



(12) 发明专利申请

(10) 申请公布号 CN 120303954 A

(43) 申请公布日 2025. 07. 11

(21) 申请号 202380083798.8

(74) 专利代理机构 永新专利商标代理有限公司
72002

(22) 申请日 2023.11.29

专利代理师 安香子

(30) 优先权数据

2022-198122 2022.12.12 JP

(51) Int.Cl.

H04R 3/02 (2006.01)

(85) PCT国际申请进入国家阶段日

2025.06.05

(86) PCT国际申请的申请数据

PCT/JP2023/042673 2023.11.29

(87) PCT国际申请的公布数据

W02024/127986 JA 2024.06.20

(71) 申请人 松下知识产权经营株式会社

地址 日本

(72) 发明人 横田健治 冈崎隆义

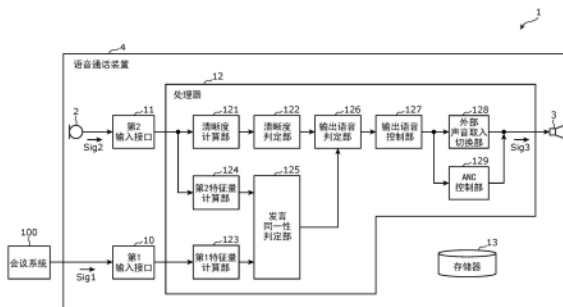
权利要求书1页 说明书15页 附图8页

(54) 发明名称

语音处理系统、语音处理方法以及程序

(57) 摘要

语音处理系统(1)具备第1输入I/F(10)、第2输入I/F(11)以及处理器(12)。第1输入I/F(10)取得经由通信线路的第1语音信号(Sig1)。第2输入I/F(11)取得基于由麦克风(2)收集到的语音的第2语音信号(Sig2)。处理器(12)将基于第1语音信号(Sig1)及第2语音信号(Sig2)的输出语音信号(Sig3)输出到扬声器(3)。处理器(12)在满足第1条件以及第2条件两者的情况下,将降低了与第1语音信号(Sig1)对应的分量的信号包含在输出语音信号(Sig3)中,第1条件为第1语音信号(Sig1)以及第2语音信号(Sig2)均包含基于同一人物发出的语音的语音信号,第2条件为第2语音信号(Sig2)清晰。



1. 一种语音处理系统,其中,具备:

第1输入接口,取得经由通信线路的第1语音信号;

第2输入接口,取得基于由麦克风收集到的语音的第2语音信号;以及

信号处理电路,将基于所述第1语音信号以及所述第2语音信号的输出语音信号输出到扬声器,

所述信号处理电路在满足第1条件以及第2条件两者的情况下,将减少了与所述第1语音信号对应的分量的信号包含在所述输出语音信号中,所述第1条件为所述第1语音信号以及所述第2语音信号均包含基于同一人物发出的语音的语音信号,所述第2条件为所述第2语音信号清晰。

2. 根据权利要求1所述的语音处理系统,其中,

所述信号处理电路在不满足所述第1条件以及所述第2条件中的至少一方的情况下,将所述第1语音信号包含在所述输出语音信号中,并且不将所述第2语音信号包含在所述输出语音信号中。

3. 根据权利要求1或2所述的语音处理系统,其中,

所述信号处理电路基于所述第1语音信号中的与元音对应的分量和所述第2语音信号中的与元音对应的分量的相关性,判定是否满足所述第1条件。

4. 根据权利要求1或2所述的语音处理系统,其中,

所述信号处理电路基于所述第2语音信号中的与元音对应的分量,判定是否满足所述第2条件。

5. 根据权利要求1或2所述的语音处理系统,其中,

所述信号处理电路在满足所述第1条件以及所述第2条件两者的情况下,将所述第2语音信号包含在所述输出语音信号中。

6. 根据权利要求2所述的语音处理系统,其中,

所述信号处理电路在不满足所述第1条件以及所述第2条件中的至少一方的情况下,还将与所述第2语音信号相反相位的语音信号包含在所述输出语音信号中。

7. 一种语音处理方法,其中,

取得经由通信线路的第1语音信号,

取得基于由麦克风收集到的语音的第2语音信号,

在满足第1条件以及第2条件两者的情况下,将降低了与所述第1语音信号对应的分量的信号包含在输出语音信号中并输出到扬声器,所述第1条件为所述第1语音信号以及所述第2语音信号均包含基于同一人物发出的语音的语音信号,所述第2条件为所述第2语音信号清晰。

