(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103824556 A
(43)申请公布日 2014.05.28

(21)申请号 201310552914.4
(22)申请日 2013.11.08
(30)优先权数据
2012-251809 2012.11.16 JP
2013-037542 2013.02.27 JP
(71)申请人 索尼公司
地址 日本东京都
(72)发明人 澁谷崇 安部素嗣 西口正之
(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 杜诚 贾萌
(51)Int.Cl.
G10L 15/08(2006.01)
G10L 15/02(2006.01)
G10L 25/54(2013.01)
G10L 25/78(2013.01)

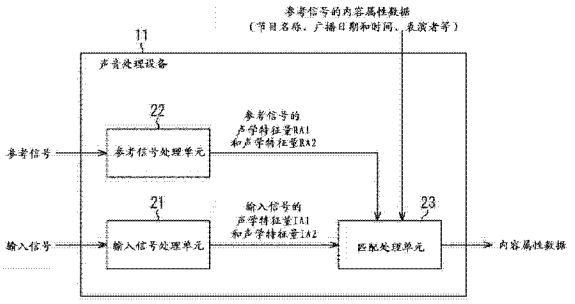
权利要求书1页 说明书13页 附图9页

(54)发明名称

声音处理设备、声音处理方法和程序

(57)摘要

提供有一种声音处理设备、声音处理方法和程序。该声音处理设备包括：输入信号处理单元，被配置为基于要识别内容的输入信号来计算第一声学特征量和与第一声学特征量不同的第二声学特征量，第一声学特征量指示在每个时频域中信号是正弦波的似然性；参考信号处理单元，被配置为基于预先准备的内容的参考信号来计算第一声学特征量和第二声学特征量；以及匹配处理单元，被配置为基于输入信号的第一声学特征量和第二声学特征量与参考信号的第一声学特征量和第二声学特征量来计算输入信号与参考信号之间的相似度。



CN 103824556 A

1. 一种声音处理设备，包括：

输入信号处理单元，被配置为基于要识别内容的输入信号来计算第一声学特征量和与所述第一声学特征量不同的第二声学特征量，所述第一声学特征量指示在每个时频域中信号是正弦波的似然性；

参考信号处理单元，被配置为基于预先准备的内容的参考信号来计算所述第一声学特征量和所述第二声学特征量；以及

匹配处理单元，被配置为基于所述输入信号的所述第一声学特征量和所述第二声学特征量与所述参考信号的所述第一声学特征量和所述第二声学特征量来计算所述输入信号与所述参考信号之间的相似度。

2. 根据权利要求 1 所述的声音处理设备，其中，所述匹配处理单元基于所述输入信号的所述第一声学特征量和所述参考信号的所述第一声学特征量来生成屏蔽图，所述屏蔽图指示在每个时频域中信号是内容的似然性，并且所述匹配处理单元基于所述屏蔽图、所述第一声学特征量以及所述第二声学特征量来计算所述相似度。

3. 根据权利要求 2 所述的声音处理设备，其中，所述匹配处理单元还计算所述输入信号的所述第一声学特征量与所述参考信号的所述第一声学特征量之间的相似度，并且基于所述屏蔽图、所述第一声学特征量之间的相似度以及所述第二声学特征量来计算所述输入信号与所述参考信号之间的相似度。

4. 根据权利要求 3 所述的声音处理设备，其中，所述匹配处理单元通过使所述参考信号对所述第一声学特征量之间的相似度的贡献率大于所述输入信号对所述第一声学特征量之间的相似度的贡献率来计算所述第一声学特征量之间的相似度。

5. 根据权利要求 4 所述的声音处理设备，其中，基于所述输入信号或所述参考信号的频谱图计算所述第二声学特征量，并且所述第二声学特征量在时间轴和频率轴上具有与所述第一声学特征量相同的粒度。

6. 一种声音处理方法，包括：

基于要识别内容的输入信号来计算第一声学特征量和与所述第一声学特征量不同的第二声学特征量，所述第一声学特征量指示在每个时频域中信号是正弦波的似然性；

基于预先准备的内容的参考信号来计算所述第一声学特征量和所述第二声学特征量；以及

基于所述输入信号的所述第一声学特征量和所述第二声学特征量与所述参考信号的所述第一声学特征量和所述第二声学特征量来计算所述输入信号与所述参考信号之间的相似度。

7. 一种用于使计算机执行下述处理的程序：

基于要识别内容的输入信号来计算第一声学特征量和与所述第一声学特征量不同的第二声学特征量，所述第一声学特征量指示在每个时频域中信号是正弦波的似然性；

基于预先准备的内容的参考信号来计算所述第一声学特征量和所述第二声学特征量；以及

基于所述输入信号的所述第一声学特征量和所述第二声学特征量与所述参考信号的所述第一声学特征量和所述第二声学特征量来计算所述输入信号与所述参考信号之间的相似度。

声音处理设备、声音处理方法和程序

[0001] 相关申请的交叉引用

[0002] 本申请要求2012年11月16日提交的日本在先专利申请JP2012-251809以及2013年2月27日提交的日本在先专利申请JP2013-037542的优先权，其全部内容通过引用合并于此。

技术领域

[0003] 本技术涉及一种声音处理设备、声音处理方法和程序。更具体地，本技术涉及一种能够以较高的准确度识别任何内容的声音处理设备、声音处理方法和程序。

背景技术

[0004] 作为示例，构成内容的声音信号被设定为参考信号，并且通过在任何设备中拾取基于参考信号再现的声音来获得输入信号。在基于这些输入信号和参考信号进行匹配检索时，可以对内容进行识别。在此情况下，从原始声源输出的声音在其中混合有混响或噪声的状态下被拾取，因此基于输入信号的声音变成混响声音或噪声被叠加在内容声音上的声音。

[0005] 作为这种内容识别技术的示例，存在这样的音乐片段识别技术：记录在CD（压缩盘）等中的无噪声音乐的信号被设定为参考信号，并且从与非音乐声音混合的输入信号中识别无噪声音乐的背景音乐。

[0006] 在该音乐片段识别技术中，通过在根据无噪声音乐的参考信号提取的声学特征量与根据输入信号提取的声学特征量之间的匹配处理来进行对音乐片段的识别。在以下描述中，假定输入信号混合有噪声，因此根据输入信号获得的声学特征量会受噪声影响。

[0007] 因此，例如，在匹配处理中使用屏蔽图。屏蔽图是表示来自构成声学特征量的各个元素之中的可靠元素的信息。在使用屏蔽图的匹配处理中，通过将构成多维声学特征量的每个元素划分为可靠元素和不可靠元素并且通过基于屏蔽图仅使用可靠元素来进行匹配。

[0008] 作为以这种方式使用屏蔽图的音乐片段识别技术，提出了例如一种执行音乐片段识别的方法，在该方法中，预先准备多个屏蔽图，以针对具有时频分量的特征矩阵屏蔽给定的时频域（例如，参考日本未审专利申请公报2009-276776号）。

[0009] 在上述方法中，通过将使用针对输入信号的特征矩阵和数据库中音乐片段的特征矩阵（即参考信号的特征矩阵）预先准备的所有屏蔽图计算出的相似度之中的最大值设定为输入信号与音乐片段之间的相似度来执行音乐片段识别。在该音乐片段识别中，存储了取决于输入信号的多个固定屏蔽图，并且使用这些屏蔽图执行匹配处理。

发明内容

[0010] 然而，在上述技术中，内容识别被专用于音乐的匹配检索，因此不可以识别任何公共使用的内容，例如，诸如广播节目的内容。例如，对于广播节目内容，可能存在需要将没有音乐的场景的声音信号检索为输入信号的情况。然而，在这种情况下，难以使用上述技术来

识别内容。

[0011] 此外，在上述技术中，未考虑声音中混响的影响并且因此可能不能以高准确度实现内容识别。换言之，输入信号受实际使用环境中的混响影响并且混响对检索造成不利影响。因此，在具有强混响的环境中，降低了对内容的匹配检索的准确度。

[0012] 此外，在日本未审专利申请公报 2009-276776 号中所公开的技术中，使用固定的屏蔽图。然而，对于包括在输入信号中的混合噪声，可能不能预测何时包括有噪声以及噪声具有哪种属性。因而，难以预先为输入信号准备最佳屏蔽图。因此，可能不能使用预先准备的屏蔽图以高准确度对内容进行识别。

[0013] 鉴于这种情形，做出了本技术的实施例。期望以较高的准确度识别任何内容。

[0014] 根据本技术的实施例，提供有一种声音处理设备，包括：输入信号处理单元，被配置为基于要识别内容的输入信号来计算第一声学特征量和与第一声学特征量不同的第二声学特征量，第一声学特征量指示在每个时频域中信号是正弦波的似然性；参考信号处理单元，被配置为基于预先准备的内容的参考信号来计算第一声学特征量和第二声学特征量；以及匹配处理单元，被配置为基于输入信号的第一声学特征量和第二声学特征量与参考信号的第一声学特征量和第二声学特征量来计算输入信号与参考信号之间的相似度。

[0015] 匹配处理单元可以基于输入信号的第一声学特征量和参考信号的第一声学特征量来生成屏蔽图，屏蔽图指示在每个时频域中信号是内容的似然性，并且匹配处理单元可以基于屏蔽图、第一声学特征量以及第二声学特征量来计算相似度。

[0016] 匹配处理单元还可以计算输入信号的第一声学特征量与参考信号的第一声学特征量之间的相似度，并且可以基于屏蔽图、第一声学特征量之间的相似度以及第二声学特征量来计算输入信号与参考信号之间的相似度。

[0017] 匹配处理单元可以通过使参考信号对第一声学特征量之间的相似度的贡献率大于输入信号对第一声学特征量之间的相似度的贡献率来计算第一声学特征量之间的相似度。

[0018] 可以基于输入信号或参考信号的频谱图计算第二声学特征量，并且第二声学特征量可以在时间轴和频率轴上具有与第一声学特征量相同的粒度。

[0019] 根据本技术的实施例，提供有包括下述步骤的声音处理方法和程序。所述步骤为：基于要识别内容的输入信号来计算第一声学特征量和与第一声学特征量不同的第二声学特征量，第一声学特征量指示在每个时频域中信号是正弦波的似然性；基于预先准备的内容的参考信号来计算第一声学特征量和第二声学特征量；以及基于输入信号的第一声学特征量和第二声学特征量与参考信号的第一声学特征量和第二声学特征量来计算输入信号与参考信号之间的相似度。

