克风来执行声源定位和声源分离的方法。该方法以一种谐波结构（由基频和其谐波组成的频率结构）为基础，该谐波结构对于通过管子（发音物）发出的诸如人声的声音是特有的。在该方法中，从对麦克风拾取的声音信号进行傅立叶变换获得的数据中探测具有不同基频的谐波结构。所探测的谐波结构的数量被用作说话者的数量，以通过利用其 IPD（耳间相位差）和 IID（耳间声强差）来准确地估测每一谐波结构的方向。如此，通过其谐波结构来估测每一源声。通过从傅立叶变换数据中探测多个谐波结构，该方法可以处理多于麦克风的声源。然而，声源的数量和方向的估测与源声的估测以谐波结构为基础，因此可处理的声源仅限于具有谐波结构的声源，诸如人声。即，该方法不能处理各种声音。

如上所述，存在矛盾（自相矛盾）的问题：（1）如果声源不受限制，则声源的数量不能大于麦克风的数量；和（2）如果声源的数量大于麦克风的数量，则这些声源仅限于，例如，谐波结构。即，还没有设计出能够处理比麦克风多的声源而不限制这些声源的方法。

发明内容

考虑到上述情形，做出了本发明，且其目的是提供用于声源定位和声源分离的声音信号处理设备和声音信号处理方法，该设备和方法

可以减少对声源的限制且可以处理多于麦克风的声源。

根据本发明一个方案的声音信号处理设备包括：声音信号输入器件，构造成用于输入在不少于两个空间上不同的点处拾取的多个声音信号；频率分解器，构造成将多个声音信号中的每一个分解以获得多个表示每一频率的相位值的频率分解数据组；相位差计算器，构造成对于多个频率分解数据组中的一对不同频率分解数据组来计算每一频率的相位差值；二维数据形成器件，构造成对于每一对，产生表示具有在二维坐标系上的坐标值的点的二维数据，在该二维坐标系中，频率的函数作为第一轴，而通过相位差计算器所计算的相位差值的函数为第二轴；图形探测器，构造成从二维数据中探测表现来自相同声源的频率与相位差之间的比例关系的图形；声源信息产生器件，构造成根据该图形来产生声源信息，该声源信息含有以下内容的至少一种，与声音信号的发生源相应的声源数量、每一声源的空间存在范围、由每一声源产生的声音的时间存在期间、由每一声源产生的声音的分量构成、对于每一声源分离出的分离声音，和由每一声源产生的声音的符号内容，该符号内容与彼此相互区别的声源相关；以及输出器件，构造成输出声源信息。

附图简述

图 1 是根据本发明实施例的声音信号处理设备的功能框图；

图 2A 和 2B 是示出在声音信号中所观察到的声源方向和到时差的示图；

图 3 是示出帧与帧移动量之间的关系关系的示图；

图 4A 至 4C 是示出 FFT 顺序和 FFT 数据的示图；

图 5 是示出二维数据形成单元和图形探测器的内部设置的功能框图；

图 6 是示出相位差计算的顺序的示图；

图 7 是示出坐标值计算的顺序的示图；

图 8A 和 8B 是示出对于相同时间间隔的频率与相位之间的比例关系和对于相同时间差的相位差与频率之间的比例关系的示图；

图 9 是用于解释相位差的循环（circularity）的示图；

图 10A 和 10B 是当存在多个声源时频率与相位差的示图；

图 11 是用于解释线性霍夫变换的图；

图 12 是用于解释通过霍夫变换由点探测直线的示图；

图 13 是示出要被投票（voted）的平均功率的函数（方程式）的示图；

图 14 是示出从实际声音中产生的频率成分、相位差绘图和霍夫投票结果的示图；

图 15 是示出从实际霍夫投票结果中获得的峰值位置和直线的示图；

图 16 是示出 θ 与 $\Delta \rho$ 之间的关系的示图；

图 17 是示出当两个人同时发声时由实际声音产生的频率成分、相位差图和霍夫投票结果的示图；

图 18 是示出仅利用 θ 轴上投票执行的峰值位置检索结果的示图；

图 19 是示出通过在由 $\Delta \rho$ 分离的几个部分中将票数相加来执行的峰值位置检索结果的示图；

图 20 是示出声源信息发生器的内部设置的功能框图；

图 21A 至 21D 是用于解释方向估测的示图；

图 22 是示出 θ 与 ΔT 之间的关系的示图；

图 23A 至 23C 是用于解释在存在多个声源时的声源成分估测（间隔阈值方法）的示图；

图 24 是用于解释最近邻法的示图；

图 25 是示出用于计算系数 α 的方程式的实例和其图表的示图；

图 26 是用于解释 ϕ 在时间轴上的踪迹的示图；

图 27 是示出由声音信号处理设备执行的处理流程的流程图；

图 28A 和 28B 是示出可以被表述的频率与时间差之间的关系的示图；

图 29 是在产生冗余点时的时间差图；

图 30 是根据包括 N 个麦克风的变型例的声音信号处理设备的功能框图；

图 31 是根据一个实施例的功能框图，该实施例使用通用计算机来执行根据本发明的声音信号处理功能；和

图 32 是示出一种记录介质的实施例的示图，该记录介质记录用于执行根据本发明的声音信号处理功能的程序。

发明详述

下面将参考附图来详述根据本发明的声音信号处理设备的实施例。

图 1 是根据本发明实施例的声音信号处理设备的功能框图。该声音信号处理设备包括：麦克风 1a、麦克风 1b、声音信号输入单元 2、频率分解器 3、二维数据形成单元 4、图形探测器 5、声源信息发生器 6、输出单元 7 和用户界面单元 8。

[根据每一频率成分的相位差的声源估测的基本原理]

麦克风 1a 和 1b 为在诸如空气的介质中间隔预定距离的两个麦克风。麦克风 1a 和 1b 为用于将在两个不同点处的介质振动（声波）转换成电信号（声音信号）的装置。麦克风 1a 和 1b 在一起被提及时被称为麦克风对。

声音信号输入单元 2 是这样一种装置，用于以预定的取样周期 F_r 周期性地对由麦克风 1a 和 1b 获得的两个声音信号进行 A/D 转换，

来以时间顺序的方式产生这两个声音信号的数字振幅数据。

假设一个声源位于远远大于麦克风的间距的距离处，如图 2A 所示，由声源 100 产生并到达麦克风对的声波的波前（wave front）101 基本上为平面。当利用麦克风 1a 和 1b 在两个不同的点处观察到该平面波时，根据声源 100 相对于连接麦克风 1a 和 1b 的线段 102（下文称之为基线）的方向 R，可能观察到在由麦克风对所转换的声音信号之间的预定到达时间差 ΔT 。值得注意的是，当声源足够远时，如果声源 100 存在于与基线 102 垂直的平面上时，到时差 ΔT 为 0，且将该方向定义为麦克风对的正面方向（front direction）。

参考文件 1 “Kaoru Suzuki et al., “Realization of “It Comes When It’s Called” Function of Home Robot by Audio-Visual Interlocking”, The 4th Automatic Measurement Control Society System Integration Department Lecture Meeting (SI2003) Papers, 2F4-5, 2003” 描述了一种方法，该方法通过利用图形校对（pattern collation）来探测一个振幅数据的哪一部分与另一振幅数据的哪一部分相似，来推导出两个声音信号（图 2B 中的 103 和 104）之间的到时差 ΔT 。仅当存在一个强的声源时该方法有效。然而，如果存在强的背景噪音或存在多个声源，则在其中多个方向上的强声音混合的波形上不出现明显相似的部分，并且图形校对有时会失败。

因此，在根据本发明的这个实施例中，当将输入振幅数据分解成每一频率成分的相位差时，分析该输入振幅数据。当存在多个声源时，对于这些声源的每一频率成分可以观察到两个数据之间与声源方向相应的相位差。如果不对声源做强制限制而可以将单个频率成分的各个相位差划分成各单个方向的组，则能够估测声源的数量、这些声源的方向以及主要由每一声源产生的声波的频率成分的特性。虽然该理论本身非常简单，但是当实际分析数据时，存在一些待解决的问题。下面将解释这些问题和用于执行该分组的功能块（频率分解器 3、二

维数据形成单元 4 和图形探测器 5)。

[频率分解器 3]

FFT (快速傅立叶变换) 为一种将振幅数据分解成频率成分的常见方法。一个公知的典型算法为例如 Cooley-Turkey DFT 算法。

如图 3 中所示, 频率分解器 3 从由声音信号输入单元 2 输入的振幅数据 110 中提取 N 个连续振幅数据作为一帧 (第 T 个帧 111), 对所提取的帧执行 FFT, 并通过将提取位置移动帧移动量 113 来重复该处理 (提取第 (T+1) 个帧 112)。

如图 4A 中所示, 组成帧的振幅数据被加窗 (windowing) 120, 且然后进行 FFT 121。结果, 在实部缓冲器 R[N]和虚部缓冲器 I[N] (122) 中产生输入帧的 FFT 数据。图 4B 示出窗函数 (加汉明 (Hamming) 窗或加汉宁 (Hanning) 窗) 124 的实例。

通过将这一帧的振幅数据分解成 $N/2$ 个频率成分来获得如此产生的 FFT 数据。如图 4C 中所示, 对于第 k 个频率成分 f_k , 在缓冲器 122 中的实部 R[k]和虚部 I[k]的数字值表示在复坐标系 123 中的点 P_k 。 P_k 离原点 O 的距离的平方为该频率成分的功率 $P_o(f_k)$ 。 P_k 离实轴的带正负号的旋转角 θ ($\theta: -\pi > \theta \geq \pi$ (弧度)) 为该频率成分的相位 $Ph(f_k)$ 。

当取样频率为 $Fr[\text{Hz}]$ 且帧的长度为 $N[\text{采样点}]$ 时, k 取从 0 到 $(N/2) - 1$ 的整数值。在这种情况下, $k=0$ 表示 $0[\text{Hz}]$ (直流电), $k=(N/2) - 1$ 表示 $Fr/2[\text{Hz}]$ (最高频率成分), 且通过以频率分辨率 $\Delta f = (Fr/2) \div ((N/2) - 1) [\text{Hz}]$ 相等地划分这两个值之间的部分来获得每一个 k 的频率。这一频率由 $f_k = k \cdot \Delta f$ 表示。

值得注意的是, 如先前所述, 频率分解器 3 以预定间隔 (帧移动量 F_s) 来连续执行这一处理, 由此按照时间顺列方式为输入振幅数据的每一频率产生包括功率值和相位值的频率分解数据组。

（二维数据形成单元 4&图形探测器 5）

如图 5 中所示,二维数据形成单元 4 包括相位差计算器 301 和坐标值测定仪 302。图形探测器 5 包括投票单元 303 和直线探测器 304。

[相位差计算器 301]

相位差计算器 301 为用于比较由频率分解器 3 在相同定时获得的两个频率分解数据组 a 和 b,并通过对于每一个频率成分计算数据组 a 和 b 的相位值之间的差来产生 a-b 相位差数据的装置。例如,如图 6 所示,通过计算在麦克风 1a 处的相位值 $Ph1(fk)$ 与在麦克风 1b 处的相位值 $Ph2(fk)$ 之间的差来获得特定频率成分 fk 的作为 2π 余数系统的相位差 $\Delta Ph(fk)$,从而这一差值满足 $\{\Delta Ph(fk): -\pi < \Delta Ph(fk) \leq \pi\}$ 。

[坐标值测定仪 302]

坐标值测定仪 302 是这样一种装置,用于根据由相位差计算器 301 获得的相位差数据,来确定用于处理通过计算每一个频率成分的两个数据组的相位值之间的差而获得的相位差数据的坐标值,作为预定 X-Y 坐标系上的一个点。由图 7 中示出的方程式来确定与特定频率成分 fk 的相位差 $\Delta Ph(fk)$ 相应的 X 坐标值 $x(fk)$ 和 Y 坐标值 $y(fk)$ 。X 坐标值为相位差 $\Delta Ph(fk)$,而 Y 坐标值为频率成分号 k 。

[相位差对相同时间差的频率比例]

通过如图 6 中示出的相位差计算器 301 计算的单个频率成分的相位差被认为表示相同的到时差,如果它们来自相同的声源(相同方向)的话。通过 FFT 获得的特定频率的相位值和各麦克风之间的相位差为通过将频率的周期设定为 2π 所计算的值。如果频率加倍,则即使对于相同的时间差,相位差也加倍。图 8A 和 8B 示出该比例关系。

如图 8A 中所示, 对于相同的时间 T , 具有频率 $f_k[\text{Hz}]$ 的波 130 包含 $1/2$ 周期, 即, 相位间隔为 π , 而具有加倍频率 $2f_k[\text{Hz}]$ 的波 131 包含一个周期, 即, 相位间隔为 2π 。这类似地适用于相位差。即, 对于相同的时间差 ΔT , 相位差与频率成比例地增加。图 8B 示出相位差与频率之间的这种比例关系。当通过如图 7 中所示的坐标值计算将从相同的声源产生且具有共同的 ΔT 的单个频率成分的相位差绘制在二维坐标系上时, 表示单个频率成分的这些相位差的坐标点 132 布置在直线 133 上。随着 ΔT 增加, 即, 随着麦克风至声源的距离间的差增加, 该直线的倾斜度增加。

[相位差的循环]

值得注意的是, 从作为分析目标的最低频率到最高频率, 仅当真实的相位差落入 $\pm\pi$ 范围内时, 如图 8 所示, 各麦克风之间的相位差在整个区域中与频率成比例。这一条件是 ΔT 小于最高频率 (采样频率的一半) $F_r/2[\text{Hz}]$ 的 $1/2$ 周期的时间, 即, 小于 $1/F_r[\text{sec}]$ 。如果 ΔT 为 $1/F_r$ 或更大, 则如下所述, 相位差仅可以获得为具有循环性的值。

可以仅在 2π 的宽度 (在这一实施例中, $-\pi$ 与 π 之间的 2π 宽度) 内获得每一频率成分的相位值作为旋转角 θ 的值, 如图 4C 中所示。这意味着: 即使当各麦克风之间该频率成分中的实际相位差为一个周期或更多时, 其不能从作为频率分解结果所获得的相位值中获知。因此, 在本实施例中, 在 $-\pi$ 与 π 之间获得相位差, 如图 6 中所示。然而, 由 ΔT 产生的真实相位差可以通过向所获得的相位差值添加 2π 或从所获得的相位差值中减去 2π 、或者进一步向所获得的值添加 4π 或 6π 或从其中减去 4π 或 6π 所计算的值得到。这在图 9 中示意性示出。参考图 9, 当频率 f_k 的相位差 $\Delta Ph(f_k)$ 为 $+\pi$ 时, 如实心圆 140 所示, 相邻的较高的频率 f_{k+1} 的相位差超过 $+\pi$, 如空心圆 141 所示。然而, 通过从原始的相位差中减去 2π 所获得的计算得到的相位差 Δ

$\text{Ph}(f_{k+1})$ 稍微大于 $-\pi$ ，如实心圆 142 所示。虽然未示出，即使三重频率（three-fold frequency）也显示出通过从实际相位差中减去 4π 所获得的相似值。因此，随着频率上升，相位差作为 2π 余数系统在 $-\pi$ 与 π 之间循环。如果 ΔT 像在这一实例中那样增加，当频率为特定频率 f_{k+1} 或更高时，则由空心圆表示的真实相位差循环到由实心圆表示的相反一侧。

[当存在多个声源时的相位差]

另一方面，当从多个声源产生声波时，频率与相位差的绘图如图 10A 和 10B 中示意性示出的那样。图 10A 和 10B 示出其中两个声源存在于相对于麦克风对的不同方向上的情况。图 10A 示出其中两个源声不包含相同频率成分的情况。图 10B 示出一些相同的频率成分均包含在两个源声中的情况。参考图 10A，每一个频率成分的相位差呈现于多条共同具有 ΔT 的直线中的一条上。即，五个点布置在具有小倾角的直线 150 上，而六个点布置在具有大倾角的直线 151（包括循环直线 152）上。参考图 10B，在包含于两个源声的两个频率成分 153 和 154 中，由于波的混合，不能正确地获得相位差，所以相位差不在任何直线上。特别地，仅三个点布置在具有小倾角的直线 155 上。

估测声源数量和方向的问题将其自身转变为在如图 10A 和 10B 中所示的绘图中寻找直线。同样，估测每一个声源的频率成分的问题将其自身转变为选择布置在所探测的直线附近的频率成分。在该实施例中，从二维数据形成单元 4 输出的二维数据为通过利用由频率分解器 3 获得的两个频率分解数据组来被确定为频率和相位差的函数的点群（dot group），或者是通过将该点群中的点布置（绘制）在二维坐标系上所获得的图像。值得注意的是，这个二维数据由不包括时间轴的两个轴来定义，因此可以定义作为二维数据的时间序列的三维

数据。作为图形，图形探测器 5 从给定为该二维数据（或作为二维数据的时间序列的三维数据）的点的布置中探测直线的布置。

[投票单元 303]

投票单元 303 是这样一种装置，用于如稍后介绍的对由坐标值确定器 302 给定 (x, y) 坐标的每一频率成分实施线性霍夫变换，并通过预定方法在霍夫投票空间中对所获得的轨迹投票。虽然霍夫变换在参考文件 2000 年 10 月 20 日出版的“Akio Okazaki, “First Step in Image Processing”, Industrial Investigation Society, pp.100-102”中描述过，但下面将再次对其进行解释。