8. 一种程序,使一个以上的处理器执行权利要求7所述的语音处理方法。

语音处理系统、语音处理方法以及程序

技术领域

[0001] 本公开涉及用于处理扬声器发出的语音的语音处理系统等。

背景技术

[0002] 例如,在专利文献1中公开了语音交流终端。该语音交流终端是对参与多点语音通信系统的多个终端中的至少一个的语音输出进行控制的装置,具有语音配置部和对话者管理部。语音配置部设定输出来自其他终端的语音时的音源配置。对话者管理部从多个终端中检测发言者与作为其对方的对话者,并基于检测出的发言者及对话者的组合来检测会话组。语音配置部根据检测到的会话组的变化来变更音源配置的设置。

[0003] 现有技术文献

[0004] 专利文献

[0005] 专利文献1:日本特开2012-108587号公报

发明内容

[0006] 发明要解决的课题

[0007] 本公开提供即使在相同的据点存在多个用户的环境中也不易损害会话的舒适性的语音处理系统等。

[0008] 用于解决课题的手段

[0009] 本公开的一个方式所涉及的语音处理系统具备第1输入接口、第2输入接口以及信号处理电路。所述第1输入接口取得经由通信线路的第1语音信号。所述第2输入接口取得基于由麦克风收集到的语音的第2语音信号。所述信号处理电路将基于所述第1语音信号以及所述第2语音信号的输出语音信号输出到扬声器。所述信号处理电路在满足第1条件以及第2条件两者的情况下,将降低了与所述第1语音信号对应的分量的信号包含在所述输出语音信号中,所述第1条件为所述第1语音信号以及所述第2语音信号均包含基于同一人物发出的语音的语音信号,所述第2条件为所述第2语音信号清晰。

[0010] 在本公开的一个方式所涉及的语音处理方法中,取得经由通信线路的第1语音信号。在所述语音处理方法中,取得基于由麦克风收集到的语音的第2语音信号。在所述语音处理方法中,在满足第1条件以及第2条件两者的情况下,将降低了与所述第1语音信号对应的分量的信号包含在输出语音信号中并输出到扬声器,所述第1条件为所述第1语音信号以及所述第2语音信号均包含基于同一人物发出的语音的语音信号,所述第2条件为所述第2语音信号清晰。

[0011] 本公开的一个方式所涉及的程序使一个以上的处理器执行所述语音处理方法。

[0012] 发明效果

[0013] 根据本公开的语音处理系统等,具有如下优点:即使在离线会话和在线会话双方混合存在的环境中,也不易损害会话的舒适性。

附图说明

- [0014] 图1是使用会议系统的交流的问题的说明图。
- [0015] 图2是表示包括实施方式所涉及的语音处理系统的整体结构的一例的框图。
- [0016] 图3是用于判定第2语音信号的清晰度的第1判定动作的说明图。
- [0017] 图4是用于判定第2语音信号的清晰度的第2判定动作的说明图。
- [0018] 图5是表示实施方式所涉及的语音处理系统的动作例的流程图。
- [0019] 图6是表示为了判定第2语音信号的清晰度而需要的参数的计算例的流程图。
- [0020] 图7是表示为了判定发言同一性而需要的参数的计算例的流程图。
- [0021] 图8是实施方式所涉及的语音处理系统的动作例的概要的说明图。
- [0022] 图9是实施方式所涉及的语音处理系统的优点的说明图。

具体实施方式

[0023] [1. 成为本公开的基础的见解]