[0020] 根据本技术的实施例，基于要识别内容的输入信号来计算第一声学特征量和与第一声学特征量不同的第二声学特征量，第一声学特征量指示在每个时频域中信号是正弦波的似然性；基于预先准备的内容的参考信号来计算第一声学特征量和第二声学特征量；以及基于输入信号的第一声学特征量和第二声学特征量与参考信号的第一声学特征量和第二声学特征量来计算输入信号与参考信号之间的相似度。

[0021] 根据本公开的一个或更多实施例，可以以较高的准确度对任何内容进行识别。

附图说明

[0022] 图 1 是用于说明屏蔽图的图；

[0023] 图 2 是例示声音处理设备的示例性配置的图；

[0024] 图 3 是例示输入信号处理单元的示例性配置的图；

[0025] 图 4 是例示参考信号处理单元的示例性配置的图；

[0026] 图 5 是例示匹配处理单元的示例性配置的图；

[0027] 图 6 是用于说明声学特征量的图；

[0028] 图 7 是用于说明匹配检索处理的流程图；

[0029] 图 8 是用于说明对声学特征量 IA1 的提取处理的流程图；

[0030] 图 9 是用于说明对声学特征量 IA2 的提取处理的流程图；以及

[0031] 图 10 是例示计算机的示例性配置的图。

具体实施方式

[0032] 在下文中，将参考附图对本公开优选实施例进行详细描述。请注意：在本说明书和附图中，用相同的附图标记来表示具有基本相同功能和结构的结构元件，并且省略对这些结构元件的重复说明。

[0033] < 第一实施例 >

[0034] < 本技术的实施例的技术特征 >

[0035] 本技术的实施例使得可以通过使用便携式终端设备（诸如多功能手机或平板型终端设备）的记录功能来识别用户用其它设备观看的任何内容，诸如电视节目、广播节目以及流分发内容。

[0036] 在从诸如电视接收器、收音机或个人计算机的设备的扬声器输出要处理的声音，并且由便携式终端设备记录所输出的声音的情况下，声音穿过设备的扬声器与便携式终端设备之间的空间。由此，通过记录获得的声音也会包括由于在空间中行进引起的声音的混响。另外，通过记录获得的声音被与不同于从设备的扬声器输出的声音的声音（在下文中，这被称为“混合噪声”）混合。

[0037] 在本技术的实施例中，期望执行内容的匹配检索，其对混响或混合噪声是鲁棒的。更一般地，期望在原始声源（干源（dry source））与叠加有通过使给定的声源穿过空间而产生的混响或混合噪声的声源之间执行匹配检索。

[0038] 现在将描述本技术实施例的技术特征。例如，本技术实施例可以具有如下特征中的一个或更多。

[0039] 技术特征（1）

[0040] 使用分别针对输入信号和参考信号计算的、指示在剪切的时频域中的每个时频域中是正弦波的似然性的指标来生成屏蔽图。

[0041] 技术特征（2）

[0042] 通过频谱形状在微小时间内的稳定性量化指示为正弦波的似然性的指标。

[0043] 技术特征（3）

[0044] 为正弦波的似然性是对混响鲁棒的指标。

[0045] 技术特征（4）

[0046] 使用输入信号的信息以及参考信号的信息生成屏蔽图。

[0047] 技术特征(5)

[0048] 在计算输入信号与参考信号之间的相似度时,通过将优先权给予参考信号而不是输入信号来计算该相似度,而不是等同对待参考信号和输入信号。

[0049] 例如,在本技术的实施例中,如图 1 所示,获得输入信号的频谱图和参考信号的频谱图。另外,在图 1 中,垂直轴指示频率而水平轴指示时间。

[0050] 在图 1 中,图的右侧指示参考信号的频谱图,而图的左侧指示输入信号的频谱图。

[0051] 在输入信号的频谱图即时频域中,由实线表示的分量指示也包括在参考信号中的声音分量,而由虚线表示的分量指示不包括在参考信号中的混合噪声分量。

[0052] 在本技术的实施例中,通过生成屏蔽图指定图中作为阴影线部分区域的可靠时频域,并且通过仅使用该可靠时频域进行输入信号与参考信号之间的匹配处理。

[0053] 根据本技术的实施例,可以获得如下一个或更多有益效果。

[0054] 有益效果(1)

[0055] 通过使用没有音乐的场景以及有音乐的场景,可以对内容进行识别。

[0056] 有益效果(2)

[0057] 即使在有混响的空间中,也可以对诸如观看节目的内容进行识别。

[0058] 有益效果(3)

[0059] 即使在输入信号中包括有不同于包括在原始参考信号中的声音的声音(混合噪声)时,也可以对诸如观看节目的内容进行识别。

[0060] < 声音处理设备的示例性配置 >

[0061] 现在将描述本技术应用于其的具体实施例。

[0062] 图 2 是例示根据本技术的实施例的声音处理设备的示例性配置的图。

[0063] 声音处理设备 11 被配置为包括输入信号处理单元 21、参考信号处理单元 22 以及匹配处理单元 23。

[0064] 将包括在预先准备的内容中的声音的参考信号和包括在要识别内容中的声音的输入信号输入到声音处理设备 11。通过在另一个设备中记录(拾取)基于从给定设备再现的参考信号的声音来获得输入信号。例如,输入信号可以是通过在声音处理设备 11 中进行记录获得的声音信号。

[0065] 此外,例如,多个内容项的声音信号被输入为参考信号。另外,参考信号的内容属性数据也被输入到声音处理设备 11。内容属性数据是包括内容名称(节目名称)、广播日期和时间、表演者等的内容相关数据。

[0066] 输入信号处理单元 21 分析所提供的输入信号以生成两种类型声学特征量,即声学特征量 IA1 和声学特征量 IA2,然后将它们提供给匹配处理单元 23。

[0067] 参考信号处理单元 22 分析所提供的作为内容的原始声源的参考信号 , 以生成两种类型的声学特征量,即声学特征量 RA1 和声学特征量 RA2,然后将它们提供给匹配处理单元 23。声学特征量 RA1 和声学特征量 RA2 分别与声学特征量 IA1 和声学特征量 IA2 相对应。

[0068] 声学特征量 IA1 和声学特征量 IA2 具有相同的特征量(相同类型的特征量),而声学特征量 RA1 和声学特征量 RA2 具有相同的特征量。在下文中,在不必要对声学特征量 IA1

和声学特征量 IA2 做出区分的情况下，将它们简称为声学特征量 A1。另外，在不必要对声学特征量 RA1 和声学特征量 RA2 做出区分的情况下，将它们简称为声学特征量 A2。

[0069] 匹配处理单元 23 基于由输入信号处理单元 21 提供的声学特征量 IA1 和声学特征量 IA2 与由参考信号处理单元 22 提供的声学特征量 RA1 和声学特征量 RA2 来进行输入信号与参考信号之间的匹配处理，以对内容进行识别。另外，匹配处理单元 23 从所提供的内容属性数据之中输出通过匹配处理所识别的内容的内容属性数据，并且还输出通过匹配处理获得的结果。

[0070] < 输入信号处理单元的示例性配置 >

[0071] 图 2 所示的输入信号处理单元 21 被更具体地配置为如图 3 所示的那样。图 3 所示的输入信号处理单元 21 被配置为包括输入信号剪切部分 51、时频转换器 52、声学特征量提取器 53 以及声学特征量提取器 54。

[0072] 输入信号剪切部分 51 从所提供的输入信号中剪切具有预定时间长度的部分，并且将所剪切的输入信号提供给时频转换器 52。时频转换器 52 对由输入信号剪切部分 51 提供的输入信号进行时频转换，以将该输入信号转换为对数 - 幅度频谱图，并且将该频谱图输出到声学特征量提取器 53 和 54。

[0073] 声学特征量提取器 53 基于由时频转换器 52 提供的对数 - 幅度频谱图来计算声学特征量 IA1，并且将所计算的声学特征量 IA1 提供给匹配处理单元 23。声学特征量提取器 54 基于由时频转换器 52 提供的对数 - 幅度频谱图来计算声学特征量 IA2，并且将所计算的声学特征量 IA2 提供给匹配处理单元 23。

[0074] 现在将描述声学特征量 IA1 和 IA2。

[0075] 例如，声学特征量 IA1 和声学特征量 IA2 都由具有分别与时间分量和频率分量相对应的两个轴的矩阵表示。每个矩阵具有以下特征。

[0076] 换言之，声学特征量 IA1 是表示在每个时频域中输入信号是正弦波的似然性的特征矩阵。

[0077] 此外，声学特征量 IA2 是用于在输入信号与参考信号之间的匹配的特征量，并且是表示信号的个性特征的特征矩阵。然而，声学特征量 IA2 的时间轴和频率轴的粒度与声学特征量 IA1 的时间轴和频率轴的粒度是相同的。

[0078] 此外，现在将详细描述输入信号处理单元 21 计算声学特征量 IA1 和声学特征量 IA2 的处理。

[0079] 输入信号剪切部分 51 从连续输入的输入信号中剪切具有一定时间长度(例如，5 秒)的信号，并且将所剪切的信号输出到时频转换器 52。时频转换器 52 将所剪切的输入信号转换为对数 - 幅度频谱图(在下文中，简称为频谱图)。

[0080] 此外，声学特征量提取器 53 将频谱图转换为通过将在划分的时频域中具有频谱图的正弦波的似然性数字化而获得的中间特征量。

[0081] 换言之，频谱图的微小时间内的稳定性被用于将作为正弦波的似然性数字化。与噪声不同，乐器的声音或人的嗓音可以被视为在微小时间(例如，0.020 秒)内频率基本上不变的正弦波，因此频谱图在形状上基本上不变。

[0082] 声学特征量提取器 53 通过使用该属性针对每个频带将频谱图在微小时间内的稳定性数字化，并且将经数字化的值视为指示作为正弦波的似然性的指标。更具体地，声学特

征量提取器 53 对于频谱图的每个时间帧执行峰检测处理，并且对于峰周围的时频域将对数 - 幅度频谱图近似为由以下式（1）表示的双二次函数 g(k,n)。

[0083] $$g(k,n)=\bar{a}k^2+\bar{b}k+\bar{c} \quad (1)$$

[0084] 在式（1）中，k 表示频谱图的频率区间（bin）编号，并且 n 表示频谱图的时间帧编号。另外，使用诸如最小二乘法的最优化技术进行对数 - 幅度频谱图的近似。

[0085] 接下来，声学特征量提取器 53 将所检测的峰周围的时频域的每个时间帧的对数 - 幅度频谱近似为由以下式（2）表示的二次函数 fn(k)。

[0086] $f_n(k)=a_nk^2 + b_nk + c_n$ (2)