[线性霍夫变换]

如图 11 中示意性示出，诸如直线 160、161 和 162 的无数直线可以穿过二维坐标系中的点 $P(x, y)$ 。然而，当从原点 O 引出到每一直线的垂线 163 的倾角为 θ ，而垂线 163 的长度为 ρ 时，对于一条直线而言， θ 和 ρ 是唯一确定的。穿过某个点 (x, y) 的直线的多对 θ 和 ρ 可以在 $\theta - \rho$ 坐标系中绘制相对于 (x, y) 的值唯一的轨迹 164 ($\rho = x \cos \theta + y \sin \theta$)。这种从 (x, y) 坐标值到穿过 (x, y) 的直线的 (θ, ρ) 的轨迹的变换被称之为线性霍夫变换。值得注意的是，当直线向左倾斜， θ 为正值，当直线竖直时， θ 为 0，而当直线向右倾斜， θ 为负值。还值得注意的是， θ 的值域不超出 $\{\theta : -\pi < \theta \leq \pi\}$ 。

对于 $X-Y$ 坐标系上的每一个点，可以独立地获得霍夫曲线。然而，如图 12 中所示，例如，可以获得穿过三个点 p_1 、 p_2 和 p_3 的直线 170 作为由点 174 的坐标 (θ_0, ρ_0) 限定的直线，在点 174 处，分别相应于点 p_1 、 p_2 和 p_3 的轨迹 171、172 和 173 彼此相交。随着直线穿过的点的数量增加，穿过表示该直线的位置 θ 和 ρ 的轨迹数量增加。如上所述，霍夫变换适合于从点中探测直线。

[霍夫投票]

被称为霍夫投票的工程方法用于从点中探测直线。在该方法中，在以 θ 和 ρ 作为其坐标轴的二维霍夫投票空间中，每一轨迹穿过的 θ 和 ρ 对被投票，由此明示出有大量轨迹穿过的 θ 和 ρ 对，即，在霍夫投票空间中具有许多得票的位置上存在直线。通常，首先准备具有用于 θ 和 ρ 的必要检索范围的尺寸的二维阵列（霍夫投票空间），并用 0 将其初始化。然后，通过霍夫变换来获得每个点的轨迹，对该轨迹通过的阵列上的值加 1。这被称为霍夫投票。当完成对所有点的轨迹的投票时，在没有得票的位置（没有轨迹通过它）不存在穿过一个点的直线，在具有一票的位置（一个轨迹通过它）存在通过一个点的直线，在具有两票的位置（两个轨迹通过它）存在通过两个点的直线，而在具有 n 票的位置（ n 个轨迹通过它）存在通过 n 个点的直线。如果霍夫投票空间的分辨率可以是无限的，则由多个轨迹穿过的点仅获得相应于穿过该点的轨迹数量的投票。然而，由于实际霍夫投票空间以 θ 和 ρ 的适当分辨率被量化，所以在多个轨迹彼此相交的位置周围产生高的投票分布。因此，需要通过从霍夫投票空间的投票分布中搜索具有最大值的位置来精确获得多个轨迹彼此相交的位置。

投票单元 303 对都满足下述投票条件的频率成分执行霍夫投票。在该条件下，仅对在预定频带中且其功率等于或高于预定阈值的频率成分执行投票。

即，投票条件 1 为频率落入预定范围（低频界限与高频界限）内。投票条件 2 为频率成分 f_k 的功率 $P(f_k)$ 等于或高于预定阈值。

投票条件 1 用于去掉通常携带暗噪音的低频，以及去掉 FFT 精确度下降的高频。根据操作可以调节低频界限和高频界限的范围。当使用最宽的频带时，优选仅去掉作为低频的 DC 成分并且仅去掉作为高频的最大值频率。

对于诸如暗噪音的非常弱的频率成分,FFT 结果的可靠性可能很低。通过利用功率的阈值处理,投票条件 2 用于防止该低可靠性频率成分参与投票。假设麦克风 1a 具有功率值 $P_{o1}(fk)$,而麦克风 1b 具有功率值 $P_{o2}(fk)$,则可以使用下述三个条件来确定要被估测的功率 $P(fk)$ 。值得注意的是,可以根据操作来选择要使用的条件。

(平均值):使用 $P_{o1}(fk)$ 和 $P_{o2}(fk)$ 的平均值。该条件需要两个功率都适当地强。

(最小值):使用 $P_{o1}(fk)$ 和 $P_{o2}(fk)$ 中较小的一个。该条件要求这两个功率都至少等于阈值。

(最大值):使用 $P_{o1}(fk)$ 和 $P_{o2}(fk)$ 中较大的一个。在该条件下,即使当一个小于阈值时,如果另一个足够强则可以执行投票。

同样,投票单元 303 可以在投票中执行下述两种加法的方法。

即,在加法方法 1 中,向轨迹穿过的一个位置添加预定的固定值(例如,1)。在加法方法 2 中,向轨迹穿过的一个位置添加频率成分 fk 的功率 $P(fk)$ 的函数值。

加法方法 1 通常用于霍夫变换直线探测问题。因为与穿过的点的数量成比例地整理投票,所以加法方法 1 适合于优先探测含有许多频率成分的直线(即,声源)。在该方法中,包含在直线中的频率成分不必具有任何谐波结构(在其中所包含的频率被相等地隔开)。因此,可以探测各种类型的声源以及人的声音。

在加法方法 2 中,即使当通过的点的数量很小时,如果含有高功率频率成分,则可以获得较高位置中的最大值。加法方法 2 适合于探测含有小量的频率成分但具有高功率、有影响力的成分的直线(即,声源)。在加法方法 2 中,功率 $P(fk)$ 的函数值被作为 $G(P(fk))$ 计算。图 13 示出在 $P(fk)$ 为 $P_{o1}(fk)$ 和 $P_{o2}(fk)$ 的平均值时用于计算 $G(P(fk))$ 的方程式。此外,如同在投票条件 2 中, $P(fk)$ 还可以计算为 $P_{o1}(fk)$ 和 $P_{o2}(fk)$ 中的最小值或最大值。即,可

以独立于投票条件 2 而根据操作来设定加法方法 2。将中间参数 V 的值作为通过向 $P(fk)$ 的对数值 $\log_{10}(P(fk))$ 添加预定偏移量 α 而获得的值来计算。如果 V 为正, 则将 $V+1$ 的值用作函数 $G(P(fk))$ 的值, 而如果 V 为零或更小, 则使用 1。通过这样至少投票 1, 也可以将加法方法 1 的由大多数决定的特性赋予加法方法 2, 即, 不仅含有高功率频率成分的直线(声源)浮置在较高位置, 而且含有许多频率成分的直线(声源)浮置在较高位置。投票单元 303 可以根据设定来执行加法方法 1 或加法方法 2。然而, 当采用后一种方法时, 还可以同时探测具有少量频率成分的声源。因此, 可以探测更多种类型的声源。

[多个 FFT 结果的共同投票]

此外, 虽然无论何时执行 FFT, 投票单元 303 就可以投票, 但是其通常对 m ($m \geq 1$) 个连续的时间序列 FFT 结果执行共同投票 (collective voting)。声源的频率成分在长时间内发生变化。然而, 通过这样执行共同投票, 通过利用在频率成分保持稳定的适当短周期期间的多个时刻从 FFT 结果中获得的大量数据, 可以获得具有较高可靠性的霍夫投票结果。值得注意的是, 可以将 m 设定为根据操作的参数。

[直线探测器 304]

直线探测器 304 是用于通过分析由投票单元 303 在霍夫投票空间上产生的投票分布来探测高功率直线的装置。值得注意的是, 在这一情况下, 通过考虑该问题的特有情形, 例如参考图 9 所阐释的相位差的循环, 可以以更高的精确度来探测直线。

图 14 示出当在室内噪音环境下处理由在相对于离麦克风对的前方的左侧大约 20° 角处的人发出的实际声音时频率成分的功率谱、从

五个 (m (上述的) = 5) 连续的 FFT 结果中获得的每一频率成分的相位差图、以及从相同的五个 FFT 结果中获得的霍夫投票结果 (投票分布)。一直到此的处理是通过从声音信号输入单元 2 至投票单元 303 的一系列功能模块执行的。

利用频率分解器 3, 对于每一个频率成分, 将通过麦克风对获得的振幅数据转换为功率值和相位值的数据。参考图 14, 180 和 181 是这样的示图, 在其每一幅中, 以横轴为时间, 由亮度表示每一频率成分的功率值的对数 (指示越暗, 值越大)。一个垂直线相应于一个 FFT 结果, 且这些线形成为沿着时间流逝 (向右) 的曲线图。上部 180 示出处理来自麦克风 1a 的信号的结果, 而下部 181 示出处理来自麦克风 1b 的信号的结果。许多频率成分在上部和下部都被探测到。根据这些频率分解结果, 相位差计算器 301 计算每一频率成分的相位差, 而坐标值确定器 302 计算相位差的 (x , y) 坐标值。参考图 14, 182 为绘制了通过从特定时刻 183 的五个连续 FFT 处理中获得的相位差的示图。在该图中, 沿着从原点向左侧倾斜的直线 184 显示出点分布。然而, 该分布没有精确地存在于直线 184 上, 且存在与直线 184 分离的大量点。投票单元 303 在霍夫投票空间内对这样分布的点投票以形成投票分布 185。值得注意的是, 该投票分布 185 是利用加法方法 2 产生的。

[限制 $\rho = 0$]

当通过声音信号输入单元 2 对来自麦克风 1a 和 1b 的信号进行相互同相的 A/D 转换时, 则要探测的直线总是满足 $\rho = 0$, 即, 总是通过 X-Y 坐标系的原点。因此, 声源估测问题将其自身转化为在霍夫投票空间中从 $\rho = 0$ 的 θ 轴上的投票分布中检索最大值的问题。图 15 示出从图 14 示出的数据中检索 θ 轴上的最大值的结果。

图 15 中示出的投票分布 190 与图 14 中示出的投票分布 185 相同。

通过提取 θ 轴 191 上的投票分布 $S(\theta, 0)$ 作为 $H(\theta)$ 来获得条线图 192。投票分布 $H(\theta)$ 具有一些峰值部分（突出部分）。只要相同的点其自身在投票分布 $H(\theta)$ 的某个位置的左和右侧连续，直线探测器 304 (1) 就执行检索，并留下一个部分，在该部分中最后仅出现少于其自身的更少量的投票。结果，提取了投票分布 $H(\theta)$ 中的峰值部分。由于该峰值部分包括具有平坦峰值的部分，在象这样的部分中最大值连续。因此，直线探测器 304 (2) 通过线细化处理（line thinning process）仅留下峰值部分的中央位置作为峰值位置 193。最后，作为直线，直线探测器 304 (3) 仅探测其中投票数量等于或大于预定阈值的峰值位置。这样，可以精确地发现具有足够投票的直线的 θ 。在图 15 中示出的实例中，上述在 (2) 中探测到的峰值位置 194、195 和 196 中，峰值位置 194 是通过线细化处理从平坦峰值部分中留下的中央位置（如果有偶数个峰值位置连续，则右侧位置优先）。而且，仅峰值位置 196 为通过获得比阈值大的投票数量探测到的直线。通过由峰值位置 196 给定的 θ 和 $\rho (=0)$ 来定义直线（参考直线）197。值得注意的是，作为线细化处理的运算法则，可以使用在用于解释霍夫变换的参考文件 2 的第 89 至 92 页中描述的“Tamura's Method”的一维形式。当这样探测一个或多个峰值位置（投票数量等于或大于预定阈值的中央位置）时，直线探测器 304 按照投票的递减顺序排列这些位置，并输出每一个峰值位置的 θ 和 ρ 的值。

[考虑相位差循环的直线组的定义]

图 15 中示出的直线 197 穿过由峰值位置 196，即 $(\theta_0, 0)$ 定义的 X-Y 坐标原点。然而，实际上，由于相位差的循环，直线 198 也示出与直线 197 相同的到时差。当图 15 中示出的直线 197 平行移动 $\Delta \rho$ 199 并从 X 轴的相反侧循环时获得直线 198。当直线 197 延伸且从 X 值区延伸出的部分从相反侧循环出现时所获得的诸如直线

198 的直线在下文中被称之为“循环延伸线”，且作为参考的直线 197 在下文中被称为“参考直线”。如果参考直线 197 进一步倾斜，循环延伸线的数量进一步增加。如果系数 a 为 0 或 0 以上的整数，则具有相同到时差的所有直线形成直线组 $(\theta, 0, a\Delta\rho)$ ，该直线组 $(\theta, 0, a\Delta\rho)$ 是由 $(\theta, 0, 0)$ 限定的参考直线 197 一次平行移动 $\Delta\rho$ 时所获得的。此外，如果通过去掉限制 $\rho=0$ 将作为起点的 ρ 推广为 $\rho=\rho_0$ ，则该直线组可以被表述为 $(\theta, 0, a\Delta\rho+\rho_0)$ 。在这种情况下， $\Delta\rho$ 为由图 16 中示出的作为直线的倾角 θ 的函数 $\Delta\rho(\theta)$ 的方程式限定的带正负号的值。

参考图 16，可以由 $(\theta, 0)$ 定义参考直线 200。由于该参考直线向右倾斜，所以根据该限定， θ 具有负值。然而，在图 16 中，将 θ 在图 16 中作为绝对值处理。图 16 中示出的直线 201 为参考直线 200 的循环延伸线，且在点 R 处与 X 轴相交。而且，参考直线 200 与循环延伸线 201 之间的间隔为由辅助线 202 表示的 $\Delta\rho$ 。辅助线 202 在点 O 处与参考直线 200 垂直相交，并且在点 U 处与循环延伸线 201 垂直相交。由于该参考直线向右倾斜，所以根据定义， $\Delta\rho$ 具有负值。然而，在图 16 中将 $\Delta\rho$ 作为绝对值处理。图 16 中的 $\triangle OQP$ 为其中边 OQ 的长度为 π 的直角三角形。与 $\triangle OQP$ 全等的三角形为 $\triangle RTS$ 。因此，边 RT 的长度也为 π ，而 $\triangle OUR$ 的斜边 OR 的长度为 2π 。由于 $\Delta\rho$ 为边 OU 的长度，所以 $\Delta\rho=2\pi\cos\theta$ 。通过考虑 θ 和 $\Delta\rho$ 的正负号来推导出图 16 中示出的方程式。

[考虑到相位差循环的峰值位置探测]

如上所述，由于相位差的循环，不应该将表示声源的直线作为一个直线处理，而应该作为包括参考直线和循环延伸线的直线组来处理。当要从投票分布中探测峰值位置时，也必须考虑这个问题。当仅在麦克风对的前方（front）附近中探测声源时，在那里不发生相位差

循环，或即使发生相位差循环，其规模也很小，仅通过在 $\rho = 0$ （或 $\rho = \rho_0$ ）条件下的投票数量（即参考直线的投票数量）来检索峰值位置的上述方法不仅在性能上是令人满意的，而且还具有缩短检索时间和增加检索精确度的效果。然而，当要探测较宽范围内的声源时，需要将在相对于特定 θ 的几个彼此间隔 $\Delta \rho$ 的部分中的投票数量加起来来检索峰值位置。下面将解释这种不同。

图 17 示出当处理室内噪音环境中由离麦克风对的前方向左呈大约 20° 角和向右呈大约 45° 角的两个人同时发出的实际声音时所获得的频率成分的功率频谱、从五个 ($m=5$) FFT 结果中获得的每一频率成分的相位差图、以及从相同的五个 FFT 结果中获得的霍夫投票结果（投票分布）。

通过频率分解器 3，对于每一频率成分，将由麦克风对获得的振幅数据转换成功率值和相位值的数据。参考图 17，210 和 211 是这样的示图，在其中每一幅中，以频率为纵坐标轴而以时间为横坐标轴，每一频率成分的功率值的对数由亮度表示（指示越暗，值越大）。一条垂直线相应于一个 FFT 结果，且这些线沿着时间流逝（向右）形成一幅图。上部 210 示出来自麦克风 1a 的信号的处理结果，而下部 211 示出来自麦克风 1b 的信号的处理结果。在上和下部中都探测到许多频率成分。根据这些频率分解结果，相位差计算器 301 计算每一频率成分的相位差，而坐标值确定器 302 计算相位差的 (x, y) 坐标值。在示图 212 中，绘制通过从特定时刻 213 的五个连续 FFT 处理获得的相位差。示图 212 示出沿着从原点向左侧倾斜的直线 214 的点分布、以及沿着向右倾斜的参考直线 215 的点分布。投票单元 303 在霍夫投票空间内对这样分布的点投票以形成投票分布 216。值得注意的是，通过利用加法方法 2 来产生投票分布 216。