[0024] 首先,以下说明发明人的着眼点。

[0025] 以往,已知例如使用经由MCU (Multipoint Control Unit:多地点控制装置)的会议系统、或者Zoom (注册商标)等Web会议服务,在多个据点间同时进行会议等交流的技术。在这样的交流中,各参与者通过佩戴具备麦克风以及扬声器的设备(例如,耳麦等),与其他参与者之间进行会话。另外,近年来,各参与者通过佩戴使用了XR(x Reality)技术的设备(例如,头戴式显示器或者智能眼镜等),也能够相同的虚拟空间上或者一边观察相同的虚拟空间一边与其他参与者之间进行会话。在这样的交流中,有时在相同的据点存在多个参与者,存在如下课题。

[0026] 图1是使用会议系统的交流的问题的说明图。在图1中,会议系统100是经由上述MCU的会议系统或Web会议服务所提供的服务器。在图1所示的例子中,表示处于第1据点A1的两个用户U1、U2、处于第2据点A2的用户U3以及处于第3据点A3的用户U4使用会议系统100在线进行会议。用户U1~U4均通过从扬声器输出基于经由会议系统100发送的语音信号的语音,能够听到其他用户发出的语音。作为一例,在用户U1发出了“你好”这样的语音的情况下,其他用户U2、U3、U4通过从扬声器输出基于经由会议系统100发送的语音信号的该语音,能够听到“你好”这样的由用户U1发出的语音。

[0027] 在此,在图1所示的例子中,在第1据点A1存在两个用户U1、U2。因此,在第1据点A1,两个用户U1、U2中的一个用户发出的语音能够不经由会议系统100而直接被另一个用户听到。在该情况下,例如在处于第1据点A1的用户U2发出了“你好”这样的语音V2的情况下,用户U1直接听到用户U2发出的语音V2,并且听到经由会议系统100发送的用户U2发出的语音V1。

[0028] 如上所述,在相同的据点存在多个用户的情况下,处于该据点的用户听到来自处于该据点的其他用户的直接语音和经由会议系统100的语音双方,因此存在难以听取其他用户发出的语音的问题。另外,来自处于该据点的其他用户的经由会议系统100的语音比来自其他用户的直接语音晚到达用户的耳朵。因此,存在如下问题:在用户听到来自其他用户的直接语音而要发言的定时,来自其他用户的经由会议系统100的语音到达用户的耳朵,从而用户的发言被妨碍而难以说话。这样,存在如下问题:在相同的据点存在多个用户的环境

中,容易损害会话的舒适性。

[0029] 鉴于以上内容,发明人完成了本公开。

[0030] 以下,参照附图对实施方式进行具体说明。此外,以下说明的实施方式均表示总括性或具体的例子。以下的实施方式所示的数值、形状、材料、构成要素、构成要素的配置位置以及连接方式、步骤、步骤的顺序等是一个例子,并非旨在限定本公开。另外,关于以下的实施方式中的构成要素中的未记载于独立权利要求的构成要素,作为任意的构成要素进行说明。

[0031] 此外,各图是示意图,不一定严格地图示。另外,在各图中,对实质上相同的结构标注相同的附图标记,有时省略或简化重复的说明。

[0032] (实施方式)

[0033] [2. 结构]

[0034] [2-1. 整体结构]

[0035] 首先,使用图2对包含实施方式所涉及的语音处理系统的整体结构进行说明。图2是表示包含实施方式所涉及的语音处理系统的整体结构的框图。语音处理系统1是用于在从外部取得了语音信号的情况下从扬声器3输出基于语音信号的语音的系统。在实施方式中,语音处理系统1由语音通话装置4实现。

[0036] 语音通话装置4能够经由因特网等网络与会议系统100进行通信。此外,语音通话装置4也可以经由LAN(Local Area Network:局域网)与会议系统100进行通信。

[0037] 语音通话装置4是佩戴于用户的头部或颈部的设备,分为密闭型的语音通话装置、开放型的语音通话装置、以及能够切换密闭型和开放型的语音通话装置。密闭型的语音通话装置是覆盖用户的耳孔(鼓膜)的设备,例如包括入耳式耳机(Earphone)型的耳麦(Headset)或头戴式耳机(Headphone)型的耳麦等。开放型的语音通话装置是不覆盖用户的耳孔的设备,例如包括颈挂式扬声器或XR用的护目镜型的可穿戴设备等。能够切换密闭型和开放型的语音通话装置是能够切换覆盖用户的耳孔的功能和不覆盖用户的耳孔的功能这两个功能的设备,例如包括能够通过壳体部的板开闭来进行切换的入耳式耳机型的耳麦、或者头戴式耳机型的耳麦等。此外,语音通话装置4中,执行语音处理等的主体部、以及包含麦克风及扬声器的耳麦部既可以一体地构成,也可以分别分体地构成。