[0087] 类似地，使用诸如最小二乘法的最优化技术进行近似。

[0088] 此外，声学特征量提取器 53 通过使用由对双二次函数 g(k,n) 和二次函数 fn(k) 的两种类型的函数进行近似所获得的系数的式（3）计算作为正弦波的似然性。

[0089] $$\eta(n,k)=1-\alpha\sqrt{\sum\{D_1(a_n,\bar{a})+D_2(b_n,\bar{b})\}} \quad (3)$$

[0090] 在式（3）中，α 是具有正值的参数。D(x,y) 即 $D_1(x,y)$ 和 $D_2(x,y)$ 表示距离函数。

[0091] 此外，在对正弦波进行时频转换时，存在二次函数的二阶系数的理论值。考虑到理论值与所计算的二阶系数的接近性，信号是正弦波的似然性可以由以下式（4）计算。

[0092] $$\eta(n,k)=1-\alpha\sqrt{\sum\{D_1(a_n,\bar{a})+D_2(b_n,\bar{b})+D_3(a_n,a)\}} \quad (4)$$

[0093] 在式（4）中，η(n,k) 意指在每个峰处信号是正弦波的似然性，因此如果 η(n,k)<0，则 η(n,k) 变成 0。由此，η(n,k) 取范围为从 0 至 1 的值。

[0094] 此外，对于与峰不对应的频率区间 η(n,k)=0，并且对于相应的时间帧获得包含每个频率区间信号是正弦波的似然性的信息的向量。信号是正弦波的似然性是对混响鲁棒的特征量，因此最后可以进行对混响鲁棒的检索。

[0095] 在使时间帧移位的同时计算以上面描述的方式获得的向量，并且所获得的向量以时间序列布置并且被沿时间轴方向进行下采样，从而获得声学特征量 IA1。为了进行下采样，使用平滑滤波器（低通滤波器）。通过滤波获得的值意指在每个频率处信号是正弦波的似然性的时间平均值。

[0096] 对于所获得的声学特征量 IA1 的每个元素，可以进行量化处理或非线性处理（诸如对数函数、指数函数或 S 形函数）。

[0097] 此外，在声学特征量提取器 54 中，将频谱图转换为声学特征量 IA2。

[0098] 例如，一阶微分滤波器沿时间轴方向应用于以与声学特征量 IA1 相似的方式计算的信号是正弦波的似然性的矩阵，并且对以这种方式获得的矩阵进行下采样，从而获得声学特征量 IA2。通过一阶微分滤波器的滤波获得的值意指在每个频率处信号是正弦波的似然性的时间变化。

[0099] 对于所获得的声学特征量 IA2 的每个元素，可以进行量化处理或非线性处理（诸如对数函数、指数函数或 S 形函数）。此外，作为声学特征量 IA2，可以使用表示信号的个性特征的值，例如，可以使用通过使一定时间间隔中的频谱的时间平均归一化而获得的值。

[0100] < 参考信号处理单元的示例性配置 >

[0101] 图 4 例示图 2 所示的参考信号处理单元 22 的更详细的配置。图 4 中所示的参考

信号处理单元22被配置为包括参考信号剪切部分81、时频转换器82、声学特征量提取器83以及声学特征量提取器84。

[0102] 参考信号剪切部分81从所提供的参考信号中剪切具有预定时间长度的部分并且将所剪切的参考信号提供给时频转换器82。时频转换器82对由参考信号剪切部分81提供的参考信号进行时频转换，以将参考信号转换为对数－幅度频谱图，并且将该频谱图输出到声学特征量提取器83和声学特征量提取器84。

[0103] 声学特征量提取器83基于由时频转换器82提供的对数－幅度频谱图来计算声学特征量RA1，并且将所计算的声学特征量RA1提供给匹配处理单元23。声学特征量提取器84基于由时频转换器82提供的对数－幅度频谱图来计算声学特征量RA2，并且将所计算的声学特征量RA2提供给匹配处理单元23。

[0104] 声学特征量提取器83和声学特征量提取器84分别与声学特征量提取器53和声学特征量提取器54相对应。声学特征量提取器83和声学特征量提取器84分别输出声学特征量RA1和声学特征量RA2。声学特征量RA1和声学特征量RA2分别具有与声学特征量IA1和声学特征量IA2相同的在时间轴和频率轴上的粒度。

[0105] 另外，根据参考信号提取的声学特征量RA1和声学特征量RA2可以被直接提供给匹配处理单元23，或者可以被提供给存储设备以被保存作为数据库。然而，在将声学特征量RA1和声学特征量RA2提供给存储设备时，声学特征量RA1和声学特征量RA2需要与参考信号的元数据(节目名称、广播日期和时间、表演者等)即内容属性数据结合地保存。

[0106] ＜匹配处理单元的示例性配置＞

[0107] 图5例示图2中所示的匹配处理单元23的更详细的配置。图5中所示的匹配处理单元23被配置为包括屏蔽图生成器111、相似度计算器112以及比较积分器113。

[0108] 屏蔽图生成器111基于由声学特征量提取器53提供的声学特征量IA1和由声学特征量提取器83提供的声学特征量RA1来生成屏蔽图。然后，屏蔽图生成器111将所生成的屏蔽图以及声学特征量A1之间的相似度输出到相似度计算器112。屏蔽图指示每个时频域中信号是内容的似然性的可靠性，即可靠的时频域。

[0109] 相似度计算器112基于由声学特征量提取器54提供的声学特征量IA2、由声学特征量提取器84提供的声学特征量RA2以及由屏蔽图生成器111提供的屏蔽图和相似度来计算输入信号与参考信号的相似度。另外，相似度计算器112将所计算的相似度和所提供的内容属性数据提供给比较积分器113。

[0110] 比较积分器113基于由相似度计算器112提供的相似度来确定参考信号的内容与输入信号中包括的内容是否彼此相同，并且将确定结果和内容属性数据输出。

[0111] 匹配处理单元23计算参考信号与输入信号之间的相似度。例如，如图6中所示，在参考信号的碎片被包括在具有一定时间段(例如，5秒)的输入信号中的情况下，输入信号的声学特征量IA1和声学特征量IA2的矩阵在时间方向上的分量的数目通常小于参考信号的声学特征量RA1和声学特征量RA2。

[0112] 因此，通过从参考信号的声学特征量RA1和声学特征量RA2的矩阵中剪切在时间方向上具有与输入信号的声学特征量IA1和声学特征量IA2的长度相同的长度的部分矩阵来计算相似度。为了剪切该部分矩阵，剪切可以被剪裁的所有部分矩阵。在屏蔽图生成器111和相似度计算器112中进行剪切处理。

[0113] 在图 6 中，垂直方向表示频率并且水平方向表示时间。另外，由箭头 Q11、Q12、Q13 以及 Q14 指示的矩形形状分别表示参考信号的声学特征量 RA1、参考信号的声学特征量 RA2、输入信号的声学特征量 IA1 以及输入信号的声学特征量 IA2。

[0114] 在此示例中，可见从参考信号提取的声学特征量 RA1 和声学特征量 RA2 与从输入信号提取的声学特征量 IA1 和声学特征量 IA2 相比，在图中的水平方向上即时间方向上更长并且在时间方向上分量的数目更多。

[0115] 因此，声学特征量 RA1 和声学特征量 RA2 的一部分被剪切成部分矩阵。该部分矩阵被用于计算相似度。

[0116] 接下来，现在将描述在匹配处理单元 23 中进行的详细处理。

[0117] 屏蔽图生成器 111 根据输入信号的声学特征量 IA1 和参考信号的声学特征量 RA1 生成屏蔽图，并且还计算声学特征量 A1 之间的相似度。屏蔽图以与声学特征量 A1 相似的方式被表示为具有时间轴和频率轴的二维矩阵。

[0118] 例如，根据输入信号的声学特征量 IA1 和参考信号的声学特征量 RA1 生成对不存在正弦波的时频域进行掩蔽的矩阵作为屏蔽图。更具体地，例如，通过计算下式(5)来生成屏蔽图。

[0119] $$W_{f(t+u)} = S^{(1)}_{fu} A^{(1)}_{f(t+u)} \qquad (5)$$

[0120] 在式(5)中，$W_{f(t+u)}$ 表示屏蔽图的矩阵元素，$S^{(1)}{}_{fu}$ 表示输入信号的声学特征量 IA1 的矩阵元素，而 $A^{(1)}{}_{f(t+u)}$ 表示参考信号的声学特征量 RA1 的部分矩阵的元素。

[0121] 另外，f 表示每个矩阵的频率分量，u 表示每个矩阵的时间分量，而 t 表示部分矩阵的时间偏移。

[0122] 以这种方式计算的屏蔽图在后续级的相似度计算器 112 中用作为每个时频域的权重。换言之，计算了相似度，该相似度将优先权给予具有屏蔽图的矩阵元素 $W_{f(t+u)}$ 的较大值的时频域。

[0123] 声学特征量 A1 之间的相似度是通过对两个特征量的接近性进行量化而获得的非负的指标，并且例如通过以下式(6)计算。

[0124] $$R^{(1)}(t) = \frac{\sum S^{(1)}_{fu} A^{(1)}_{f(t+u)}}{\left(\sum S^{(1)p}_{fu}\right)^{1/p} \cdot \left(\sum A^{(1)\ q}_{f(t+u)}\right)^{1/q}} \qquad (6)$$

[0125] 在式(6)中，$R^{(1)}(t)$ 表示 $S^{(1)}{}_{fu}$ 与 $A^{(1)}{}_{f(t+u)}$ 之间的相似度。另外，p 和 q 是用于调整对输入信号的声学特征量 IA1 与参考信号的声学特征量 RA1 之间的相似度的贡献率的参数。换言之，p 和 q 是具有 1 或更大的值的满足 1/p+1/q=1 的权重系数。

[0126] 例如，通过使 p 大于 q，计算将优先权给予参考信号中包括的声音的相似度，并且即使在输入信号中包括有与参考信号无关的混合噪声的情况下，也可以进行其中该噪声的影响被减弱的匹配。此外，作为声学特征量之间的相似度，除了上面描述的相似度之外，可以使用基于诸如平方误差或绝对误差的两个矩阵的差而计算的值。

[0127] 此外，相似度计算器 112 通过使用输入信号的声学特征量 IA2、参考信号的声学特征量 RA2、屏蔽图以及声学特征量 A1 之间的相似度来计算最终相似度。