图 18 是示出仅通过 θ 轴上的投票数量来检索峰值位置的结果的示图。图 18 中的投票分布 220 与图 17 中示出的投票分布 216 相同。

通过提取 θ 轴 221 上的投票分布 $S(\theta, 0)$ 作为 $H(\theta)$ 来获得条线图 222。投票分布 $H(\theta)$ 具有一些峰值部分（突出部分）。通常， θ 的绝对值越大，投票数量越小。从投票分布 $H(\theta)$ 中，探测到四个峰值位置 224、225、226 和 227，如峰值位置示图 223 所示。在这些峰值位置中，仅峰值位置 227 具有大于阈值的投票数量。以这种方式，探测到一个直线组（参考直线 228 和循环延伸线 229）。通过探测从麦克风对的正面向左呈约 20° 角的声音来获得该直线组。然而，没有探测从麦克风对的正面向右呈约 45° 角的声音。随着穿过原点的参考直线的角度增大，该参考直线在超出 X 值域之前能够穿过的频带数量减小。因此，该参考直线穿过的频带的宽度随着 θ 而改变（即，存在不公平）。限制 $\rho=0$ 仅使参考直线的投票数量在该不公平的条件下相互竞争。因此，直线的角度越大，投票竞争的缺陷越大。这就是为什么不能探测向右呈约 45° 角的声音的原因。

图 19 示出通过将在几个彼此分离 $\Delta \rho$ 的部分中的投票数量加起来来检索峰值位置的结果。在图 19 的 240 中，在图 17 所示的投票分布 216 上，穿过原点的直线一次平行移动 $\Delta \rho$ 时所得到的 ρ 的位置由点线 242 至 249 示出。 θ 轴 241 和点线 242 至 245 与 θ 轴 241 和点线 246 至 249 以 $\Delta \rho(\theta)$ 的自然数倍数相等地间隔开。值得注意的是，对于 $\theta=0$ 的条件，没有点线存在，在该条件下所述直线不超过 X 值域，而是可靠地延伸至该示图的顶部。

当在 $\theta = \theta_0$ 的位置垂直观察时，将特定的 θ_0 的投票 $H(\theta_0)$ 作为在 θ 轴 241 上以及点线 242 至 249 上的投票总值来计算，即， $H(\theta_0) = \sum \{S(\theta_0, a\Delta \rho(\theta_0))\}$ 。该处理等同于将 $\theta = \theta_0$ 时的参考直线的投票和其循环延伸线的投票加合起来。图 19 中的 250 示出作为条线图的投票分布 $H(\theta)$ 。与图 18 的 222 中不同，在这一分布中，即使当 θ 的绝对值增加，投票的数量也不减少。这是因为：通过在投票计算中加入循环延伸线，可以对所有的 θ 值使用相同的频

带。在投票分布 250 中，探测到在图 19 的 251 中示出的十个峰值位置。在这些峰值位置中，峰值位置 252 和 253 各自获得大于阈值的投票数量，并且探测到两个直线组。即，通过探测从麦克风对的前方向左呈约 20° 角的声音来探测一个直线组（与峰值位置 253 相应的参考直线 254 和循环延伸线 255），并通过探测从麦克风对的前方向右呈约 45° 角的声音来探测另一个直线组（与峰值位置 252 相应的参考直线 256 和循环延伸线 257 和 258）。通过加合在彼此分开 $\Delta \rho$ 的部分中的投票来如此检索峰值位置，能够稳定地从具有小角度的直线到具有大角度的直线中探测直线。

[考虑到不同相的峰值位置探测：一般化]

如果没有通过声音信号输入单元 2 将来自麦克风 1a 和 1b 的信号进行相互同相的 A/D 转换，则要被探测的直线为 $\rho = 0$ ，即，不穿过 X-Y 坐标原点。在这种情况下，必须通过除去限制 $\rho = 0$ 来检索峰值位置。

当将从其中除去限制 $\rho = 0$ 的参考线一般性地描述为 (θ_0, ρ_0) 时，可以将其直线组（参考直线和循环延伸线）描述为 $(\theta_0, a\Delta\rho(\theta_0) + \rho_0)$ 。 $\Delta\rho(\theta_0)$ 为由 θ_0 确定的循环延伸线的平行移动量。当声源来自特定方向时，对于相应于声源的 θ_0 ，仅存在一个最强的直线组。通过使用 ρ_0 的值 $\rho_{0\max}$ ，将该直线组表示为 $(\theta_0, a\Delta\rho(\theta_0) + \rho_{0\max})$ ，其中 $\rho_{0\max}$ 表示，当 ρ_0 有各种改变时该直线组的投票 $\sum\{S(\theta_0, a\Delta\rho(\theta_0) + \rho_0)\}$ 为最大值。因此，通过利用每一个 θ 的投票 $H(\theta)$ 作为该 θ 的最大投票 $\sum\{S(\theta_0, a\Delta\rho(\theta_0) + \rho_0)\}$ ，能够利用与施加限制 $\rho = 0$ 时所使用的算法相同的峰值位置探测算法来执行直线探测。

值得注意的是，这样探测的直线组的数量为声源的数量。

[声源信息发生器 6]

如图 20 中所示, 声源信息发生器 6 包括方向估测器 311、声源成分估测器 312、声源再合成器 313、时间序列跟踪单元 314、持续时间鉴别器 315、相位匹配单元 316、自适应阵列处理器 317 和语音识别单元 318。

[方向估测器 311]

方向估测器 311 是用于接收通过上述直线探测器 304 获得的直线探测结果的装置, 即, 接收每一个直线组的 θ 值, 并计算与每一直线组相应的声源的存在范围。所探测的直线组的数量为声源的数量 (所有候选项)。如果离声源的距离远远长于麦克风对的基线, 则声源存在范围为与该麦克风对的基线成特定角度的圆锥。下面参考图 21 将对其进行阐释。

麦克风 1a 和 1b 之间的到时差 ΔT 可以在 $\pm \Delta T_{\max}$ 的范围内改变。当声音从前方入射, 如图 21A 中所示, 则 ΔT 为 0, 而该声源距离前方的方位角 ϕ 为 0° 。当声音从右侧呈直角入射时, 即, 在如图 21B 中所示的麦克风 1b 的方向上入射时, ΔT 等于 $+\Delta T_{\max}$, 而当假设从前方顺时针旋转为正方向时, 声源的方位角 ϕ 为 $+90^\circ$ 。同样, 当声音从左侧呈直角入射时, 即, 在如图 21C 中所示的麦克风 1a 的方向上入射时, ΔT 等于 $-\Delta T_{\max}$, 而方位角 ϕ 为 -90° 。如上所述, 如此定义 ΔT 以便于当声音从右侧入射时其为正, 而当声音从左侧入射时其为负。

根据上述定义, 下面将描述如图 21D 中所示的一般情况。假设麦克风 1a 的位置为 A, 麦克风 1b 的位置为 B, 而声音从线段 PA 的方向入射, 则 $\triangle PAB$ 为直角三角形, 其顶点 P 为直角。在这种情况下, 假设麦克风之间的中心为 O, 线段 OC 为麦克风对的前方方向, 且 OC 方向的方位角为 0° , 将逆时针旋转为正的角定义为方位角 ϕ 。

由于 $\triangle QOB$ 相似于 $\triangle PAB$ ，所以方位角 ϕ 的绝对值等于 $\angle OBQ$ ，即， $\angle ABP$ ，且其符号与 ΔT 的符号匹配。同样，可以将 $\angle ABP$ 作为 PA 与 AB 的比率的 $\sin^{-1}$ 来计算。当线段 PA 的长度由相应的 ΔT 来表示时，线段 AB 的长度等同于 $\Delta T_{\max}$ 。因此，可以通过 $\phi = \sin^{-1}(\Delta T / \Delta T_{\max})$ 来计算该方位角和其正负号。将声源存在范围估测为圆锥 260，该圆锥 260 以点 O 作为其顶点，并以基线 AB 作为其轴，并张开 $(90 - \phi)^\circ$ 。声源在圆锥 260 上的某处。

如图 22 中所示，通过将麦克风间距离 $L[m]$ 除以声速 $V_s[m/sec]$ 来计算 $\Delta T_{\max}$ 。声速 V_s 可以近似为温度 $t[^\circ C]$ 的函数。假设通过直线探测器 304 探测到的直线 270 具有霍夫倾角 θ 。由于直线 270 向右倾斜，所以 θ 为负值。如果 $y=k$ （频率 f_k ），则可以通过 k 和 θ 的函数 $k \cdot \tan(-\theta)$ 来计算由直线 270 表示的相位差 ΔPh 。在这种情况下， $\Delta T [sec]$ 为通过将频率 f_k 的周期 $(1/f_k) [sec]$ 乘以相位差 $\Delta Ph(\theta, k)$ 与 2π 的比率而获得的时间。由于 θ 为带正负号的量，所以 ΔT 也为带正负号的量。即，当声音从图 21D 中的右侧入射时（当相位差 ΔPh 具有正值时）， θ 具有负值。同样，当声音从图 21D 中的左侧入射时（当相位差 ΔPh 具有负值时）， θ 具有正值。因此， θ 的正负号被颠倒。值得注意的是，实际计算仅需要通过利用 $k=1$ （刚刚高于 DC 成分 $k=0$ 的频率）来执行。

[声源成分估测器 312]

声源成分估测器 312 是这样一种装置，用于估测由坐标值确定器 302 给定的每一频率成分的 (x, y) 坐标值与由直线探测器 304 所探测的直线之间的距离，由此探测位于直线附近的点（即频率成分）作为该直线（即声源）的频率成分，并根据该探测结果来估测每一声源的频率成分。

[通过距离阈值方法的探测]

图 23A 至 23C 示意性示出在存在多个声源时声源成分估测的原理。图 23A 是与图 9 相同的频率与相位差的示图，并示出了在相对于麦克风对的不同方向上存在两个声源的情况。在图 23A 中，参考数字 280 表示一个直线组；而 281 和 282 表示另一直线组。图 23A 中的实心圆表示单个频率成分的相位差位置。

如图 23B 中所示，与直线组 280 相应的声源的频率成分被探测为位于夹在直线 284 与 285 之间的区 286 中的频率成分（实心圆），直线 284 与 285 分别从直线 280 向左和向右偏离水平距离 283。当特定频率成分被探测为特定直线的成分时，在下面的说明中将使用这种表达，即，该频率成分归属于（或属于）该直线。

相似地，如图 23C 中所示，与直线组 281 相应的声源的频率成分被探测为位于夹在从左和从右面与直线 281 分离水平距离 283 的直线之间的区 287 中的频率成分（实心圆），而与直线组 282 相应的声源的频率成分被探测为位于夹在从左和从右面与直线 282 分离水平距离 283 的直线之间的区 288 中的频率成分（实心圆）。

值得注意的是，两个点，即频率成分 289 和原点（DC 成分）包含在区 286 和 288 中，所以它们被两次探测为这两个声源的成分（多重归属）。下文将该方法称为“距离阈值方法”，即，通过对各频率成分与各直线之间的水平距离执行阈值处理来为每一直线组（声源）选择存在于阈值范围内的频率成分，并直接使用所获得的功率和相位作为该源声的成分。

[通过最近邻法的探测]

图 24 是示出处理结果的示图，通过该处理，允许图 23B 中示出的多重归属的频率成分 289 仅归属到最近的直线组。当比较频率成分 289 离直线 280 和 282 的水平距离时，发现频率成分 289 最接近直

线 282。在这种情况下，频率成分 289 包含在接近直线 282 的区 288 中。因此，如图 24B 中所示，频率成分 289 被探测为属于直线组（281 和 282）的成分。对于每一频率成分选择具有最短水平距离的直线（声源）、且如果该水平距离存在于预定阈值范围内，则直接使用该频率成分的功率和相位作为该源声的成分的这种方法在下文中被称为“最近邻法”。值得注意的是，作为例外，允许 DC 成分（原点）归属到这两个直线组（声源）。

[通过距离系数方法的探测]

在上述两种方法中，仅选择相对于形成直线组的直线在预定水平距离阈值范围内的频率成分，并且直接将所选择的频率成分的功率和相位用作与该直线组相应的源声的频率成分。另一方面，在下述的“距离系数方法”中，计算随着频率成分与直线之间的水平距离 d 的增加而单调下降的非负系数 α ，并将该频率成分的功率乘以该非负系数 α 。因此，一个成分离一个直线的水平距离越长，该成分贡献于该源声的功率越弱。

在该方法中，不需要利用水平距离执行任何阈值处理。即，获得每一频率成分相对于特定直线组的水平距离 d （到直线组中最近直线的水平距离），并将通过将该频率成分的功率乘以根据水平距离 d 确定的系数 α 而计算得到的值用作该直线组中的该频率成分的功率。用于计算随着水平距离 d 增加而单调下降的非负系数 α 的表达式可以为任意表达式。一个实例为在图 25 中示出的 S 形（S 形曲线）函数 $\alpha = \exp(-(B \cdot d)^C)$ 。如图 25 中所示，如果 B 为正值（在图 25 中为 1.5）而 C 为大于 1 的值（在图 25 中为 2.0），则当 $d=0$ 时 $\alpha=1$ ，而当 $d \rightarrow \infty$ 时 $\alpha \rightarrow 0$ 。如果非负系数 α 中的下降陡峭，即，如果 B 很大，则可以容易地排除直线组外侧的成分，而这加强了声源方向的方向性。另一方面，如果非负系数 α 的下降缓和，即如果 B 小，则方

向性降低。

[多个 FFT 结果的处理]

如上面已经描述过的，投票单元 303 可以对每一个 FFT 执行投票，且还可以对 m 个 ($m \geq 1$) 连续的 FFT 结果共同投票。因此，在直线探测器 304 之后的处理霍夫投票结果的那些功能块，在每一个执行一次霍夫变换的周期内运行。如果在 $m \geq 2$ 的条件下执行霍夫投票，则将在多个时间点的 FFT 结果分类为每一源声的成分，所以不同时刻的相同频率成分可能归属到不同的源声。为了防止这一情形，不论 m 值如何，坐标值确定器 302 向每一个频率成分（即，图 24 中示出的实心圆）给定帧起始时间作为获取时间信息，在该帧中获取了该频率成分。这使得能够查阅在哪一时刻的哪一频率成分归属到哪一声源。即，将源声作为其频率成分的时间序列数据来单独提取。

[功率保存选项]

在上述每一种方法中，对于属于多个 (N) 直线组（声源）的频率成分（在最近邻法中只有 DC 成分，而在距离系数方法中的所有频率成分），还可以将在相同时刻下被分布到单独声源的这些频率成分的功率归一化并将其划分成 N 个部分，以便于这些功率的总和等于在该分布之前的相同时刻下的功率值 $P_o(fk)$ 。以这种方式，可以将整个声源的总功率保持为与在相同时刻下的单独频率成分的输入相同。这被称为“功率保存选项”。该分布方法具有下述两种方式：

- (1) 分割成 N 等分（可适用于距离阈值方法和最近邻法），和
- (2) 相应于到每一直线组的距离的分布（可适用于距离阈值方法和距离系数方法）。

- (1) 是通过分割成 N 等分来自动达到归一化的分布方法。方法
- (1) 可适用于距离阈值方法和最近邻法，这两种方法中的每一种不

考虑距离而确定分布。

(2) 是这样一种分布方法，其通过与距离系数方法中相同的方式确定系数，然后将这些系数归一化以使系数的总和等于 1，从而保存总功率。方法 (2) 可适用于距离阈值方法和距离系数方法，在这两种方法中的每一种中，除原点之外发生多重归属。

值得注意的是，声源成分估测器 312 可以根据设置来执行距离阈值方法、最近邻法和距离系数方法中的任何一种。还能够在距离阈值方法和最近邻法中选择上述功率保存选项。

[声源再合成器 313]

声源再合成器 313 在相同的获取时间对形成每一源声的频率成分执行反 FFT，由此在帧的间隔中再合成源声（振幅数据），其起始时间为所述获取时间。如图 3 所示，一个帧与下一帧重叠，其时间差与它们之间的帧移动量相对应。在其中多个帧这样彼此重叠的间隔内，可以将所有重叠帧的振幅数据平均为最终的振幅数据。通过该处理，可以分离地提取源声作为其振幅数据。

[时间序列跟踪单元 314]

如上所述，每当投票单元 303 执行霍夫投票，直线探测器 304 都获得直线组。对 m ($m \geq 1$) 个连续 FFT 结果执行一次霍夫投票。结果，在 m 个帧的时间的周期内（下文称之为“图形探测周期”）按照时间序列方式获得直线组。同样，与通过方向估测器 305 计算的声源方向 ϕ 一一对应地获得直线组的 θ 。因此，无论声源静止或移动，与稳定声源相应的 θ （或 ϕ ）的时间轴上的轨迹可推定为是连续的。另一方面，取决于阈值的设定，由直线探测器 304 探测的直线组有时包括与背景噪音相应的直线组（下文中被称之为“噪音直线组”）。然而，该噪音直线组的 θ （或 ϕ ）的时间轴上的轨迹被认为是不连续的，或

者即使连续也很短。

时间序列跟踪单元 314 是用于将在每一个图形探测周期中这样获得的 ϕ 分割成在时间轴上连续的多个组, 由此获得 ϕ 在时间轴上的轨迹。下面参考图 26 将阐述分割成多个组的该方法。