[0038] 语音处理系统1能够应用于密闭型的语音通话装置、开放型的语音通话装置、以及能够切换密闭型和开放型的语音通话装置中的任一种。以下,作为一例,说明语音通话装置4是密闭型的语音通话装置的情况。

[0039] 如上所述,会议系统100例如是经由MCU的会议系统或Web会议服务所提供的服务器。会议系统100如果接收到从任意用户佩戴的语音通话装置4输出的语音信号,则对接收到的语音信号执行适当的修正处理,然后将修正后的语音信号发送给一个以上的其他用户分别佩戴的一个以上的语音通话装置4。修正处理可以包括例如用于降低接收到的语音信号中包括的噪声的噪声抑制处理。另外,修正处理例如可以包括对接收到的语音信号中的人的可听频域进行强调的频率修正处理。此外,会议系统100也可以不对接收到的语音信号执行修正处理。

[0040] [2-2. 语音通话装置(语音处理系统)的结构]

[0041] 接着,对语音通话装置4(语音处理系统1)的结构进行具体说明。如图2所示,语音

通话装置4具备麦克风2、第1输入接口(以下,记载为“第1输入I/F(Interface)”)10、第2输入接口(以下,记载为“第2输入I/F”)11、处理器12、存储器13以及扬声器3。

[0042] 麦克风2是取得语音通话装置4的周围的语音并基于取得的语音输出第2语音信号Sig2的收音设备。具体而言,麦克风2是电容式麦克风、动态麦克风或MEMS(Micro Electro Mechanical Systems:微机电系统)麦克风等,但没有特别限定。另外,麦克风可以是无指向性的,也可以具有指向性。

[0043] 扬声器3输出基于从处理器12输出的输出语音信号Sig3的语音。扬声器3是朝向佩戴了语音通话装置4的用户的耳孔发出声波的扬声器,例如可以是骨传导扬声器。

[0044] 第1输入I/F10例如是无线通信接口,基于Wi-Fi(注册商标)等无线通信标准,经由网络与会议系统100进行通信,由此接收从会议系统100发送的第1语音信号Sig1。换言之,第1输入I/F10经由通信线路取得第1语音信号Sig1。第1语音信号Sig1是主要基于其他用户发出的语音的语音信号。第1输入I/F10将所取得的第1语音信号Sig1输出到处理器12。

[0045] 第2输入I/F11是接收从麦克风2输出的第2语音信号Sig2的接口。换言之,第2输入I/F11取得基于由麦克风2收集到的语音的第2语音信号Sig2。第2输入I/F11将所取得的第2语音信号Sig2输出到处理器12。

[0046] 处理器12例如是CPU(Central Processing Unit:中央处理器)或DSP(Digital Signal Processor:数字信号处理器)等。处理器12进行将基于由第1输入I/F10取得的第1语音信号Sig1以及由第2输入I/F11取得的第2语音信号Sig2的输出语音信号Sig3输出到扬声器3的信息处理。上述的信息处理通过由处理器12执行存储于存储器13的计算机程序来实现。处理器12是语音处理系统1的信号处理电路的一例。

[0047] 在处理器12中,作为功能性的构成要素,包括清晰度计算部121、清晰度判定部122、第1特征量计算部123、第2特征量计算部124、发言同一性判定部125、输出语音判定部126、输出语音控制部127、外部声音取入切换部128以及ANC(Active Noise Cancelling:主动噪声消除)控制部129。上述的各功能例如通过由处理器12执行存储于存储器13的计算机程序来实现。