[0128] 通过将具有在时频域中信号是正弦波的似然性的信息的屏蔽图视为每个时频域

中的可靠性，并且通过对所获得的屏蔽图进行加权和量化来获得由相似度计算器 112 计算的相似度。另外，由相似度计算器 112 计算的相似度是时频域中输入信号的声学特征量 IA2 与参考信号的声学特征量 RA2 之间的接近性的指标。此外，考虑到声学特征量 A1 之间的相似度，例如，相似度 R(t) 通过以下式（7）的计算而被计算。

[0129] $$R(t)=\frac{\sum W_{f(t+u)}\exp\left(-\beta\left(S^{(2)}_{fu}-A^{(2)}_{f(t+u)}\right)^2\right)}{\sum W_{f(t+u)}}R^{(1)}(t) \qquad (7)$$

[0130] 在式（7）中，$A^{(2)}_{f(t+u)}$ 表示参考信号的声学特征量 RA2 的部分矩阵，而 $S^{(2)}_{fu}$ 表示输入信号的声学特征量 IA2 的矩阵。另外，β 是具有正值的参数。

[0131] 此外，除了通过式（7）的计算之外，可以使用基于两个矩阵（输入信号的声学特征量 IA2 和参考信号的声学特征量 RA2）的差（诸如平方误差或绝对误差）计算的值计算相似度。

[0132] 比较积分器 113 基于由相似度计算器 112 计算的相似度来确定参考信号的内容与输入信号中包括的内容是否彼此相同。

[0133] 确定关于内容是否彼此相同的方法是确定是下述内容的方法：在所述内容中，具有针对多个参考信号获得的相似度之中超过了预定阈值的最大相似度的参考信号被包括在输入信号中。另外，如果参考信号的任何相似度均不超过阈值，则确定在参考信号中不存在目标内容。

[0134] 此外，此处使用的阈值通常可以是固定值或者可以根据从输入信号和多个参考信号中获得的多个相似度以统计方法设定。

[0135] ＜对匹配检索处理的描述＞

[0136] 在将输入信号和参考信号提供给声音处理设备 11 的情况下，如果存在内容识别的指令，则声音处理设备 11 进行匹配检索处理，并且然后进行内容识别。参考图 7 的流程图，现在将描述由声音处理设备 11 进行的匹配检索处理。

[0137] 在步骤 S11 中，输入信号剪切部分 51 剪切所提供的输入信号并且将所剪切的输入信号提供给时频转换器 52。例如，剪切了具有一定时间长度的输入信号。

[0138] 在步骤 S12 中，时频转换器 52 对由输入信号剪切部分 51 提供的输入信号进行时频转换，以将输入信号转换为对数－幅度频谱图，然后将该对数－幅度频谱图提供给声学特征量提取器 53 和声学特征量提取器 54。

[0139] 在步骤 S13 中，声学特征量提取器 53 进行对声学特征量 IA1 的提取处理以计算输入信号的声学特征量 IA1，然后将所计算的声学特征量 IA1 提供给匹配处理单元 23 的屏蔽图生成器 111。

[0140] 在下文中，参考图 8 的流程图，将描述由声学特征量提取器 53 进行的对声学特征量 IA1 的提取处理。该提取处理与步骤 S13 的处理相对应。

[0141] 在步骤 S51 中，声学特征量提取器 53 为由时频转换器 52 提供的对数－幅度频谱图选择时间帧。

[0142] 在步骤 S52 中，声学特征量提取器 53 对对数－幅度频谱图的所选择的时间帧进行峰检测。

[0143] 在步骤 S53 中，声学特征量提取器 53 将所检测的峰周围的时频域的对数 - 幅度频谱图近似为两种类型的二次函数。例如，对数 - 幅度频谱图被近似为式（1）和式（2）所示的函数。

[0144] 在步骤 S54 中，声学特征量提取器 53 将近似的二次函数的系数转换为指示信号是正弦波的似然性的指标并且保存该指标。例如，式（3）的 η(n,k) 被计算作为指示信号是正弦波的似然性的指标。

[0145] 在步骤 S55 中，声学特征量提取器 53 确定是否对输入信号的所有时间帧都进行了处理。如果在步骤 S55 中确定还未对输入信号的所有时间帧进行处理，则处理返回到步骤 S51，并且重复上述处理。

[0146] 另一方面，在步骤 S55 中，如果确定对输入信号的所有时间帧都进行了处理，然后，在步骤 S56 中，声学特征量提取器 53 通过以时间序列布置信号是正弦波的似然性的指标的保存的向量来形成矩阵。

[0147] 在步骤 S57 中，声学特征量提取器 53 对形成为矩阵的指示信号是正弦波的似然性的指标，即信号是正弦波的似然性的矩阵，在时间轴方向上进行滤波，然后计算信号是正弦波的似然性的时间平均量。例如，通过使用平滑滤波器进行滤波。

[0148] 在步骤 S58 中，声学特征量提取器 53 对通过滤波获得的信号是正弦波的似然性的时间平均量在时间轴方向上进行重新采样，并且将重新采样的结果视为声学特征量 IA1。当声学特征量提取器 53 将以这种方式从输入信号提取的声学特征量 IA1 提供给屏蔽图生成器 111 时，终止对声学特征量 IA1 的提取处理。此后，处理前进到图 7 的步骤 S14。

[0149] 在步骤 S14 中，声学特征量提取器 54 通过进行提取处理来计算输入信号的声学特征量 IA2，并且然后将计算的声学特征量 IA2 提供给匹配处理单元 23 的相似度计算器 112。

[0150] 在下文中，参考图 9 的流程图，将描述由声学特征量提取器 54 进行的对声学特征量 IA2 的提取处理。该提取处理与步骤 S14 的处理相对应。另外，步骤 S91 至步骤 S96 的处理与图 8 中步骤 S51 至步骤 S56 的处理相似，因此省略了对其的描述。

[0151] 在进行步骤 S96 的处理之后，获得信号是正弦波的似然性的矩阵。在步骤 S97 中，声学特征量提取器 54 对信号是正弦波的似然性的矩阵在时间方向上进行滤波，并且计算信号是正弦波的似然性的时间变化量。例如，通过一阶差分滤波器进行滤波。

[0152] 在步骤 S98 中，声学特征量提取器 54 对通过滤波获得的具有正弦波的似然性的时间变化量在时间轴方向上进行重新采样，并且将重新采样的结果视为声学特征量 IA2。在声学特征量提取器 54 将以这种方式从输入信号提取的声学特征量 IA2 提供给相似度计算器 112 时，对声学特征量 IA2 的提取处理终止。此后，处理进行到图 7 的步骤 S15。

[0153] 返回参考图 7 的流程图，在步骤 S15 中，参考信号剪切部分 81 剪切提供的参考信号并且将剪切的信号提供给时频转换器 82。

[0154] 在步骤 S16 中，时频转换器 82 对由参考信号剪切部分 81 提供的参考信号进行时频转换，并且将参考信号转换为对数 - 幅度频谱图，并且将该对数 - 幅度频谱图提供给声学特征量提取器 83 和声学特征量提取器 84。

[0155] 在步骤 S17 中，声学特征量提取器 83 进行对声学特征量 RA1 的提取处理，以计算参考信号的声学特征量 RA1，并且然后将计算的声学特征量 RA1 提供给匹配处理单元 23 的屏蔽图生成器 111。

[0156] 另外，在步骤 S18 中，声学特征量提取器 84 进行对声学特征量 RA2 的提取处理，以计算参考信号的声学特征量 RA2，并且然后将所计算的声学特征量 RA2 提供给匹配处理单元 23 的相似度计算器 112。

[0157] 此外，步骤 S17 和步骤 S18 的处理与步骤 S13 和步骤 S14 的处理相似，因此省略了对其的描述。然而，在步骤 S17 和步骤 S18 的处理中，待处理的信号是参考信号而不是输入信号。

[0158] 在步骤 S19 中，屏蔽图生成器 111 基于由声学特征量提取器 53 提供的声学特征量 IA1 和由声学特征量提取器 83 提供的声学特征量 RA1 来生成屏蔽图。例如，屏蔽图生成器 111 通过进行式（5）的计算来生成屏蔽图。

[0159] 在步骤 S20 中，屏蔽图生成器 111 计算声学特征量 A1 之间的相似度。例如，屏蔽图生成器 111 通过使用式（6）来计算声学特征量 A1 之间的相似度。屏蔽图生成器 111 将生成的屏蔽图以及声学特征量 A1 之间的相似度提供给相似度计算器 112。

[0160] 在步骤 S21 中，相似度计算器 112 基于由声学特征量提取器 54 提供的声学特征量 IA2、由声学特征量提取器 84 提供的声学特征量 RA2 以及由屏蔽图生成器 111 提供的屏蔽图和相似度来计算输入信号与参考信号之间的最终的相似度。

[0161] 例如，相似度计算器 112 通过进行式（7）的计算来计算输入信号与参考信号之间即输入信号的内容与参考信号的内容之间的相似度，并且将所计算的相似度和内容属性数据提供给比较积分器 113。

[0162] 在步骤 S22 中，比较积分器 113 基于由相似度计算器 112 提供的相似度来确定参考信号的内容与输入信号中所包括的内容是否彼此相同。

[0163] 例如，比较积分器 113 从针对多个参考信号获得的相似度之中指定超过预定阈值的最大的相似度，并且将具有指定的相似度的参考信号的内容视为输入信号的内容。比较积分器 113 输出以这种方式指定的输入信号的内容的内容属性数据和通过对内容识别的确定而获得的结果，然后终止了匹配检索。

[0164] 如以上所述，声音处理设备 11 根据输入信号和参考信号来计算指示信号是正弦波的似然性的声学特征量 A1，并且根据声学特征量 A1 生成屏蔽图。声音处理设备 11 基于屏蔽图和指示信号的个性特征的声学特征量 A2 来计算相似度。

[0165] 因此，在基于根据输入信号获得的声学特征量 IA1 和根据参考信号获得的声学特征量 RA1 生成屏蔽图时，可以获得作为对混响或混合噪声鲁棒的屏蔽图。因此，可以以较高的准确度对内容进行识别。

[0166] 以上所述的一系列处理可以由硬件执行，也可以由软件执行。在由软件执行一系列处理时，将构成这种软件的程序安装到计算机中。此处，表达“计算机”包括能够在安装有各种程序时执行各种功能的其中结合有专用硬件的计算机和通用个人计算机等。

[0167] 图 10 是示出根据程序执行较早描述的一系列处理的计算机的硬件的示例性配置的框图。

[0168] 在计算机中，中央处理单元（CPU）701、只读存储器（ROM）702 以及随机存取存储器（RAM）703 通过总线 704 相互连接。

[0169] 输入 / 输出接口 705 也连接到总线 704。输入单元 706、输出单元 707、记录单元 708、通信单元 709 以及驱动器 710 被连接到输入 / 输出接口 705。