(1) 准备轨迹数据缓冲器。该轨迹数据缓冲器为一个轨迹数据阵列。一个轨迹数据 K_d 可以保存起始时间 T_s 、结束时间 T_e 、形成该轨迹的直线组数据 L_d 的阵列(直线组列表)和标签号 L_n 。一个直线组数据 L_d 为包含形成该轨迹的一个直线组的 θ 值和 ρ 值(由直线探测器 304 获得)、表示相应于该直线组的声源方向的 ϕ 值(由方向估测器 311 获得)、相应于该直线组的频率成分(通过声源成分估测器 312 获得)、以及这些频率成分的获取时间的数据组。值得注意的是, 轨迹数据缓冲器初始为空。还值得注意的是, 准备新的标签号作为用于发行标签号的参数, 且将该新的标签号的初始值设定为 0。

(2) 在特定时刻 T , 对于每一个新获得的 ϕ (下文中称之为 ϕ_n ; 在图 26 中, 获得由实心圆 303 和 304 表示的两个 ϕ_n), 参考保持在轨迹数据缓冲器中的轨迹数据 K_d (图 26 中的矩形 301 或 302) 的直线组数据 L_d (图 26 中的每一矩形中布置的实心圆), 由此探测具有 L_d 的轨迹数据, 其中 ϕ 值与 ϕ_n 之间的差(图 26 中的 305 或 306) 等于或小于预定的角阈值 $\Delta\phi$, 而获取时间之间的差(图 26 中的 307 或 308) 等于或小于预定的时间阈值 Δt 。结果, 对于实心圆 303 探测到轨迹数据 301。然而, 对于实心圆 304, 即使最接近的轨迹数据 302 也不满足上述条件。

(3) 如果像实心圆 303 的情况那样发现满足条件(2)的轨迹数据, 则确定 ϕ_n 形成与该轨迹相同的轨迹, 所以将该 ϕ_n 和相应于 ϕ_n 的 θ 值、 ρ 值、频率成分以及当前时间(present time) T 作为轨迹 K_d 的新直线组数据添加到直线组列表, 并将当前时间 T 设置为该轨迹的新的结束时间 T_e 。如果发现多个轨迹, 则确定所有这些轨迹形

成相同轨迹，所以这些轨迹集成为具有最小标签号的轨迹数据，并从轨迹数据缓冲器中删去剩余的轨迹。该集成轨迹数据的起始时间 T_s 为在集成之前的单个轨迹数据的最早起始时间，该集成轨迹数据的结束时间 T_e 为在集成之前的单个轨迹数据的最后结束时间，而该直线组列表为在集成之前的单个轨迹数据的直线组列表的并集。结果，将实心圆 303 添加到轨迹数据 301 中。

(4) 如果像实心圆 304 的情况那样没有发现满足条件 (2) 的轨迹数据，则确定新轨迹开始，所以在轨迹数据缓冲器的空区域中形成新的轨迹数据。另外，将起始时间 T_s 和结束时间 T_e 设定在当前时间 T ，将 ϕ_n 和相应于 ϕ_n 的 θ 值、 ϕ 值、频率成分以及当前时间 T 设置为直线组列表中的第一直线组数据，将新标签号的值定为这一轨迹的标签号 L_n ，且将该新标签号增加 1。值得注意的是，如果新标签号达到预定最大值，则其返回到 0。结果，将实心圆 304 作为新的轨迹数据寄存在轨迹数据缓冲器中。

(5) 如果对于保持在轨迹数据缓冲器中的轨迹数据，在最新更新轨迹数据之后（即，在轨迹数据的结束时间 T_e 之后）和在当前时间 T 之前已经流逝上述预定时间 Δt ，则确定该轨迹数据为没有发现对其要添加的新 ϕ_n 的轨迹，即，该轨迹数据为被完全跟踪的轨迹。因此，在下一个阶段被输出到持续时间评估器 315 之后，从轨迹数据缓冲器中删除该轨迹数据。参考图 26，轨迹数据 302 就是这样一个轨迹数据。

[持续时间评估器 315]

根据轨迹数据的起始时间和结束时间，持续时间评估器 315 计算由从时间序列跟踪单元 314 中输出的被完全跟踪的轨迹数据表示的轨迹的持续时间。如果该持续时间超过预定阈值，则持续时间评估器 315 确定该轨迹数据是基于源声；否则，持续时间评估器 315 确定该

轨迹数据是基于噪音的。基于源声的轨迹数据在下文中被称为声源流信息。该声源流信息包含源声的起始时间 T_s 和结束时间 T_e ，以及 θ 、 ρ 和表示声源方向的 ϕ 的时间序列轨迹数据。值得注意的是，通过图形探测器 5 获得的直线组的数量给出了声源的数量，但是该数量包括噪音源。通过持续时间评估器 315 获得的多条声源流信息的数量给出了除基于噪音的那些声源之外的可靠声源的数量。

[相位匹配单元 316]

相位匹配单元 316 参照通过时间序列跟踪单元 314 获得的声源流信息，并获得在声源方向 ϕ 上的该流的时间转换。根据 ϕ 的最大值 $\phi_{\max}$ 和最小值 $\phi_{\min}$ ，相位匹配单元 316 计算中间值 $\phi_{\text{mid}} = (\phi_{\max} + \phi_{\min}) / 2$ 以获得宽度 $\phi_w = \phi_{\max} - \phi_{\text{mid}}$ 。然后，相位匹配单元 316 从比该流的起始时间 T_s 早预定时间的的时间到比结束时间 T_e 晚预定时间的的时间中提取两个频率分解数据组 a 和 b 的时间序列数据，作为该声源流信息的基础。相位匹配单元 316 通过校正这些时间序列数据的相位来将它们匹配，以便于消除通过利用中间值 ϕ_{mid} 的反操作所计算的到时差。

通过利用由方向估测器 311 获得的在每一时刻下的声源方向 ϕ 作为 ϕ_{mid} ，还能够总是将两个频率分解数据的时间序列数据的相位匹配。通过操作模式来确定是参照声源流信息还是参照每一时刻的 ϕ ，并且该操作模式可以作为参数被设置和改变。

[自适应阵列处理器 317]

自适应阵列处理将其中心方向性指向前方 0° ，且具有通过向 $\pm \phi_w$ 添加预定余量而获得的值以作为跟踪范围。自适应阵列处理器 317 对那些被提取的并且相互同相位的两个频率分解数据组 a 和 b 的时间序列数据执行这一自适应阵列处理，由此精确地分离和提取这一

流的源声的频率成分的时间序列数据。虽然所述各方法是不同的，但是该处理以与声源成分估测器 312 相同的方式运行，即，其中频率成分的时间序列数据被分离地提取。因此，源声再合成器 313 也可以从通过自适应阵列处理器 317 获得的源声的频率成分的时间序列数据中再合成源声的振幅数据。

值得注意的是，作为自适应阵列处理，通过使用被称为成束器（beam former）形成方法的“Griffith-Jim type generalized side lob canceller”作为两个消除器（主和次消除器）中的每一个，可以采用在设定的方向性范围内清晰地分离和提取声音的方法，如参考文件 3 “2004 年的 Toshiba Review 2004, Vol.59, No.9 中的 Tadashi Amada 等人的“Microphone Array Technique for Voice Recognition””中所述。

自适应阵列处理通常用于接收仅在预设跟踪范围的方向上的声音。因此，需要准备大量的具有不同跟踪范围的自适应阵列，以便于接收所有方向上的声音。然而，在该实施例，在实际获得声源的数量和方向后，仅可以操作数量与声源相同的自适应阵列。因为还可以将跟踪范围设置在与声源方向相应的预定狭窄范围内，所以可以高质量地有效地分离和提取数据。

同样，由于两个频率分解数据组 a 和 b 的时间序列数据的相位预先被匹配，所以可以仅通过将自适应阵列的跟踪范围设置在前方附近来处理所有方向上的声音。

[语音识别单元 318]

语音识别单元 318 分析并校对由声源成分估测器 312 或自适应阵列处理器 317 所提取的源声的频率成分的时间序列数据，由此提取该流的符号内容，即提取表示语言含义、声源类型或讲话者身份的符号（序列）。

值得注意的是，如果需要，从方向估测器 311 到语音识别单元

318 的功能块可以通过图 20 中未示出的连接来交换信息。

[输出单元 7]

输出单元 7 是用于输出作为通过声源信息发生器 6 获得的声源信息的信息的装置，该输出的信息包含以下内容中的至少一种，所述内容包括通过图形探测器 5 获得的作为直线组数量的声源数量、由方向估测器 311 估测的作为声音信号产生源的每一个声源的空间存在范围（确定圆锥的角 ϕ ）、由声源成分估测器 312 估测的由每一声源产生的声音的成分（每一频率成分的功率和相位的时间序列数据）、由源声再合成器 313 合成的对于每一声源分离的分离声音（振幅值的时间序列数据）、根据时间序列跟踪单元 314 和持续时间评估器 315 来确定的除噪音源之外的声源数量、通过时间序列跟踪单元 314 和持续时间评估器 315 确定的由每一声源产生的声音的时间存在期间、通过相位匹配单元 316 和自适应阵列处理器 317 获得的每一声源的分离声音（振幅值的时间序列数据）、以及通过语音识别单元 318 获得的每一源声的那些符号内容。

[用户界面单元 8]

用户界面单元 8 是用于向用户呈现上述声音信号处理所需要的各种设置内容、接受由用户输入的设置、将设置内容保存在外部存储器件中、从该外部存储器件中读出设置内容、以及通过使它们可视化来向用户呈现各种处理结果和中间结果的装置。例如，用户界面单元 8（1）显示每一麦克风的频率成分，（2）显示相位差（或者时间差）示图（即，显示二维数据），（3）显示各种投票分布，（4）显示峰值位置，和（5）在如图 17 或 19 中所示的示图上显示直线组，（6）如图 23 或 24 所示，显示归属到直线组的频率成分，和（7）如图 26 中所示，显示轨迹数据。用户界面单元 8 还是用于允许用户选择期望

的数据并使所选择的数据详细地可视化的装置。用户界面单元 8 允许用户,例如,检验根据该实施例的声音信号处理设备的操作、调节设备以能够执行期望的操作,并随后在该调节状态下使用设备。

[处理的流程图]

图 27 是示出根据本实施例的声音信号处理设备执行的处理的流程的流程图。该处理包括初始化步骤 S1、声音信号输入步骤 S2、频率分解步骤 S3、二维数据形成步骤 S4、图形探测步骤 S5、声源信息产生步骤 S6、输出步骤 S7、终止确定步骤 S8、确认确定步骤 S9、信息呈现/设置接收步骤 S10 以及终止步骤 S11。

初始化步骤 S1 是执行上述用户界面单元 8 的部分处理的处理步骤。在该步骤中,从外部存储器件中读取声音信号处理所需的各种设置内容,以将设备初始化为预定设置状态。

声音信号输入步骤 S2 是执行上述声音信号输入单元 2 的处理的处理步骤。在该步骤中,输入在空间上不同的两个位置处拾取的两个声音信号。

频率分解步骤 S3 是执行上述频率分解器 3 的处理的处理步骤。在该步骤中,在声音信号输入步骤 S2 中输入的每一个声音信号被分解为频率成分,并至少计算每一频率的相位值(和功率值,如果需要)。

二维数据形成步骤 S4 是执行上述二维数据形成单元 4 的处理的处理步骤。在该步骤中,将在频率分解步骤 S3 中计算的输入声音信号的单个频率的那些相位值进行比较,以计算所述两个信号的每个频率的相位差值。将每个频率的该相位差值转换成由所述频率和其相位差唯一确定的 (x, y) 坐标值作为 X-Y 坐标系上的一个点,在该 X-Y 坐标系中,频率的函数为 Y 轴而相位差值的函数为 X 轴。

图形探测步骤 S5 是执行上述图形探测器 5 的处理的处理步骤。在该步骤中,从在二维数据形成步骤 S4 中形成的二维数据中探测预

定图形。

声源信号产生步骤 S6 是执行上述声源信息发生器 6 的处理的处理步骤。在该步骤中, 根据在图形探测步骤 S5 中探测到的图形的信息来产生声源信息。该声源信息包含作为声音信号的产生源的声源的数量、每一声源的空间存在范围、每一声源产生的声音的成分、每一声源的分离声音、由每一声源产生的声音的时间存在期间、以及由每一声源产生的声音的符号内容中的至少一种。

输出步骤 S7 是执行上述输出单元 7 的处理的处理步骤。在该步骤中, 输出在声源信息产生步骤 S6 中产生的声源信息。

终止确定步骤 S8 是执行用户界面单元 8 的部分处理的处理步骤。在该步骤中, 检验来自用户的终止命令的存在/不存在。如果存在终止命令, 则流程前进到终止步骤 S11 (向左的分支)。如果不存在终止命令, 则流程前进到确认确定步骤 S9 (向上的分支)。

确认确定步骤 S9 是执行用户界面单元 8 的部分处理的处理步骤。在该步骤中, 检验来自用户的确认命令的存在/不存在。如果存在确认命令, 则流程前进到信息呈现/设定接收步骤 S10 (向左的分支)。如果没有确认命令存在, 则流程返回到声音信号输入步骤 S2 (向上的分支)。

信息呈现/设定接收步骤 S10 是响应于来自用户的确认命令来执行用户界面单元 8 的部分处理的处理步骤。在该步骤中, 向用户呈现声音信号处理所需的各种设置内容, 接收用户输入的设置, 通过保存命令将设置内容保存在外部存储器件中, 通过读取命令从外部存储器件中读取设置内容、使各种处理结果和中间结果可视化并将其呈现给用户, 并且由用户选择期望的数据并使其详细地可视化。如此, 用户可以检查声音信号处理的操作、调整处理以能够执行期望的操作、并随后在调节后的状态下继续处理。

终止步骤 S11 是响应于来自用户的终止命令来执行用户界面单

元 8 的部分处理的处理步骤。在该步骤中，将声音信号处理所需的各种设置内容自动保存在外部存储器件中。

[变型]下面将阐述上述实施例的变型。

[垂直线的探测]

如图 7 中所示，二维数据形成单元 4 的坐标值确定器 302 通过利用作为相位差 $\Delta \text{Ph}(fk)$ 的 X 坐标值并利用作为频率成分号 k 的 Y 坐标值来产生点。X 坐标值还可以是根据相位差 $\Delta \text{Ph}(fk)$ 对每一频率计算的到时差的估测值 $\Delta T(fk) = (\Delta \text{Ph}(fk) / 2\pi) \times (1/fk)$ 。当使用到时差来代替相位差时，具有相同到时的点，即来自相同声源的点被布置在垂直直线上。

在该情况下，频率越高，可以通过 $\Delta \text{Ph}(fk)$ 表达的时间差 $\Delta T(fk)$ 越小。如图 28A 中示意性地示出，令字母 T 为由具有频率 fk 的波 290 的一个周期表示的时间，可以由具有两倍频率 $2fk$ 的波 291 的一个周期表示的时间为 $T/2$ 。当如图 28A 中所示将时间差绘制在 X 轴上时，时间差的范围为 $\pm T_{\max}$ ，而在该范围外没有观察到时间差。在等于或低于阈值频率 292 的低频处，从相位差 $\Delta \text{Ph}(fk)$ 中唯一获得到时差 $\Delta T(fk)$ ，其中在该阈值频率 292， $T_{\max}$ 为 $1/2$ 周期（即， π ）或更小。然而，在超出阈值频率 292 的高频下，所计算的 $\Delta T(fk)$ 小于理论上可能的 $T_{\max}$ 。因此，如图 28B 中所示，仅可以表达直线 293 与 294 之间的范围。这是与先前描述的相位差循环问题相同的问题。

因此，为了解决该相位差循环问题，如图 29 中示意性示出，对于超出阈值频率 292 的频率区，坐标值确定器 302 在 $\pm T_{\max}$ 范围内通过相对一个 $\Delta P(fk)$ 添加或减去例如 2π 、 4π 或 6π ，来在对应于相位差的 ΔT 的位置中产生冗余点，由此形成二维数据。所产生的点为图 29 中示出的实心圆。在超出阈值频率 292 的频率区中，对于一个频率绘制多个实心圆。

在这种情况下,根据作为相对于一个相位差值的一个或多个点而产生的二维数据,投票单元 303 和直线探测器 304 可以通过霍夫投票来探测大体上穿过许多点的垂直线(图 29 中的 295)。由于该垂直线为在霍夫投票空间中 $\theta = 0$ 的直线,通过在霍夫投票之后的投票分布中探测投票数量等于或大于 $\theta = 0$ 的 ρ 轴上的预定阈值的峰值位置,可以解决探测垂直线的问题。所探测的峰值位置的 ρ 值给出了垂直线与 X 轴的交点,即到声源 ΔT 的估测值。值得注意的是,在对投票单元 303 的解释中描述的投票条件和加法方法可以直接用于投票中。还值得注意的是,与声源相应的直线不是直线组而是单个垂直线。

还可以通过在一维投票分布(在 Y 轴方向上投影投票的外围分布)中探测投票数量等于或大于预定阈值的峰值位置,来解决获取峰值位置的这个问题,其中,在该一维投票分布中上述冗余点的 X 坐标值被投票。当由此使用到时差来替换相位差作为 X 轴时,所有表示存在于不同方向上的声源的根据(evidence)被投影在具有相同倾角的直线上(即垂直线上)。这使得可以通过外围分布而不利用任何霍夫变换而容易地进行探测。