[0170] 输入单元 706 由键盘、鼠标、麦克风、成像设备等配置。输出单元 707 由显示器、扬声器等配置。记录单元 708 由硬盘、非易失性存储器等配置。通信单元 709 由网络接口等配置。驱动器 710 驱动可移除介质 711，诸如磁盘、光盘、磁光盘、半导体存储器等。

[0171] 在如上所述配置的计算机中，CPU701 将例如在记录单元 708 中存储的程序经由输入 / 输出接口 705 和总线 704 加载到 RAM703 上，并且执行该程序。因此，进行了上述的一系列处理。

[0172] 待由计算机（CPU701）执行的程序被设置成记录在作为封装式介质等的可移除介质 711 中。此外，程序可以经由有线或无线传输媒介诸如局域网、因特网或数字卫星广播等被提供。

[0173] 在计算机中，通过将可移除介质 711 插入到驱动器 710 中，程序可以经由输入 / 输出接口 705 被安装在记录单元 708 中。此外，程序可以经由有线或无线传输媒介被通信单元 709 接收并且被安装在记录单元 708 中。此外，程序可以被预先安装在 ROM702 或记录单元 708 中。

[0174] 应当注意：由计算机执行的程序可以是以根据本说明书中所描述的顺序的时间序列被处理的程序或者被并行处理的程序或在必要时如在调用时被处理的程序。

[0175] 本领域的技术人员应当理解，取决于设计要求和其它因素可以发生各种修改、组合、子组合以及变化，只要所述修改、组合、子组合以及变化落在所附权利要求书及其等效方案的范围内即可。

[0176] 例如，本公开内容可以采用通过由多个装置经由网络进行分配并且连接一个功能进行处理的云计算的配置。

[0177] 此外，上述流程图中所描述的每个步骤都可以由一个装置或者通过分配多个装置而被执行。

[0178] 另外，在单个步骤中包括有多个处理的情况下，包括在该一个步骤中的多个处理可以由一个装置或者通过在多个装置之中共享而被执行。

[0179] 此外，本技术也可以被配置如下。

[0180] （1）一种声音处理设备，包括：

[0181] 输入信号处理单元，被配置为基于要识别内容的输入信号来计算第一声学特征量和与第一声学特征量不同的第二声学特征量，第一声学特征量指示在每个时频域中信号是正弦波的似然性；

[0182] 参考信号处理单元，被配置为基于预先准备的内容的参考信号来计算第一声学特征量和第二声学特征量；以及

[0183] 匹配处理单元，被配置为基于输入信号的第一声学特征量和第二声学特征量与参考信号的第一声学特征量和第二声学特征量来计算输入信号与参考信号之间的相似度。

[0184] （2）根据（1）所述的声音处理设备，其中，匹配处理单元基于输入信号的第一声学特征量和参考信号的第一声学特征量来生成屏蔽图，屏蔽图指示在每个时频域中信号是内容的似然性，并且匹配处理单元基于屏蔽图、第一声学特征量以及第二声学特征量来计算相似度。

[0185] （3）根据（2）所述的声音处理设备，其中，匹配处理单元还计算输入信号的第一声学特征量与参考信号的第一声学特征量之间的相似度，并且基于屏蔽图、第一声学特征量

之间的相似度以及第二声学特征量来计算输入信号与参考信号之间的相似度。

[0186] (4)根据(3)所述的声音处理设备,其中,匹配处理单元通过使参考信号对第一声学特征量之间的相似度的贡献率大于输入信号对第一声学特征量之间的相似度的贡献率来计算第一声学特征量之间的相似度。

[0187] (5)根据(1)至(4)中任一项所述的声音处理设备,其中,基于输入信号或参考信号的频谱图计算第二声学特征量,并且第二声学特征量在时间轴和频率轴上具有与第一声学特征量相同的粒度。

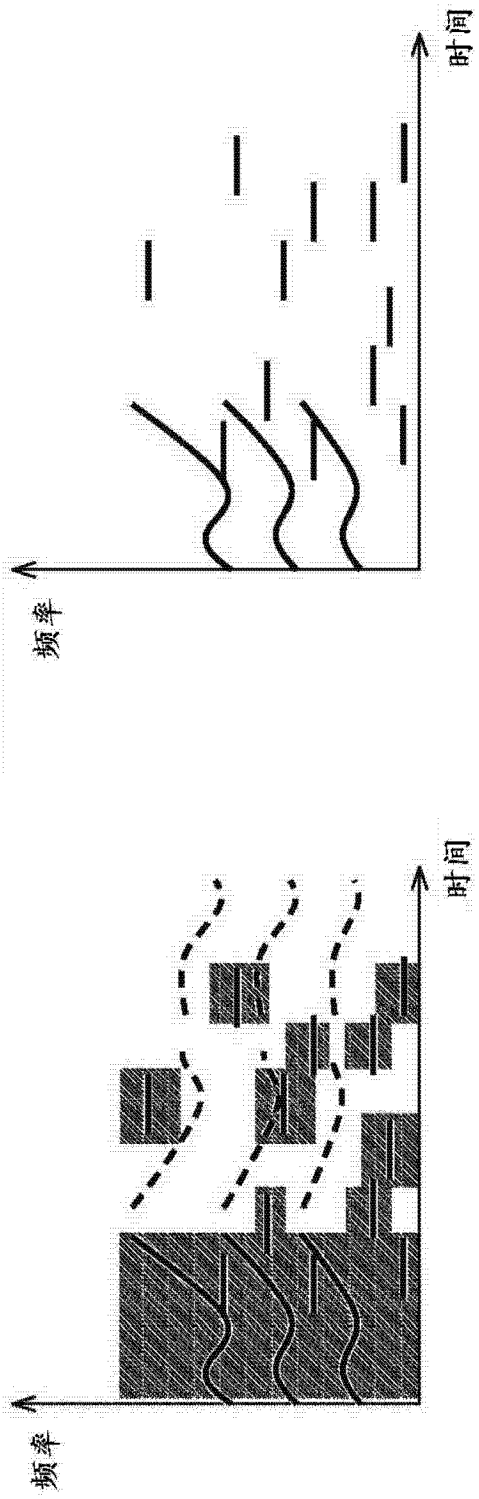

时间
频率
时间
频率

图 1

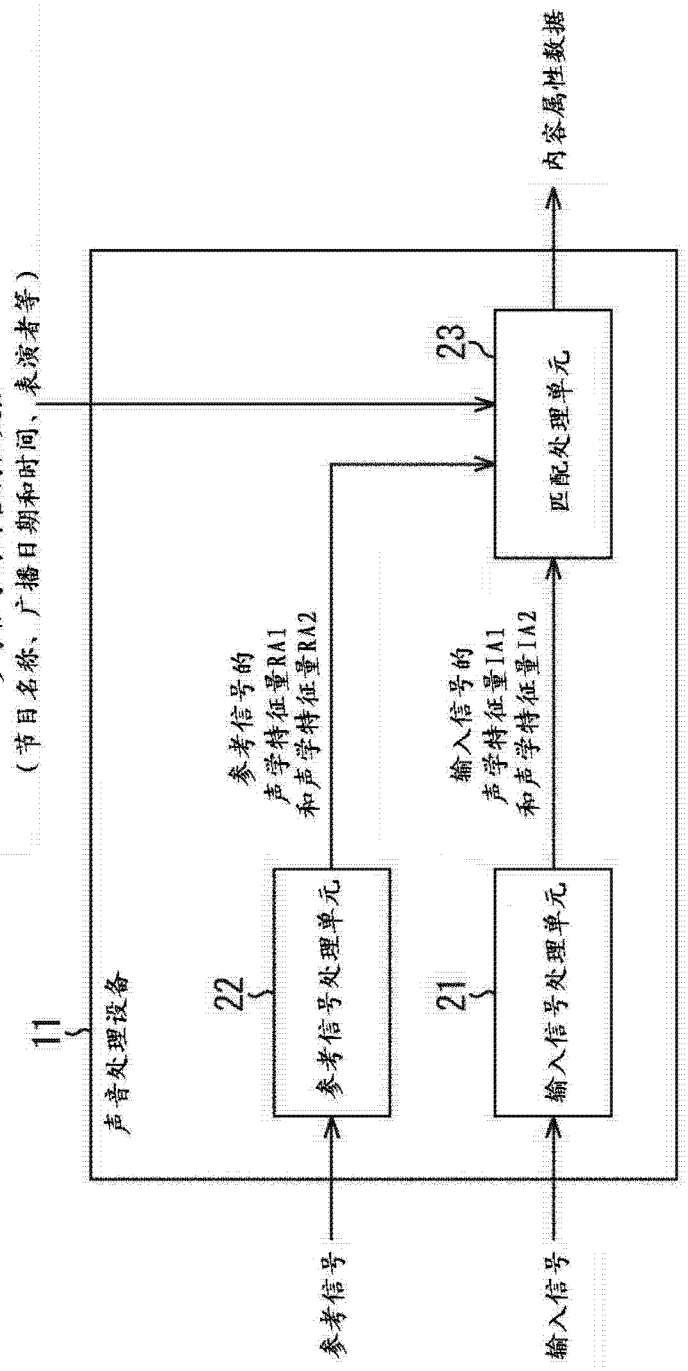

参考信号的内容属性数据
（节目名称、广播日期和时间、表演者等）
11
声音处理设备
22
参考信号处理单元
参考信号的
声学特征量RA1
和声学特征量RA2
21
输入信号处理单元
输入信号的
声学特征量IA1
和声学特征量IA2
23
匹配处理单元
内容属性数据
参考信号
输入信号