通过垂直线获得的声源方向的信息是作为 ρ 而非 θ 所获得的到时差 ΔT 。因此,方向估测器 311 可从 ΔT 立即计算出声源方向 ϕ ,而不使用 θ 。

如上所述,由二维数据形成单元 4 形成的二维数据不限于一种类型,且图形探测器 5 的图形探测方法也不限于一种类型。值得注意的是,图 29 中示出的利用到时差的点的绘图和所探测的垂直线也是要通过用户界面单元 8 呈现给用户的信息。

[多个系统的并行布置]

通过包括两个麦克风的最简单布置来阐述上述实施例。然而,如图 30 中所示,还能够通过利用 N ($N \geq 3$) 个麦克风来布置最多 M

($1 \leq M \leq N C_2$) 个麦克风对。

在图 30 中, 参考数字 11 至 13 表示 N 个麦克风; 20 表示用于输入通过 N 个麦克风获得的 N 个声音信号的装置; 21 表示用于分解所输入的 N 个声音信号的频率的装置; 22 为用于对 N 个声音信号的 M ($1 \leq M \leq N C_2$) 个对中的每一个产生二维数据的装置; 23 为用于从所产生的 M 个二维数据对中的每一个中探测预定图形的装置; 24 为用于根据所探测的 M 对图形信息中的每一个来产生声源信息的装置; 25 为用于输出所产生的声源信息的装置; 和 26 为用于向用户呈现包含形成每一对的麦克风信息的各种设定值、接收由用户输入的设置、将设定值保存在外部存储器件中、从外部存储器件中读取设定值以及向用户呈现各种处理结果的装置。对于每一个麦克风对的处理与上述实施例中的相同, 且对于多个麦克风对并行执行该处理。

在该布置中, 虽然一个麦克风对不能覆盖所有方向, 但是通过由多个麦克风对覆盖所有方向, 可以减小不能获得正确声源信息的可能性。

[利用通用计算机的执行: 程序]

如图 31 中所示, 根据本发明的这一实施例还可以作为能够执行用于实施根据本发明的声音信号处理功能的程序的通用计算机来实践。在图 31 中, 参考数字 31 至 33 表示 N 个麦克风; 40 表示用于输入通过 N 个麦克风获得的 N 个声音信号的 A/D 转换装置; 41 表示执行程序命令以处理所输入的 N 个声音信号的 CPU; 而 42 至 47 表示形成计算机的标准装置, 即, RAM 42、ROM 43、HDD 44、鼠标/键盘 45、显示器 46 和 LAN 47。参考数字 50 至 52 表示用于从外部经由存储介质向计算机提供程序和数据的驱动器, 即 CDROM 50、FDD 51 和 CF/SD 卡 52; 48 表示用于输出声音信号的 D/A 转换装置; 而 49 表示连接到 D/A 转换装置 48 的输出端子的扬声器。通过存储

用于执行图 27 中示出的处理步骤的声音信号处理程序、将程序读出到 RAM 42、并通过 CPU 41 执行该程序，该计算机设备用作声音信号处理设备。通过利用作为外部存储器件的 HDD 44、用于接收输入操作的鼠标/键盘 45 和作为信息呈现装置的显示器 46 和扬声器 49，该计算机设备还实施上述用户界面单元 8 的功能。此外，该计算机设备将通过声音信号处理所获得的声源信息保存在 RAM 42、ROM 43 和 HDD 44 中，或通过经由 LAN 47 的通信来输出该信息。

[记录介质]

如图 32 中所示，还能够将本发明作为计算机可读记录介质来实现。在图 32 中，参考数字 61 表示记录根据本发明的声音信号处理程序的、由 CD-ROM、CF 或 SD 卡、或软盘来实现的记录介质。通过将该记录介质 61 插入诸如电视或计算机的电子设备 62 或 63、或插入机器人 64 中来执行该程序。通过通信将该程序从提供了该程序的电子设备 63 输送到另一个电子设备 65 或机器人 64，还可以在所述电子设备 65 或机器人 64 上执行该程序。

[通过温度传感器的声速修正]

还可以通过将用于测量大气温度的温度传感器附着于该设备、并根据温度传感器所测量的温度数据来修正图 22 中示出的声速 V_s ，由此获得精确的 T_{max} ，从而来实现本发明。

或者，可以通过将以预定间隔分隔开的声波发射装置和接收装置附着到该设备、并利用测量装置来测量通过发射装置产生的声波到达接收装置所需要的时间、由此直接计算并修正声速 V_s 且获得精确的 T_{max} ，从而来实现本发明。

[使 θ 的间隔不等以获得相等的 ϕ 间隔]

在本发明中，当执行霍夫变换以获得直线组的倾角时， θ 被量化为例如间隔 1° 。然而，当这样以相等间隔量化 θ 时，可以被估测的声源方向 ϕ 的值以不等的间隔被量化。为了防止这样的情况，本发明还可以如此实施，通过量化 θ 以使得 ϕ 以相等间隔被量化，从而声源方向的估测精确度不轻易改变。

在 2001 年 6 月的 Artificial Intelligence Society AI Challenge Research Meeting, SIG-Challenge-0113-5 的第 35-42 页中的 Kazuhiro Nakadai 等人的“Real-time Active Person Tracking by Hierarchical Integration of Audiovisual Information”中描述的方法通过从频率分解数据中探测形成谐波结构的基频成分和其谐波成分来估测声源的数量、方向和成分。由于假设了谐波结构，所以该方法专用于人的声音。然而，在实际环境中，存在许多不具有谐波结构的声源，例如门打开和关闭的声音。该方法不能处理这种源声。

而且，在 2004 年 4 月的 Measurement and Control, Vol.43, No.4 的第 325-330 页中的 Futoshi Asano 的“Separating Sounds”中描述的方法不限于任何具体模型。然而，只要使用两个麦克风，则可以被处理的声源的数量被限制为 1 个。

另一方面，通过采用霍夫变换将频率成分的相位差分割成独立声源的组，本发明的实施例可以通过利用两个麦克风来实施定位和分离两个或多个声源的功能。由于不使用像谐波结构这样的限制模型，所以本发明可适用于具有各种特性的声源。

下面将概述通过本发明的实施例获得的其他功能和效果。

- 当执行霍夫投票时，通过利用适合于探测具有许多频率成分的声源或大功率的声源的投票方法，可以稳定地探测各种类型的声源。
- 通过施加限制 $\rho = 0$ 并在直线探测期间考虑相位差循环，可以有效且精确地探测声源。
- 通过利用直线探测结果可以获得有用的声源信息，该信息包括

作为声音信号的产生源的声源的空间存在范围、由该声源产生的源声的时间存在期间、源声的成分、源声的分离声音以及源声的符号内容。

- 当要估测单个源声的频率成分时，通过简单地选择直线附近的成分、确定哪一成分归属到哪一直线、并执行与每一直线与成分之间的距离相应的系数乘法运算，可以将这些源声容易地分离。

- 通过预先探测每一声源的方向来自适应地设置自适应阵列处理的方向性范围，可以更加精确地分离声源。

- 通过精确地分离和识别源声，可以确定每一个源声的符号内容。

- 用户可以检查该设备的操作、调节该设备以能够执行期望的操作、然后在调节后的状态下使用该设备。

另外的优点和变型对于本领域技术人员来说是容易想到的。因此，在更广的方面，本发明不限于这里示出和描述的具体细节和代表性实施例。相应地，在不脱离由附属权利要求及其等同物限定的总的发明原理的精神和范围的情况下，可以做出各种变形。

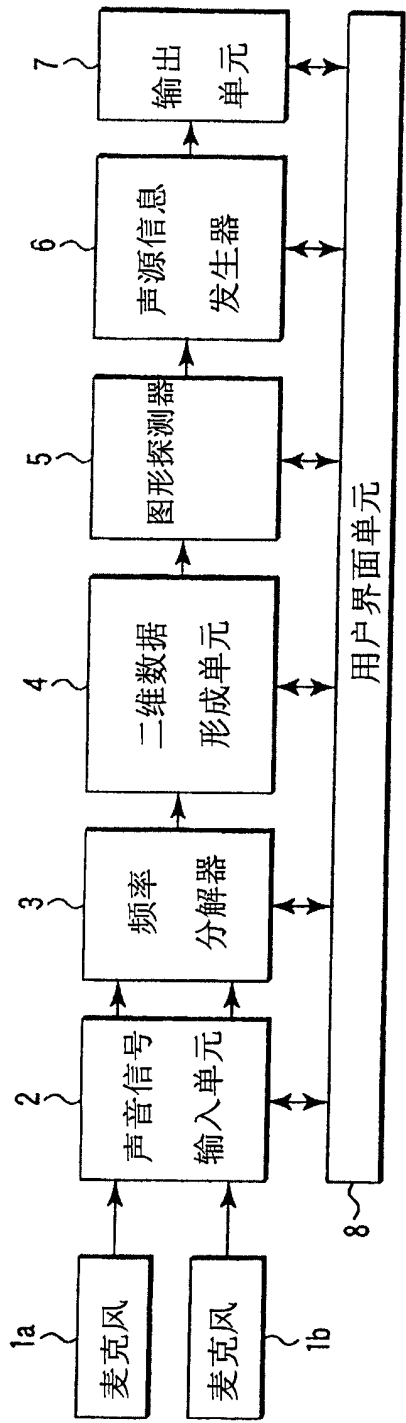


图1

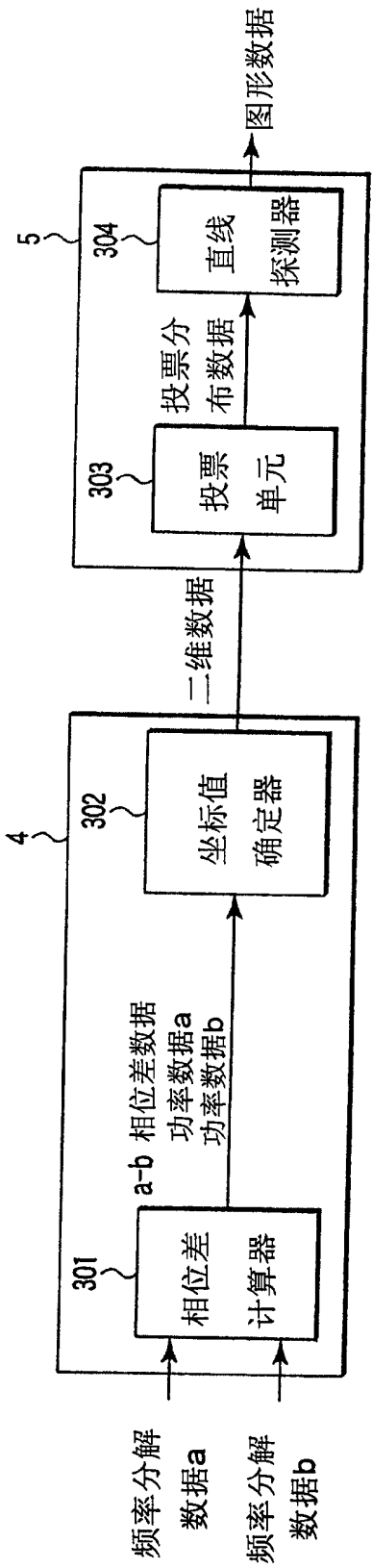


图5

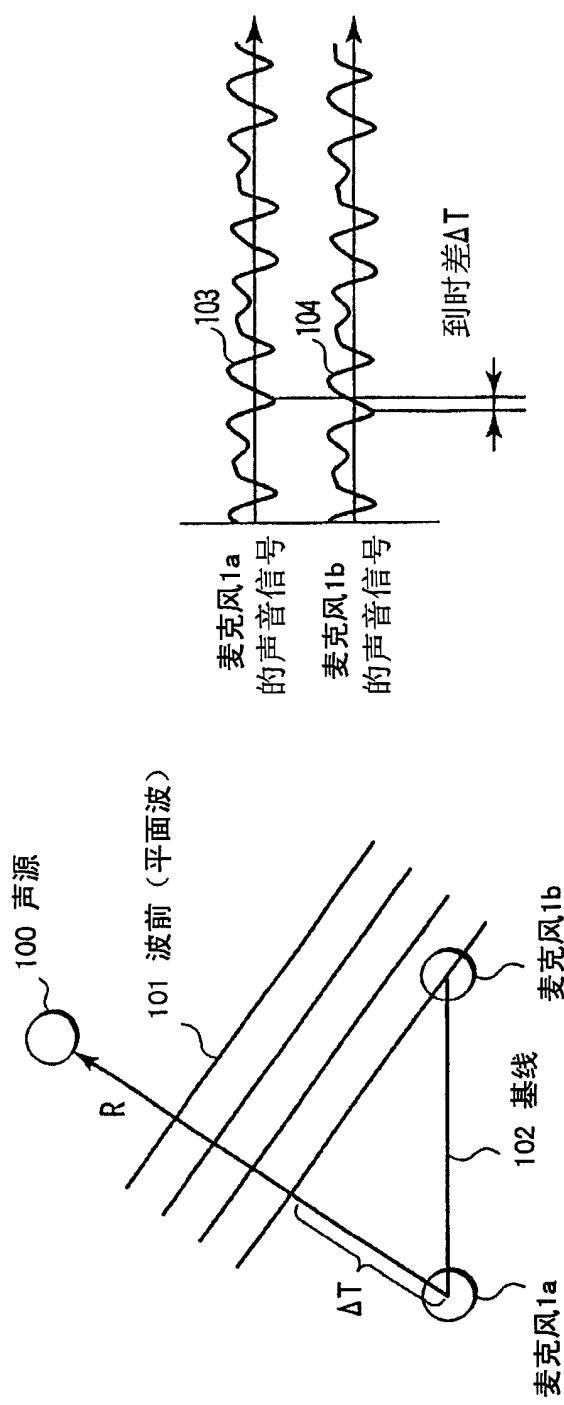


图2A

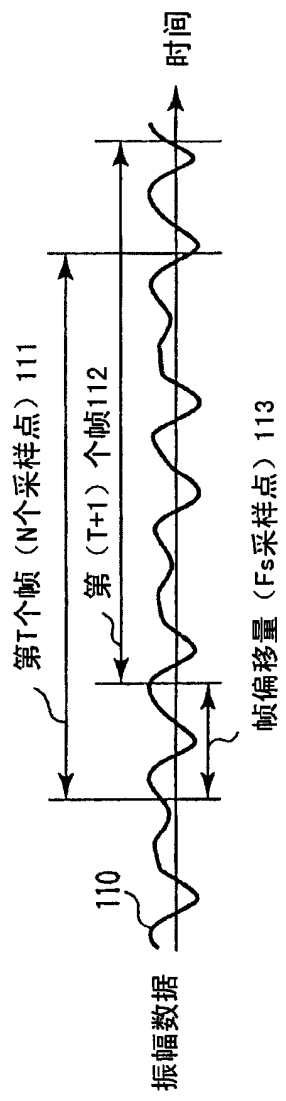
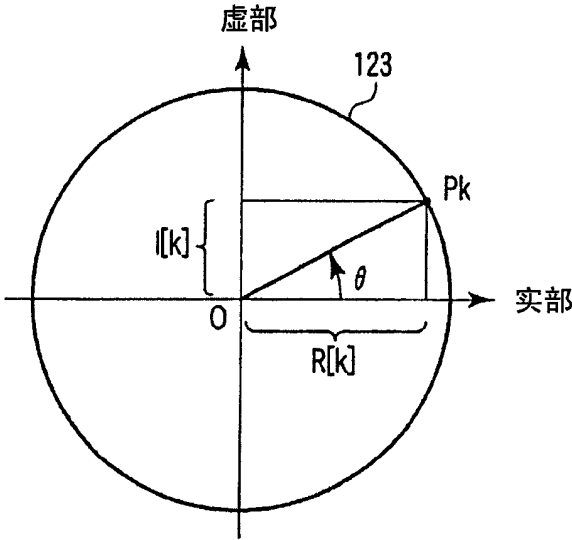
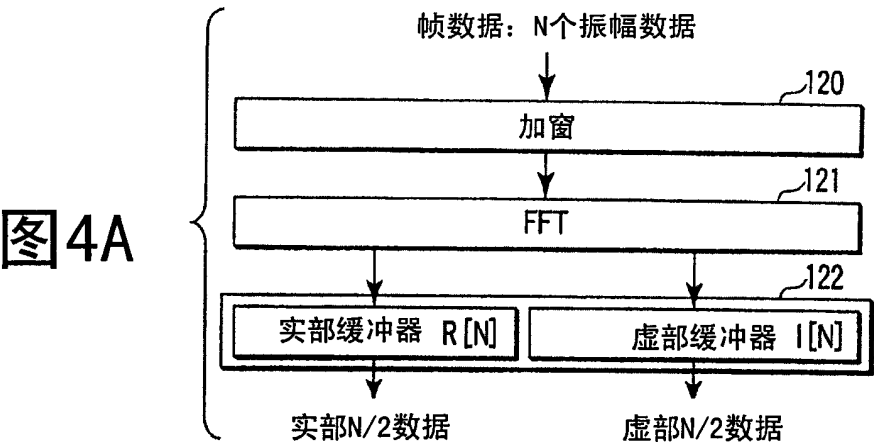


图2B



窗函数

$$w = \cos((2\pi \times n) \div (N-1))$$
$$\text{Hamming}[n] = 0.54 - (0.46 \times w)$$
$$\text{Hanning}[n] = 0.5 - (0.5 \times w)$$

其中n是帧数据中的数据数 {n: 0 ≤ n ≤ N-1}

124

图4C

```
相位差  
 $\Delta Ph(fk) = Ph2(fk) - Ph1(fk);$   
while(1){  
    if( $\Delta Ph(fk) \leq -\pi$ ){ $\Delta Ph(fk) = \Delta Ph(fk) + 2\pi$ ; continue;}  
    break;  
}  
while(1){  
    if( $\Delta Ph(fk) > \pi$ ){ $\Delta Ph(fk) = \Delta Ph(fk) - 2\pi$ ; continue;}  
    break;  
}  
其中
```

$Ph1(fk)$ 是麦克风1a的频率成分 fk 的相位值,
 $Ph2(fk)$ 是麦克风1b的频率成分 fk 的相位值,并且
 $\Delta Ph(fk)$ 的范围是 $\{\Delta Ph(fk): -\pi < \Delta Ph(fk) \leq \pi\}$

图6

```
坐标值  
 $x(fk) = \Delta Ph(fk)$   
 $y(fk) = k$ 
```

图7

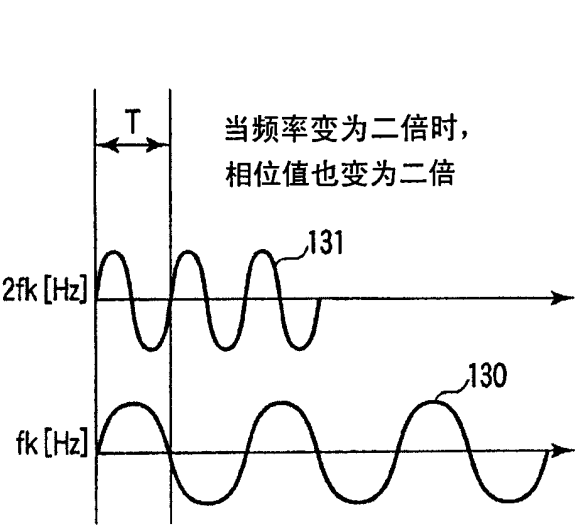


图8A

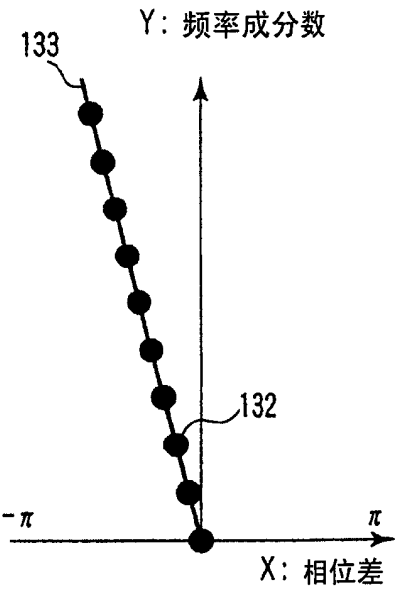


图8B

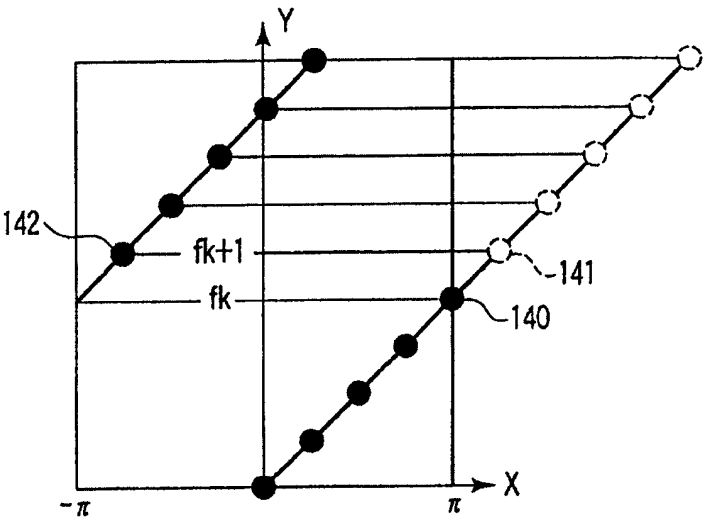


图9