图 2

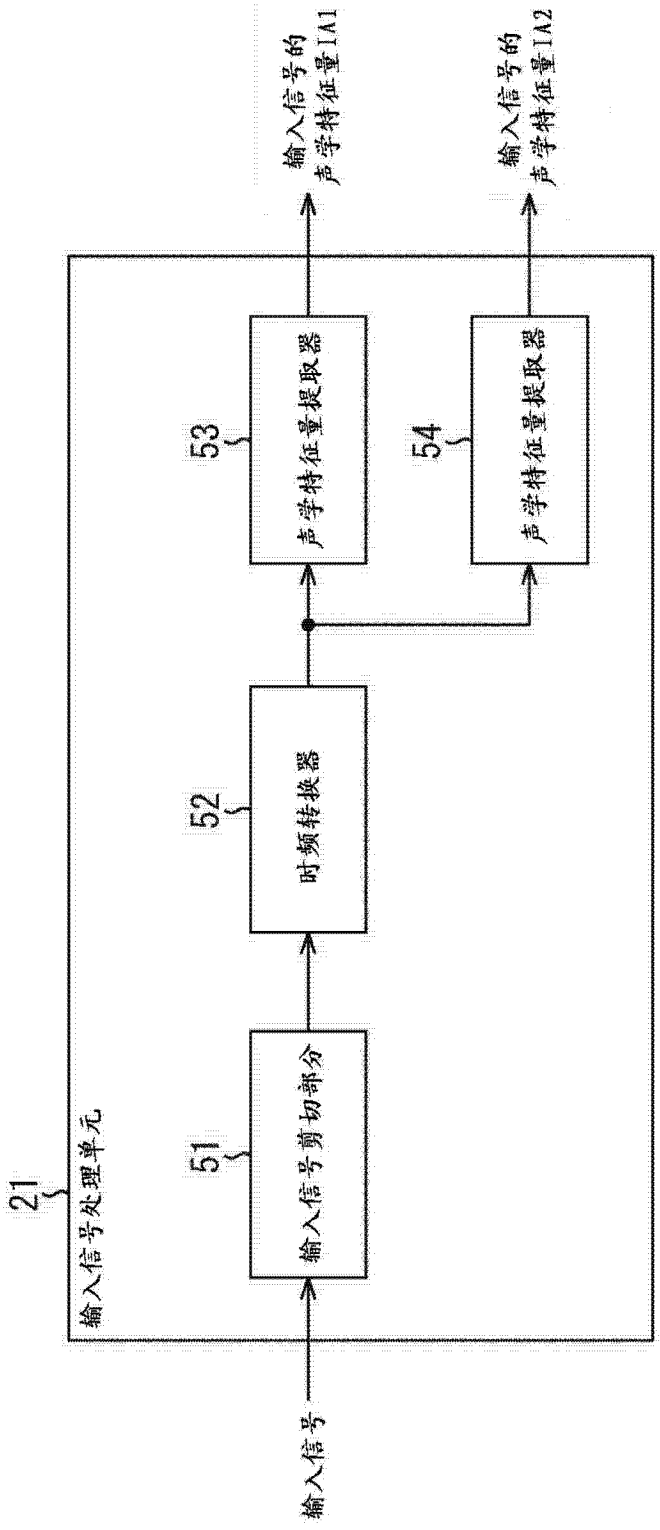

输入信号
21
输入信号处理单元
51
输入信号剪切部分
52
时频转换器
53
声学特征量提取器
54
声学特征量提取器
输入信号的
声学特征量IA1
输入信号的
声学特征量IA2

图 3

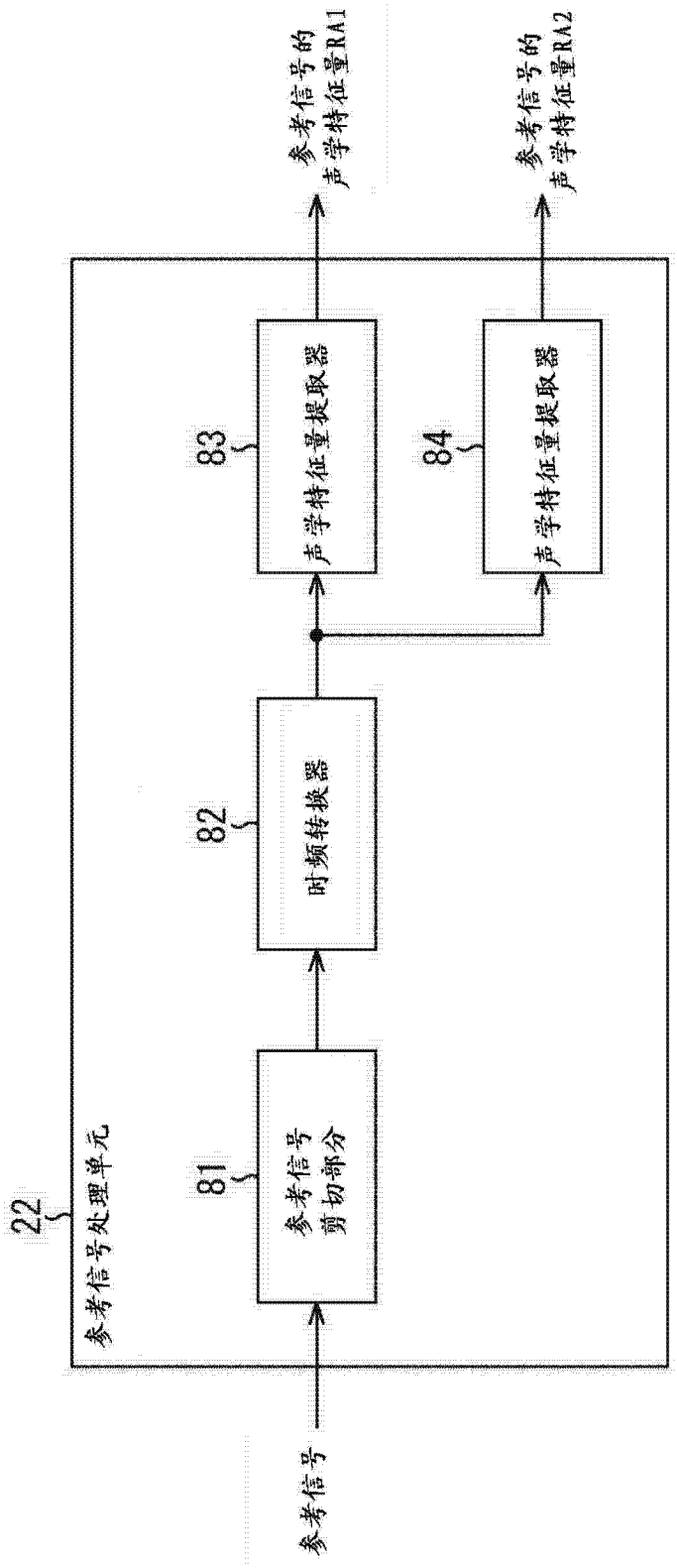

参考信号的
声学特征量RA1
参考信号的
声学特征量RA2
83
声学特征量提取器
84
声学特征量提取器
82
时频转换器
81
参考信号
剪切部分
22
参考信号处理单元
参考信号

图 4

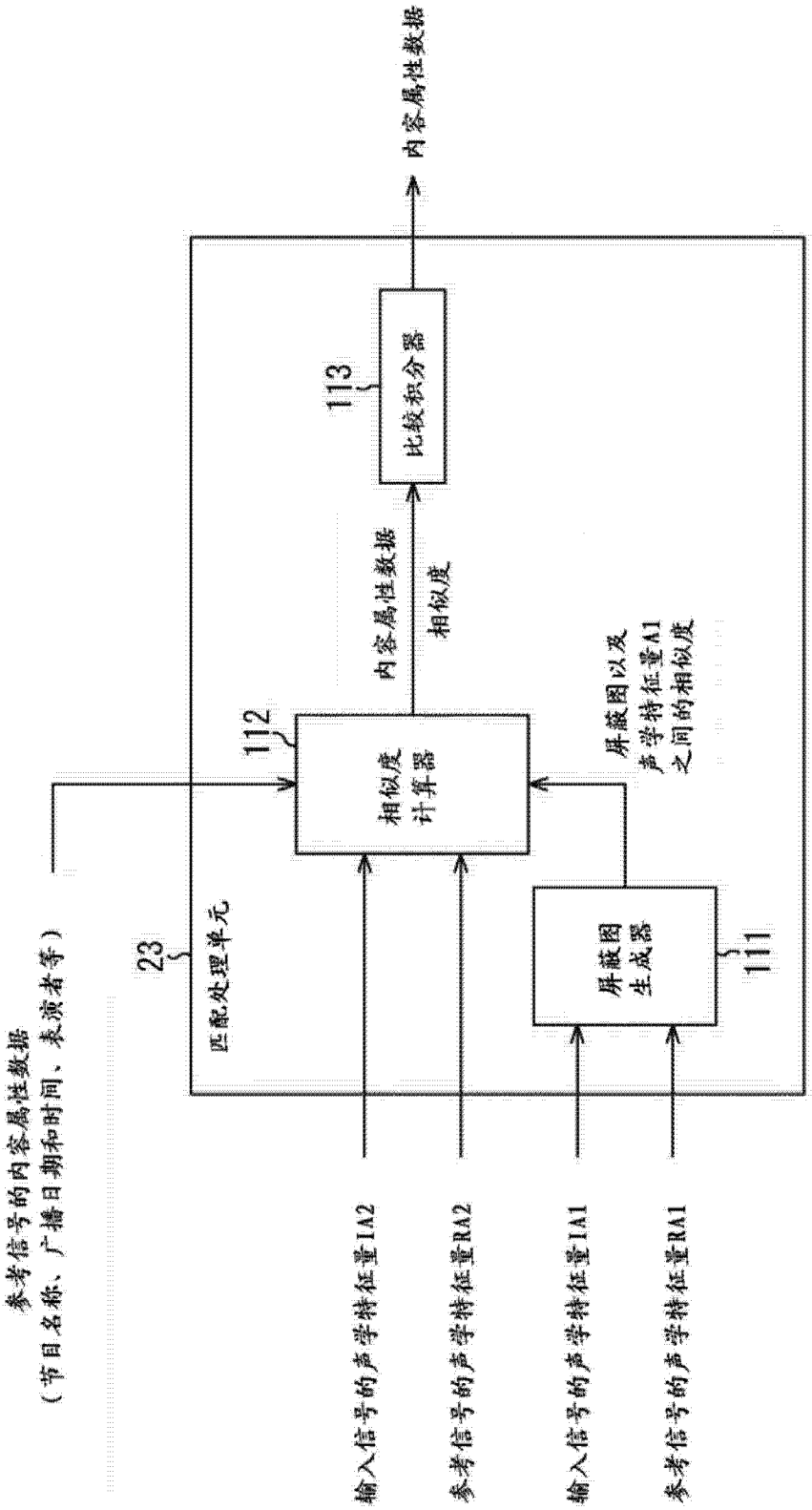

内容属性数据
比较积分器
113
内容属性数据
相似度
112
相似度
计算器
屏蔽图以及
声学特征量A1
之间的相似度
匹配处理单元
23
屏蔽图
生成器
111
参考信号的内容属性数据
（节目名称、广播日期和时间、表演者等）
输入信号的声学特征量IA2
参考信号的声学特征量RA2
输入信号的声学特征量IA1
参考信号的声学特征量RA1