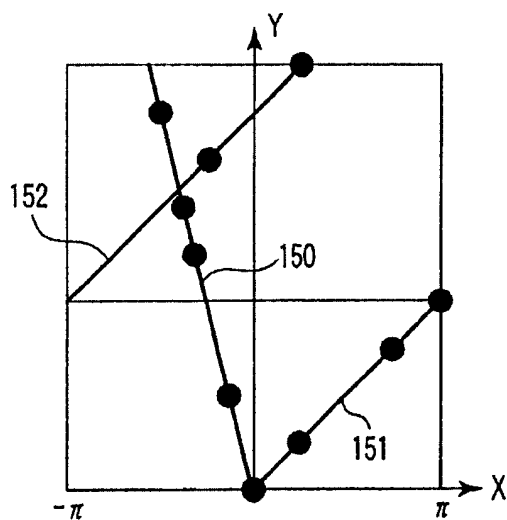


图10A

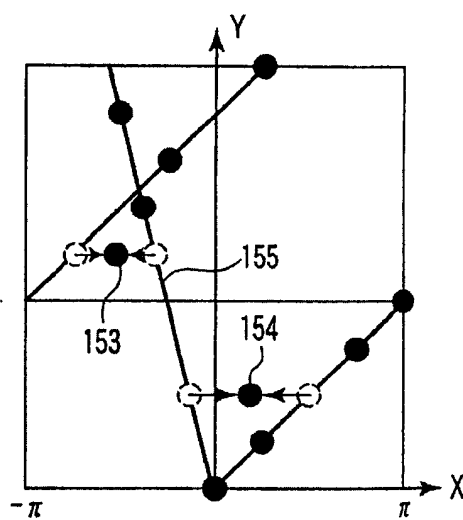


图10B

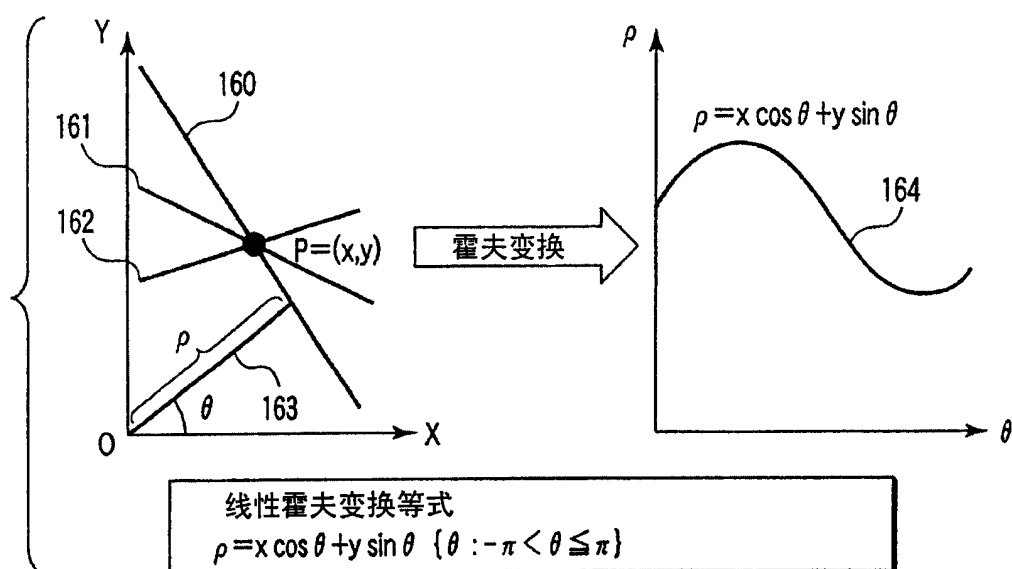


图11

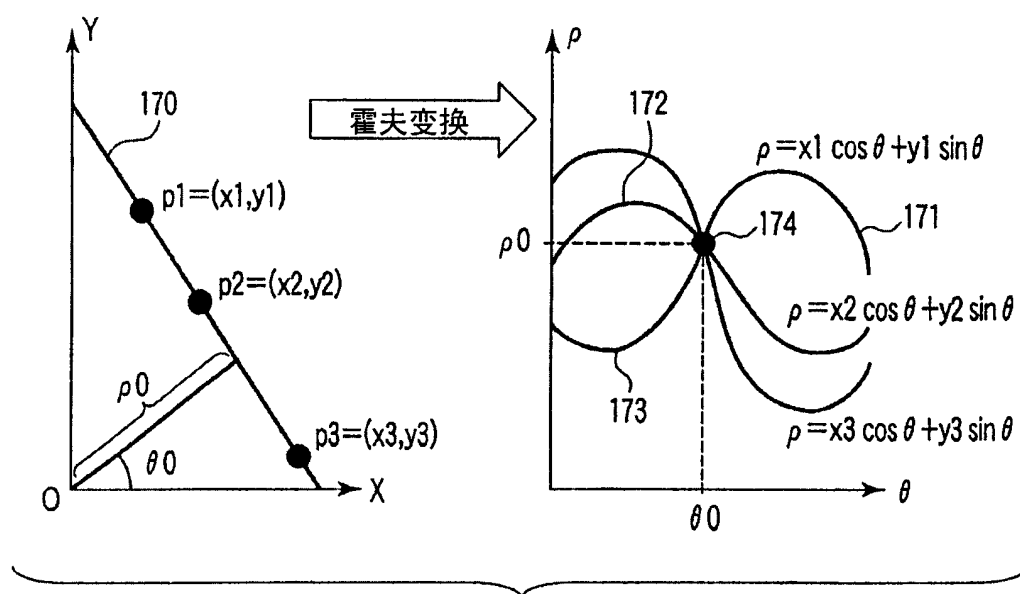


图12

功率函数

$$G(P(fk)) = V + 1 : V > 0$$

$$G(P(fk)) = 1 : V \leq 0$$

其中

$$V = \log_{10}(P(fk)) + \alpha$$

$$P(fk) = (P_{o2}(fk) + P_{o1}(fk)) / 2$$

图13

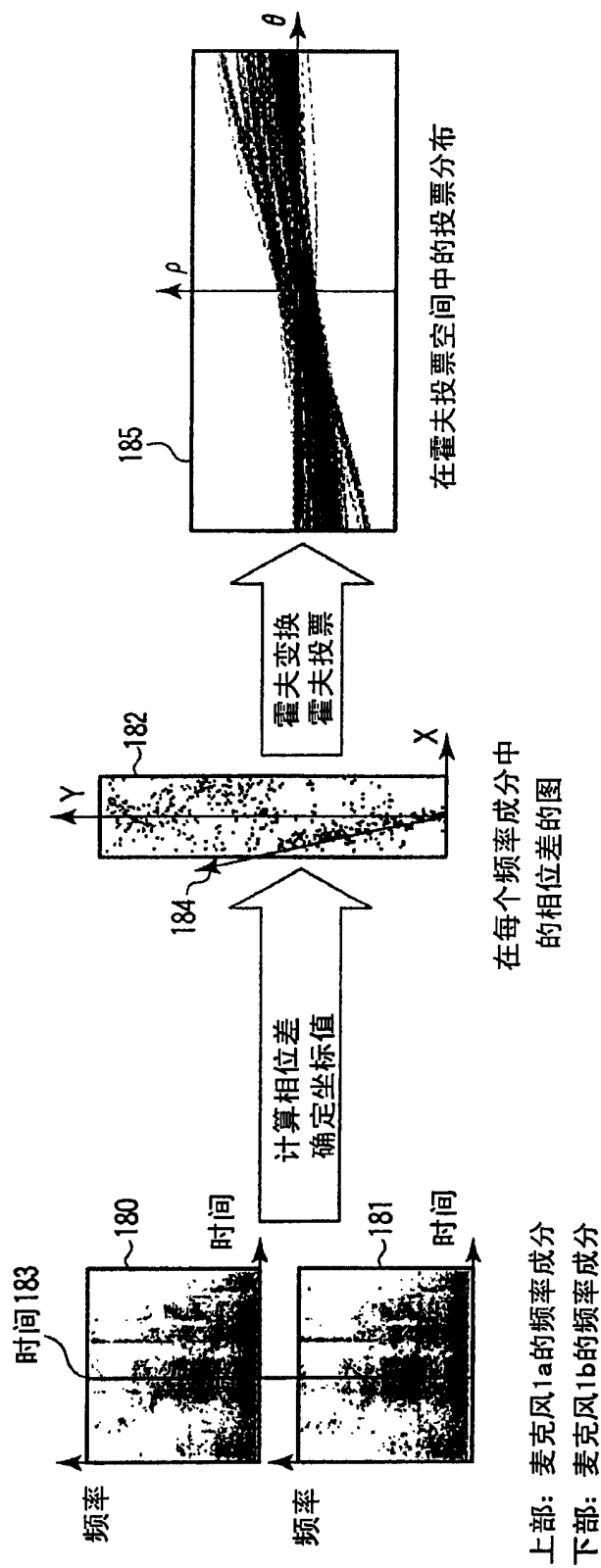


图14

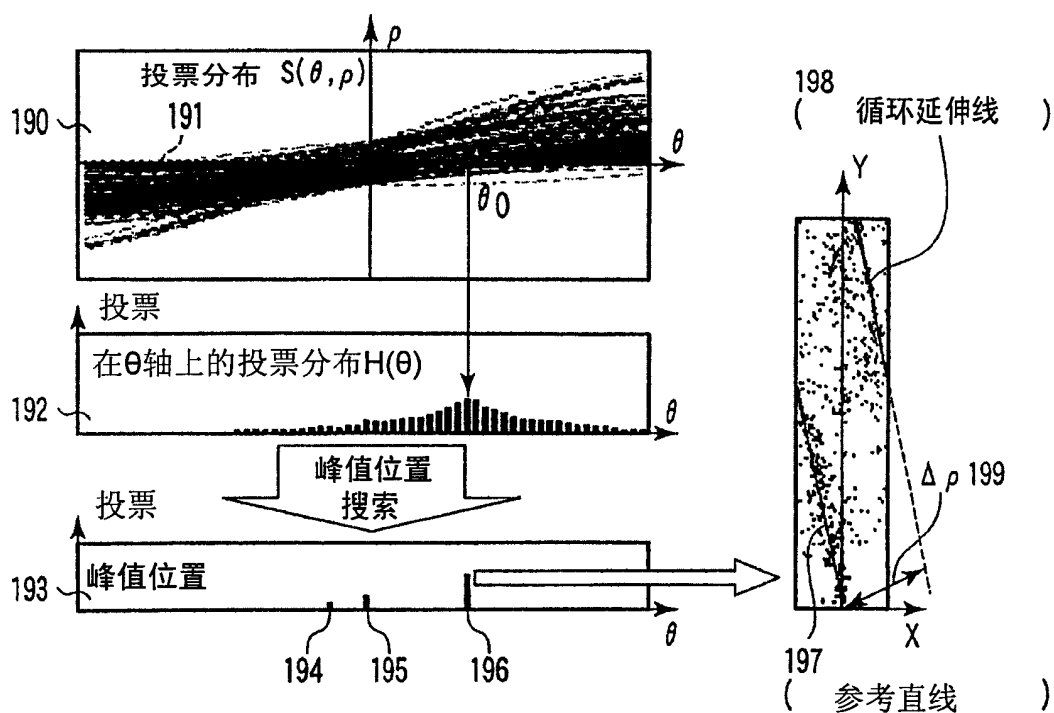


图15

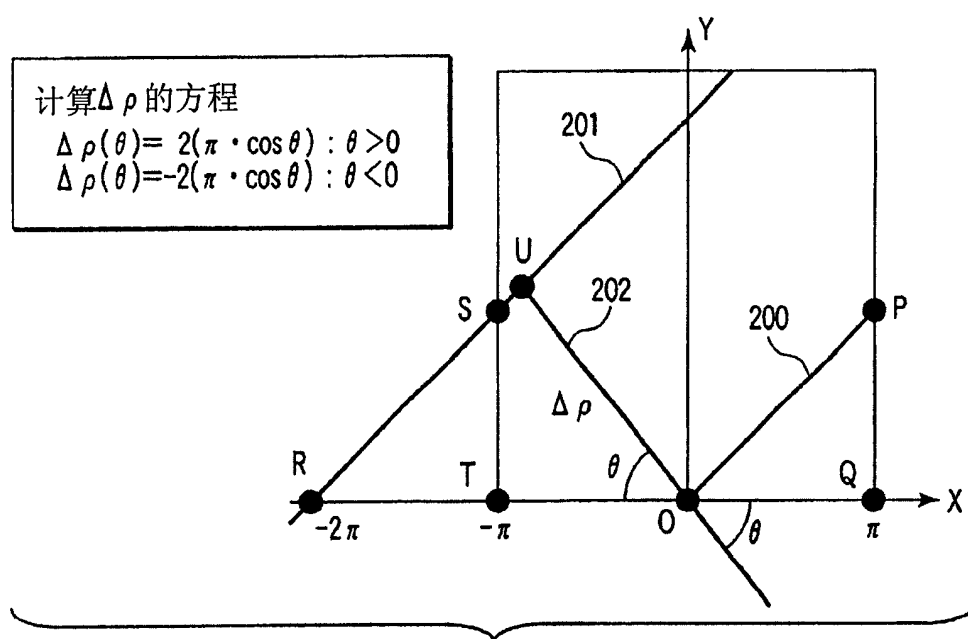


图16

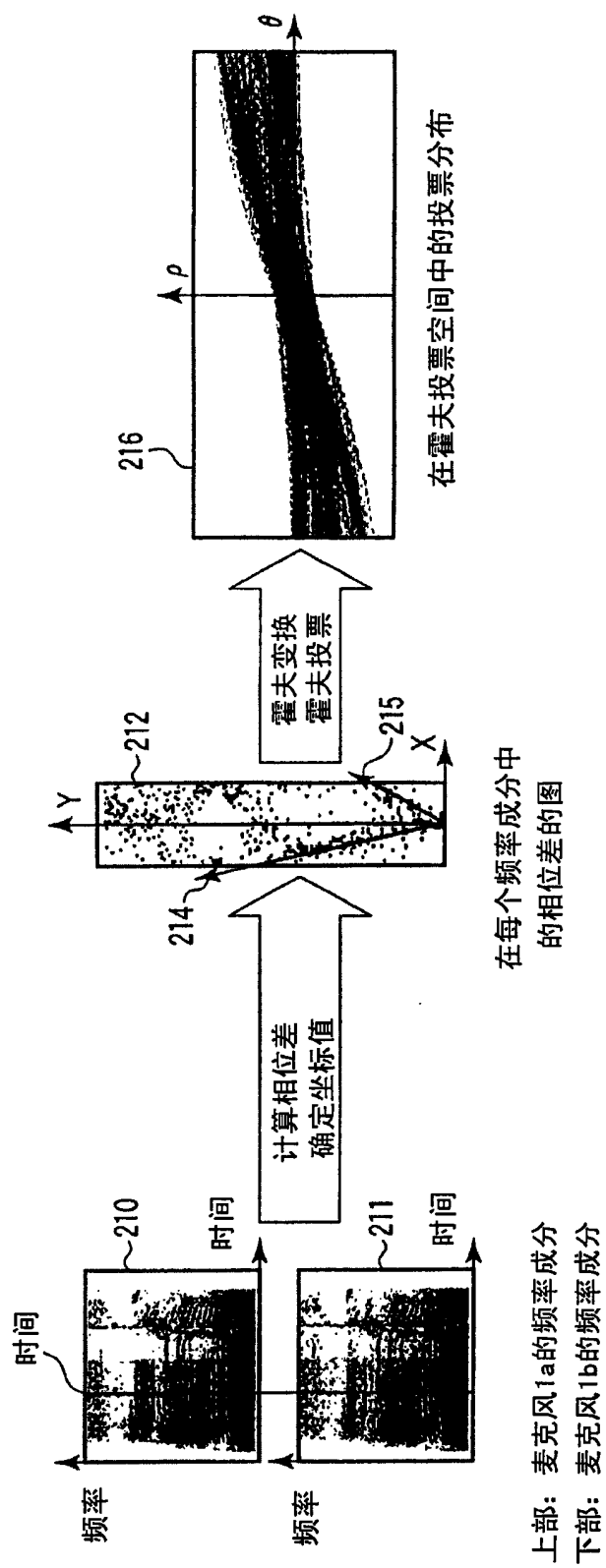


图17

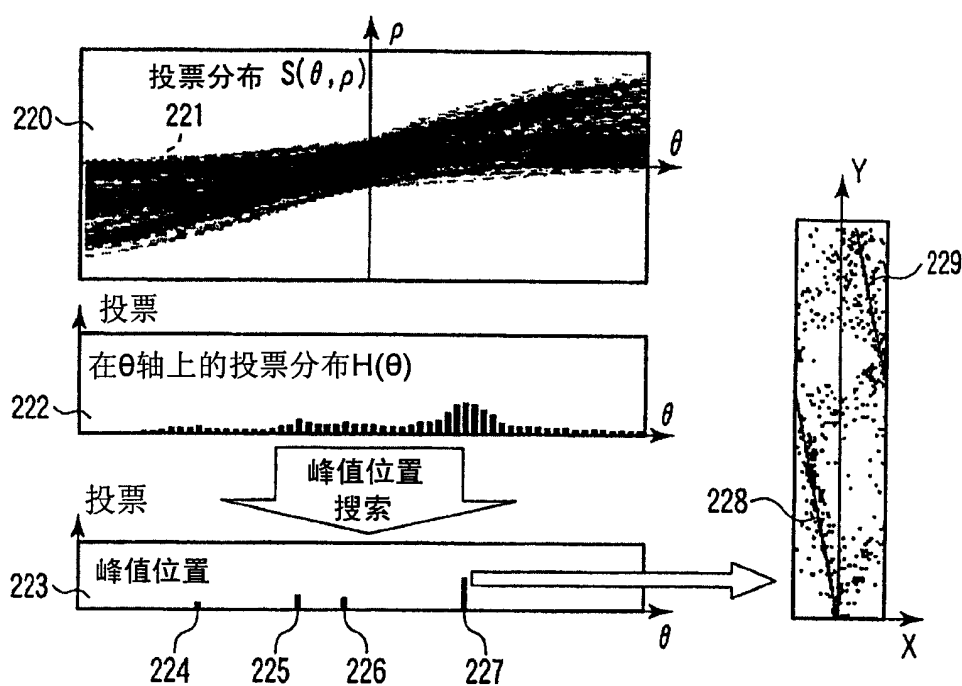


图18

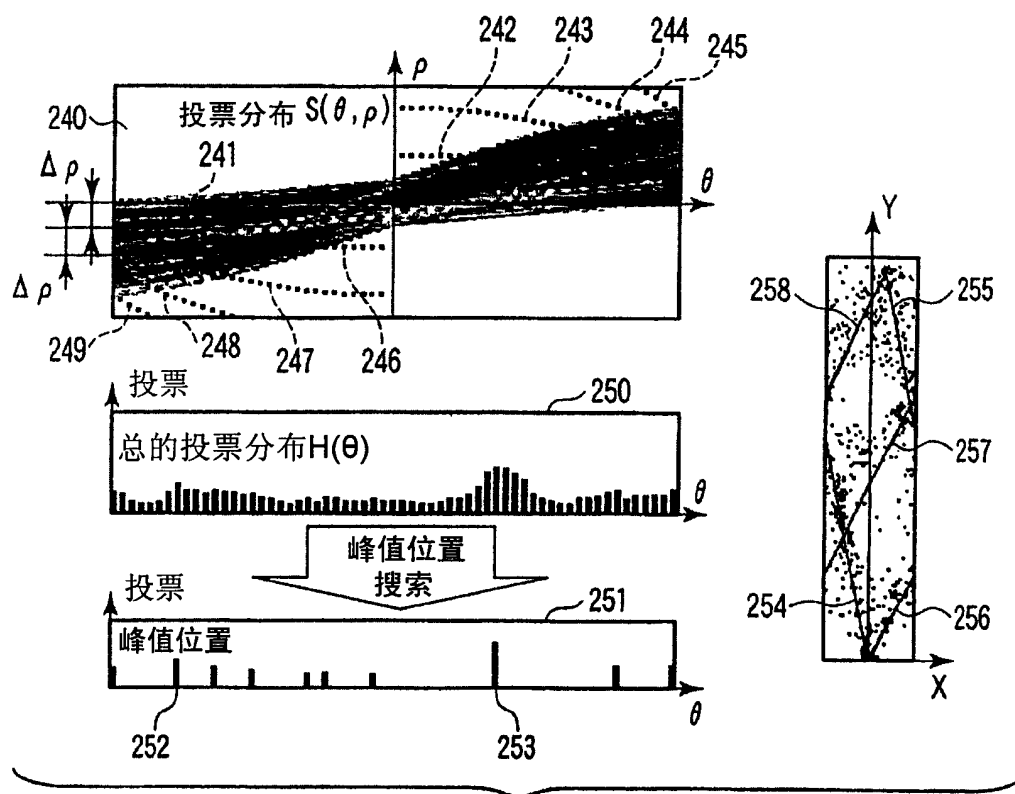


图19

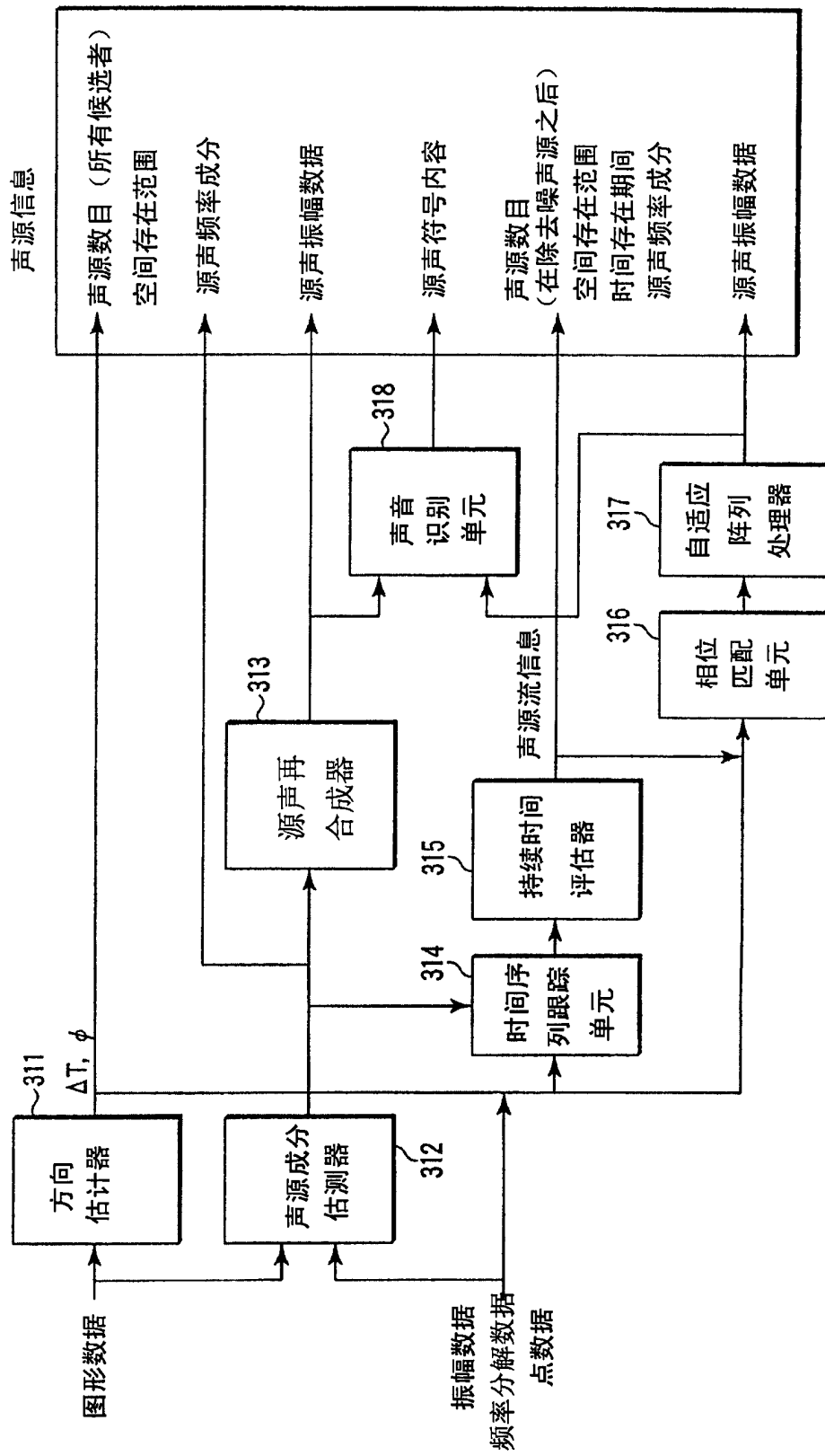
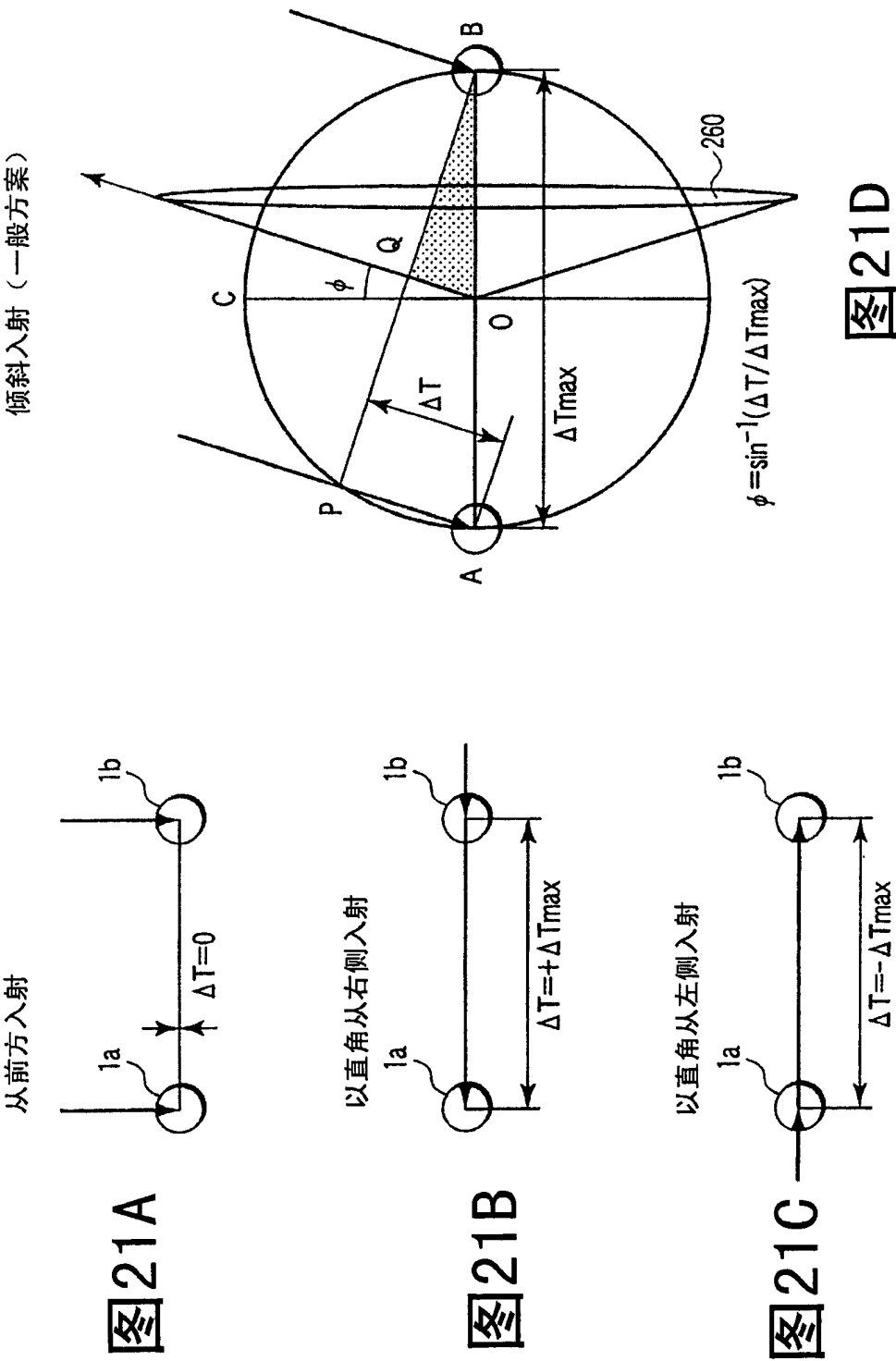