图 5

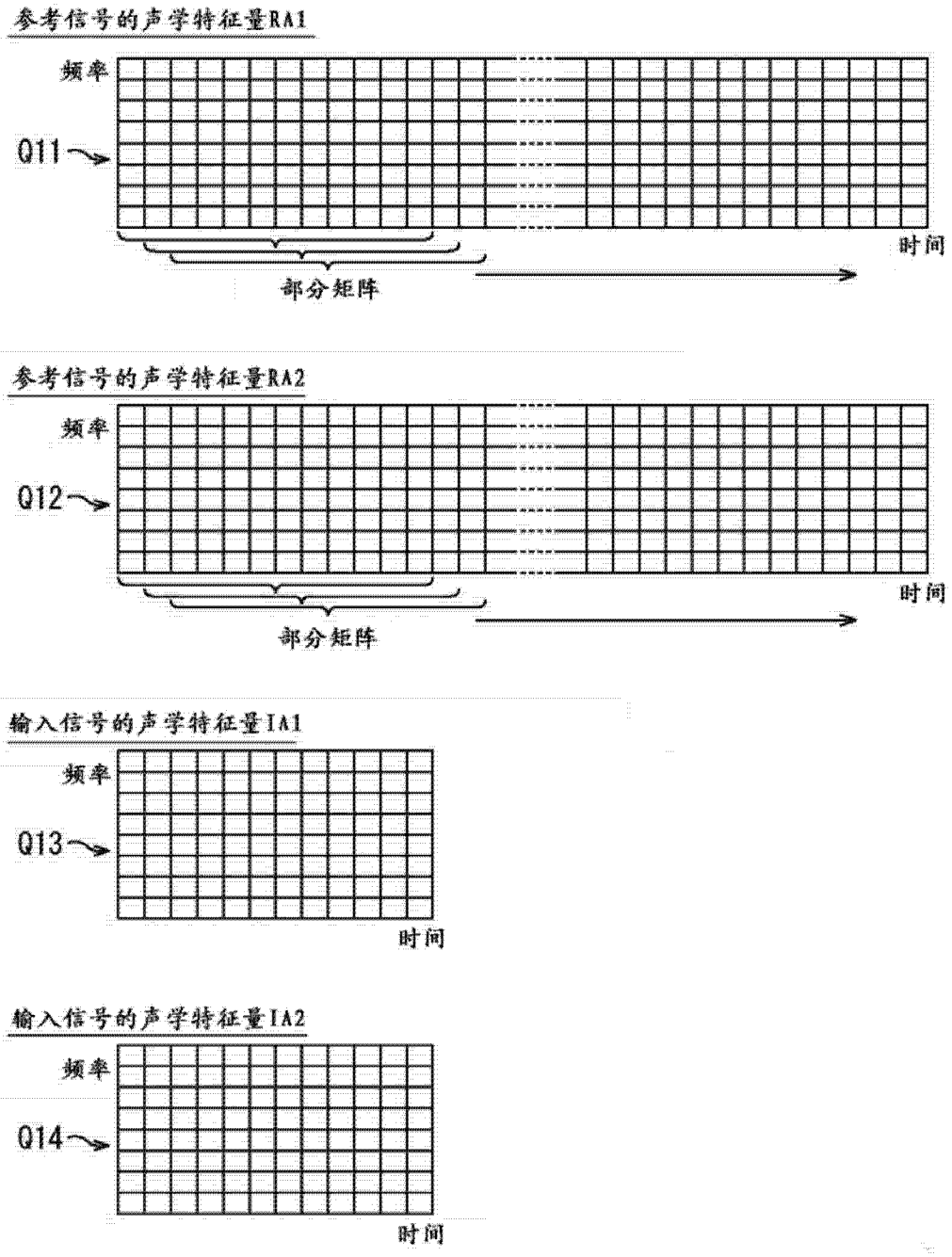

参考信号的声学特征量RA1
频率
Q11
时间
部分矩阵
参考信号的声学特征量RA2
频率
Q12
时间
部分矩阵
输入信号的声学特征量IA1
频率
Q13
时间
输入信号的声学特征量IA2
频率
Q14
时间

图 6

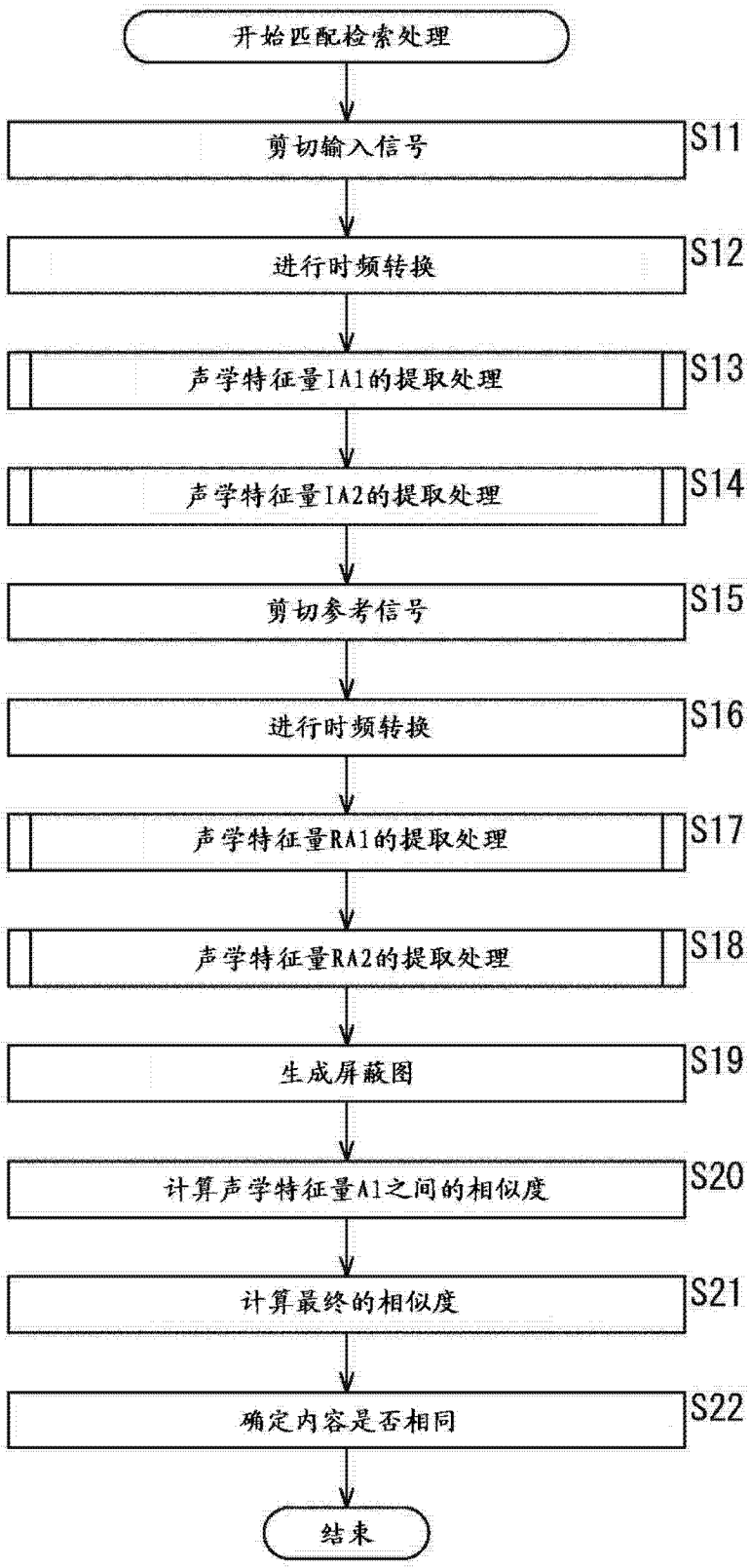

开始匹配检索处理
剪切输入信号
S11
进行时频转换
S12
声学特征量IA1的提取处理
S13
声学特征量IA2的提取处理
S14
剪切参考信号
S15
进行时频转换
S16
声学特征量RA1的提取处理
S17
声学特征量RA2的提取处理
S18
生成屏蔽图
S19
计算声学特征量A1之间的相似度
S20
计算最终的相似度
S21
确定内容是否相同
S22
结束