图20



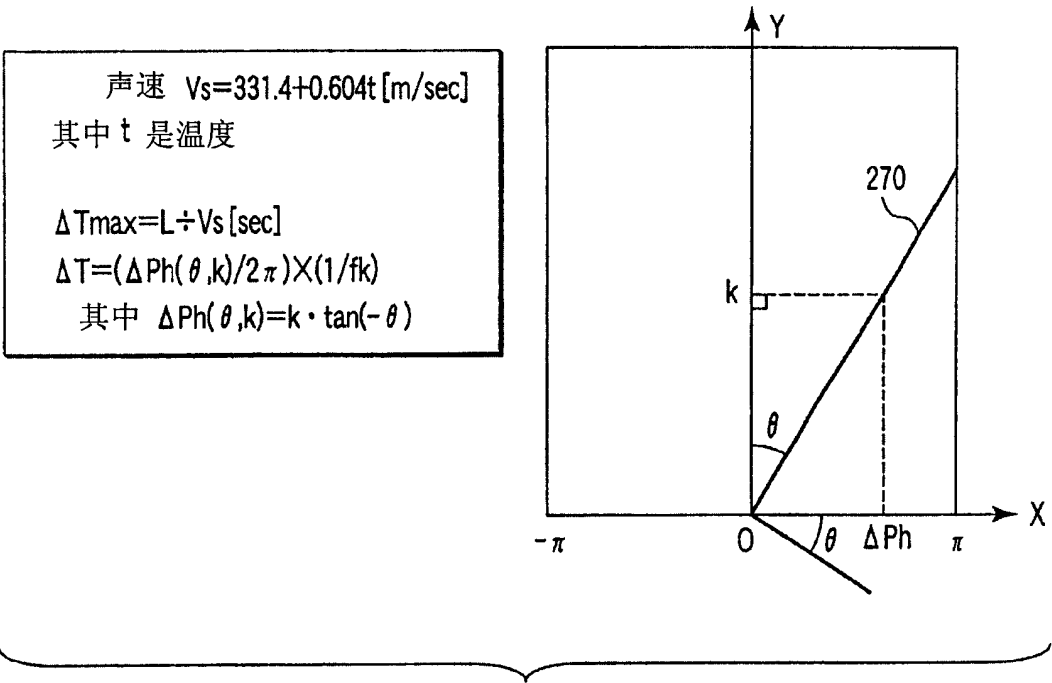
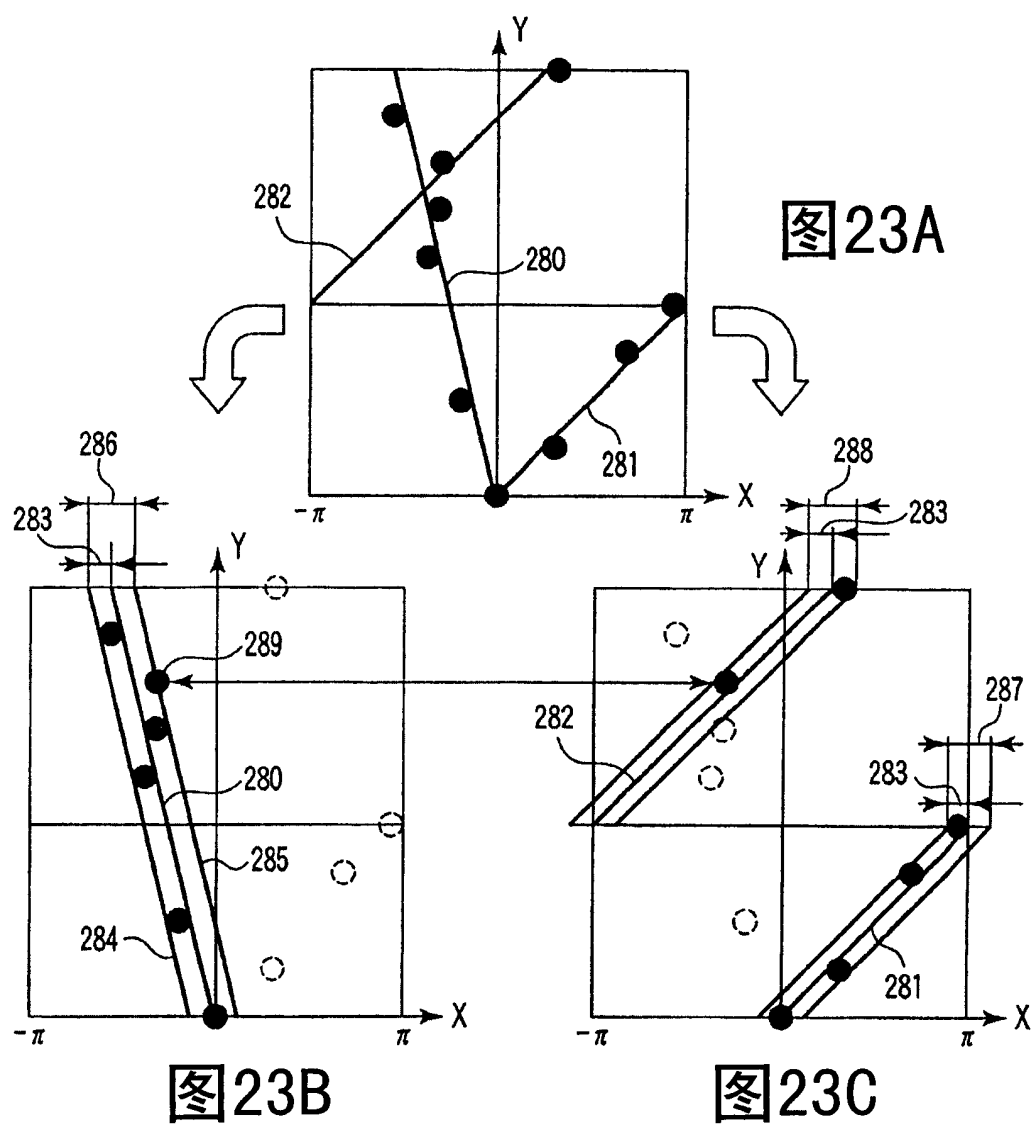


图22



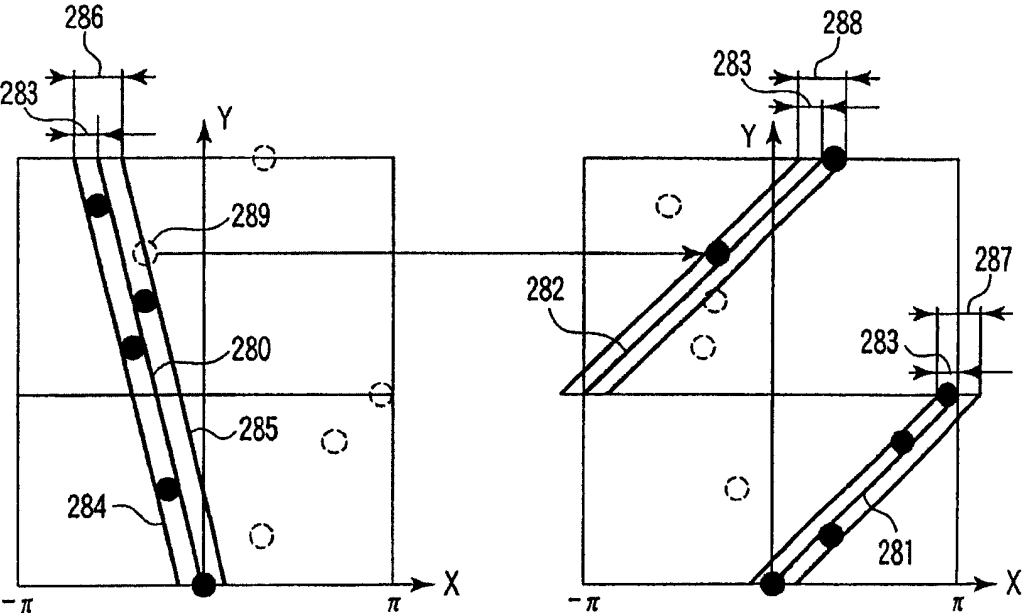


图24

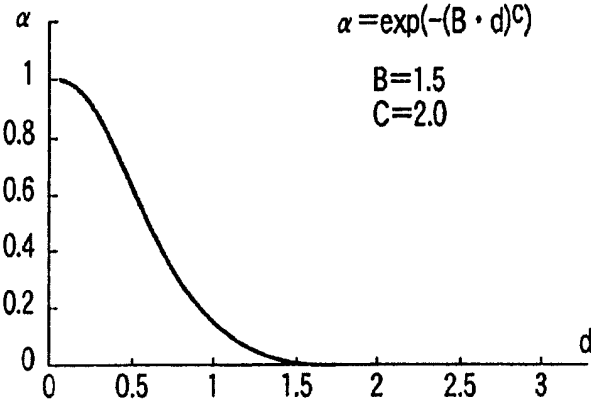


图25

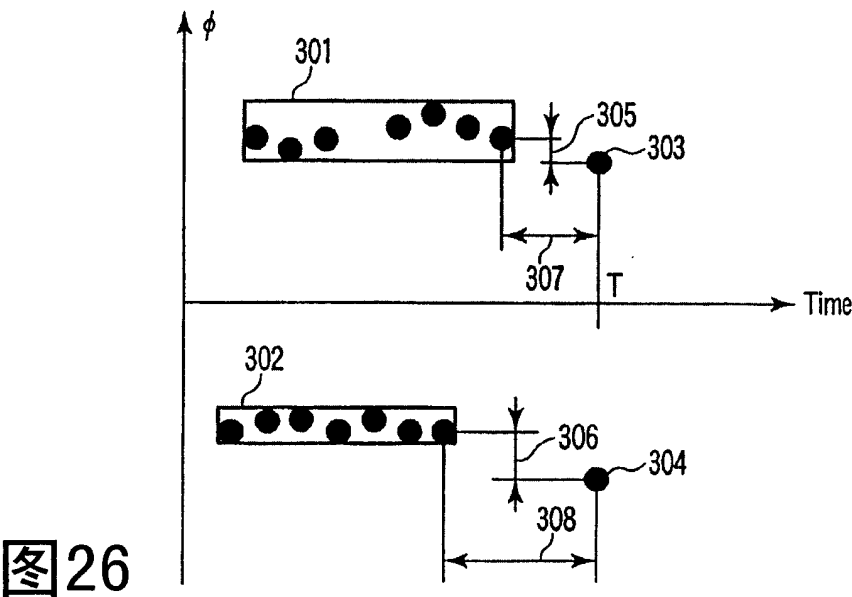


图26

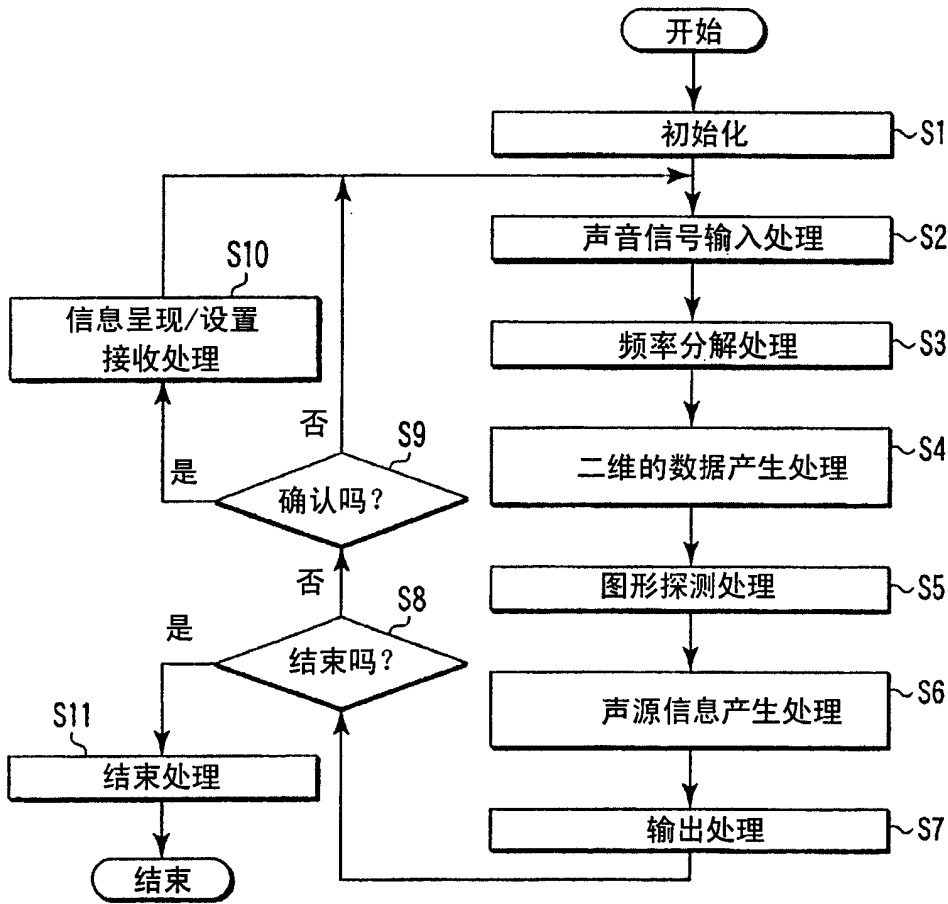


图27

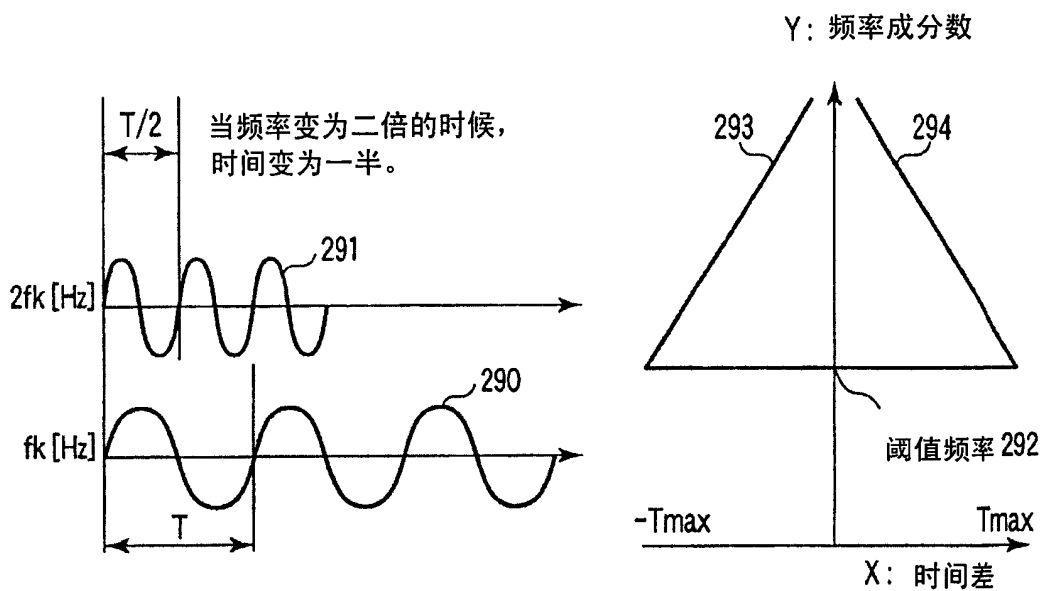


图28A

图28B

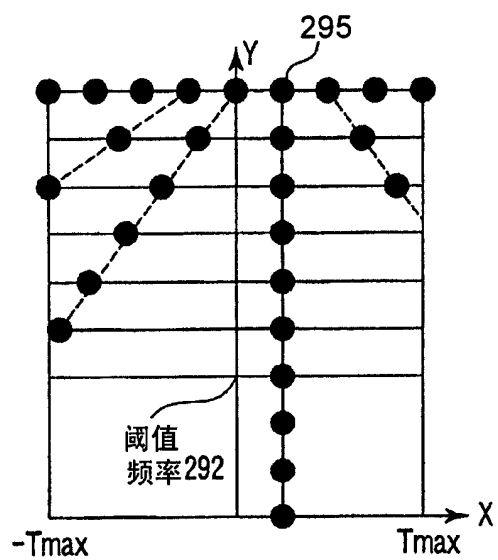


图29

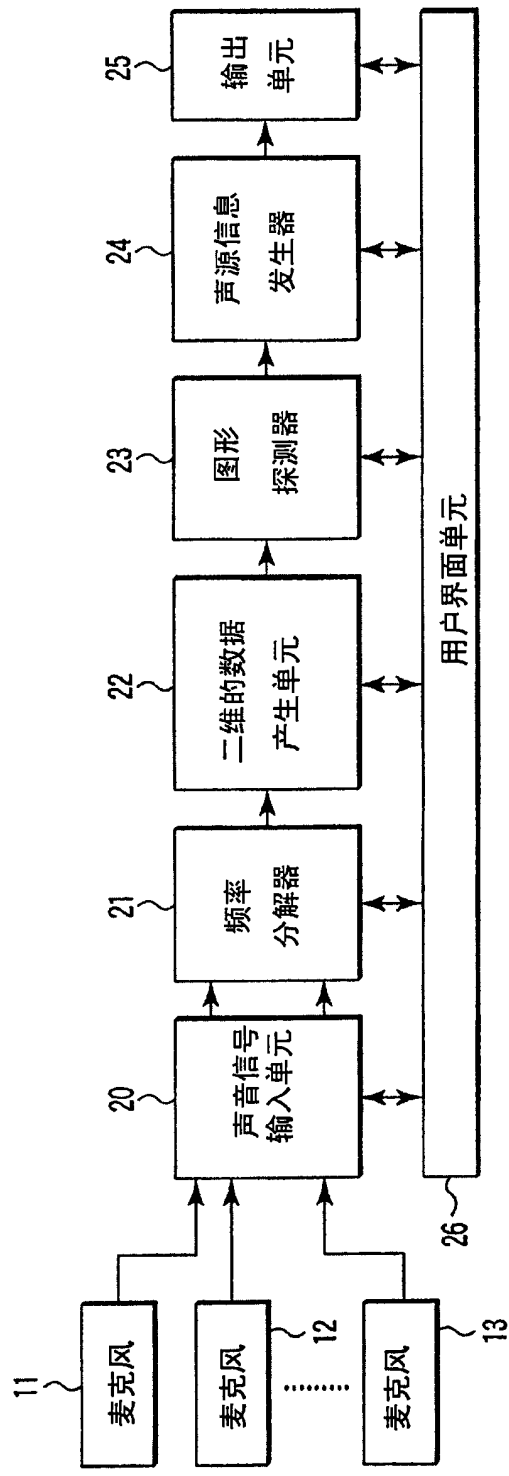


图30