图 7

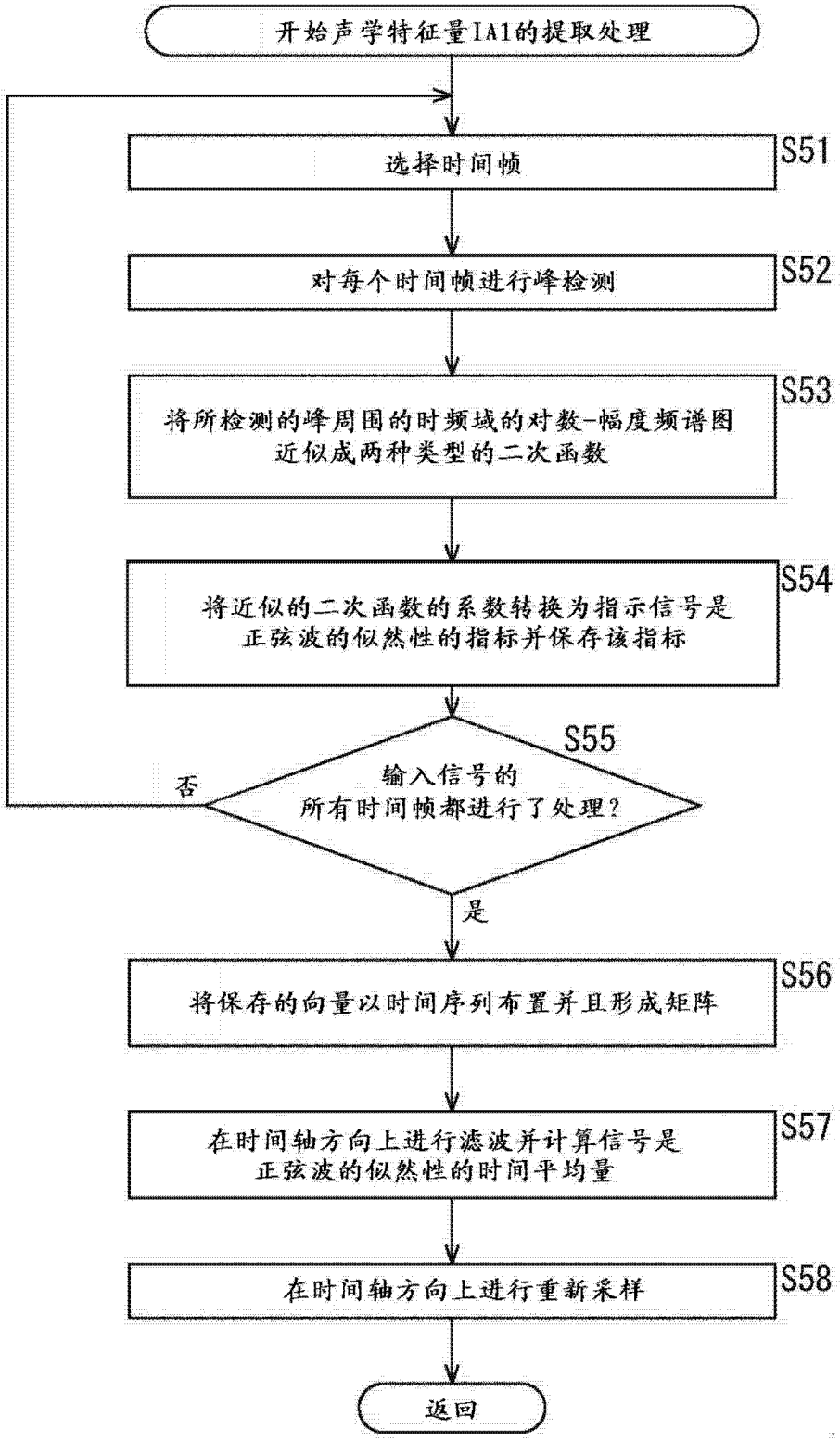

开始声学特征量IA1的提取处理
选择时间帧
S51
对每个时间帧进行峰检测
S52
将所检测的峰周围的时频域的对数-幅度频谱图
近似成两种类型的二次函数
S53
将近似的二次函数的系数转换为指示信号是
正弦波的似然性的指标并保存该指标
S54
S55
输入信号的
所有时间帧都进行了处理？
否
是
将保存的向量以时间序列布置并且形成矩阵
S56
在时间轴方向上进行滤波并计算信号是
正弦波的似然性的时间平均量
S57
在时间轴方向上进行重新采样
S58
返回

图 8

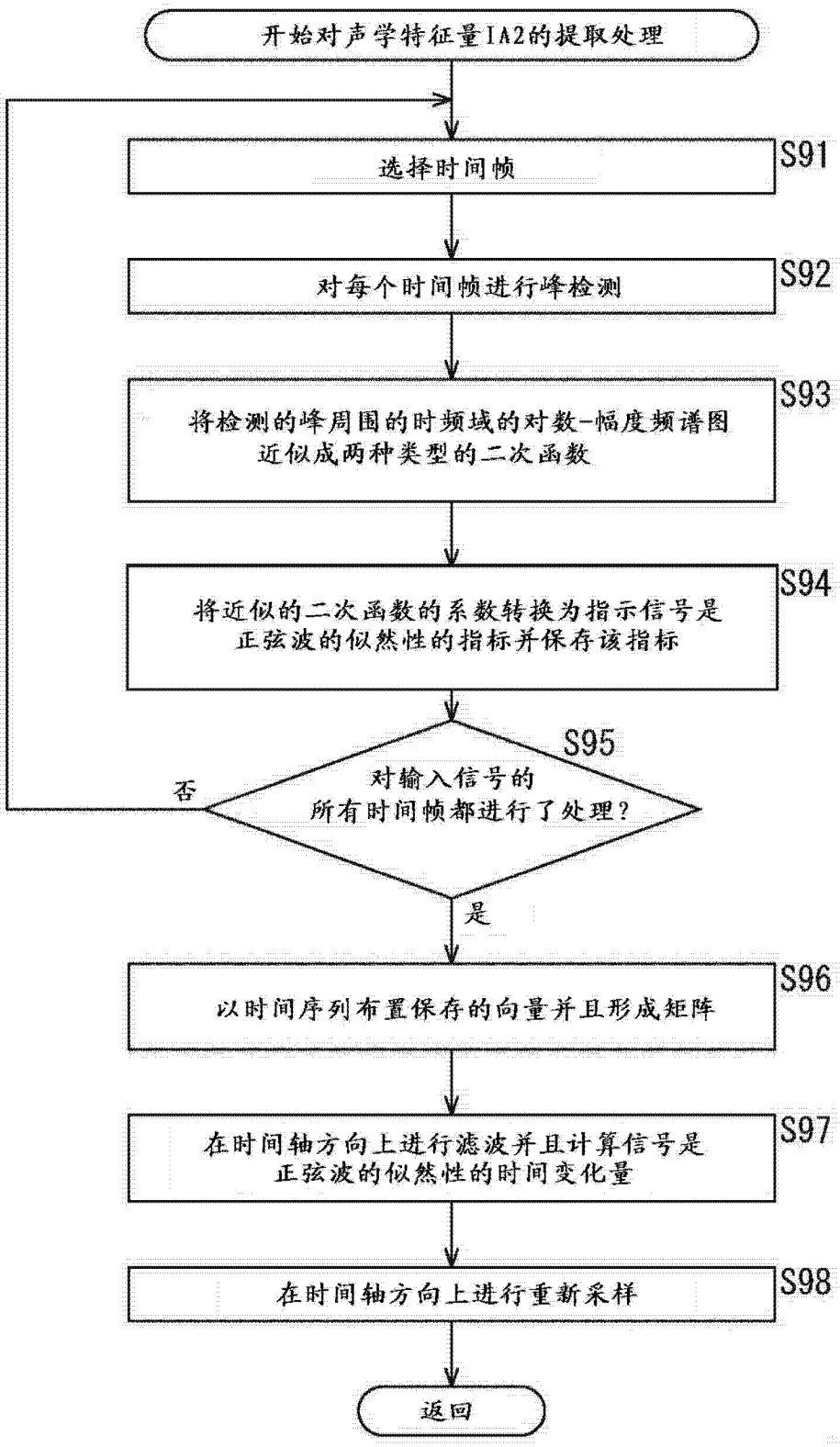

开始对声学特征量IA2的提取处理
选择时间帧
S91
对每个时间帧进行峰检测
S92
将检测的峰周围的时频域的对数-幅度频谱图
近似成两种类型的二次函数
S93
将近似的二次函数的系数转换为指示信号是
正弦波的似然性的指标并保存该指标
S94
对输入信号的
所有时间帧都进行了处理？
S95
否
是
以时间序列布置保存的向量并且形成矩阵
S96
在时间轴方向上进行滤波并且计算信号是
正弦波的似然性的时间变化量
S97
在时间轴方向上进行重新采样
S98
返回

图 9

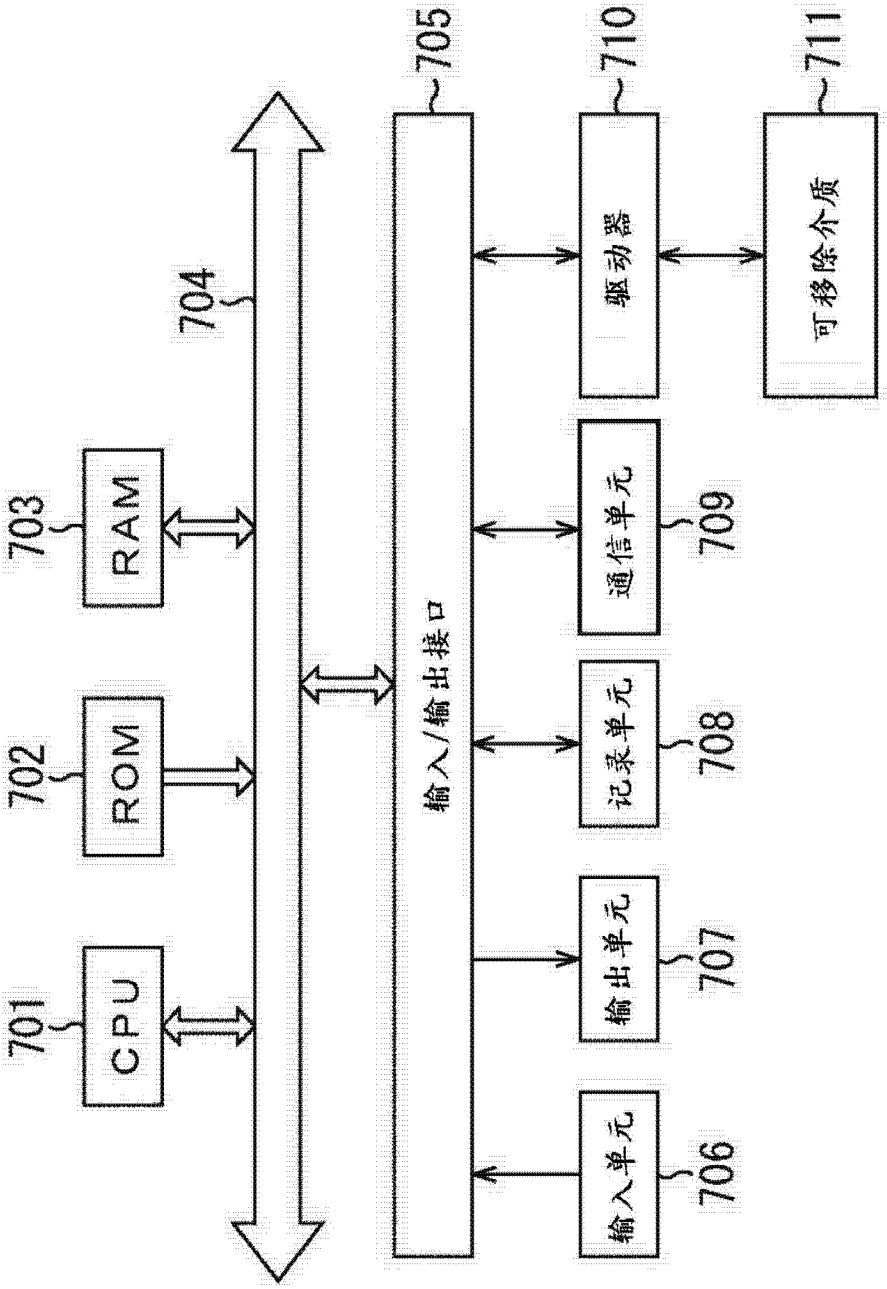

701
CPU
702
ROM
703
RAM
704
705
输入/输出接口
706
输入单元
707
输出单元
708
记录单元
709
通信单元
710
驱动器
711
可移除介质

图 10