[0048] 清晰度计算部121计算在由清晰度判定部122判定第2语音信号Sig2是否清晰时使用的第2语音信号Sig2的特征量。在此,第2语音信号Sig2清晰是指,第2语音信号Sig2中的与人的语音对应的频带(以下,称为“语音频带”)的SNR(Signal to Noise Ratio:信噪比)比阈值高且人的语音的特征明确。换言之,第2语音信号Sig2清晰是指,在人听到从扬声器3输出的基于第2语音信号Sig2的语音的情况下,该人能够理解其内容的程度。

[0049] 清晰度计算部121计算上述SNR和第2语音信号Sig2的频谱包络,作为第2语音信号Sig2的特征量。具体而言,清晰度计算部121通过对第2语音信号Sig2进行适当的信号处理,计算第2语音信号Sig2的频谱对比度,并基于计算出的频谱对比度,计算第2语音信号Sig2中的语音频带的SNR。另外,清晰度计算部121计算第2语音信号Sig2的MFCC(Mel-Frequency Cepstral Coefficient,梅尔频率倒谱系数)。MFCC是在语音识别等中用作特征量的倒谱的系数,通过将使用梅尔滤波器组进行压缩后的功率谱变换为对数功率谱,并对对数功率谱应用逆离散余弦变换而得到。MFCC相当于频谱包络。

[0050] 清晰度判定部122使用由清晰度计算部121计算出的第2语音信号Sig2的特征量,判定是否满足第2语音信号Sig2清晰这一第2条件。关于清晰度判定部122的判定动作,在后

述的[2-3.清晰度的判定]中详细说明。

[0051] 第1特征量计算部123计算在由发言同一性判定部125判定第1语音信号Sig1以及第2语音信号Sig2是否都是基于同一人物发出的语音的语音信号时使用的第1语音信号Sig1的特征量。第1特征量计算部123计算第1语音信号Sig1的基本频率和第1语音信号Sig1的频谱包络作为第1语音信号Sig1的特征量。具体而言,第1特征量计算部123计算第1语音信号Sig1的倒谱,根据计算出的倒谱计算第1语音信号Sig1的基本频率。倒谱是通过应用傅里叶变换来计算第1语音信号Sig1的功率谱,将计算出的功率谱变换为对数功率谱,并对对数功率谱进一步应用傅里叶变换而得到的。另外,第1特征量计算部123通过计算第1语音信号Sig1的MFCC来计算频谱包络。另外,第1特征量计算部123根据计算出的频谱包络,计算第1语音信号Sig1中元音出现的定时。

[0052] 第2特征量计算部124计算在由发言同一性判定部125判定第1语音信号Sig1以及第2语音信号Sig2是否都是基于同一人物发出的语音的语音信号时使用的第2语音信号Sig2的特征量。第2特征量计算部124计算第2语音信号Sig2的基本频率和第2语音信号Sig2的频谱包络,作为第2语音信号Sig2的特征量。具体而言,第2特征量计算部124计算第2语音信号Sig2的倒谱,根据计算出的倒谱计算第2语音信号Sig2的基本频率。另外,第2特征量计算部124通过计算第2语音信号Sig2的MFCC来计算频谱包络。另外,第2特征量计算部124根据计算出的频谱包络,计算第2语音信号Sig2中元音出现的定时。

[0053] 另外,第2语音信号Sig2的频谱包络仅由清晰度计算部121和第2特征量计算部124中的任意一方计算即可。在实施方式中,设为由第2特征量计算部124计算第2语音信号Sig2的频谱包络而进行说明。因此,清晰度计算部121也可以不计算第2语音信号Sig2的频谱包络。另外,在仅由清晰度计算部121和第2特征量计算部124中的某一方计算第2语音信号Sig2的频谱包络的情况下,计算出的频谱包络由另一方共享。

[0054] 发言同一性判定部125使用由第1特征量计算部123计算出的第1语音信号Sig1的特征量和由第2特征量计算部124计算出的第2语音信号Sig2的特征量,判定是否满足第1条件,该第1条件为第1语音信号Sig1以及第2语音信号Sig2均包含基于同一人物发出的语音的语音信号。在实施方式中,发言同一性判定部125在(i)第1语音信号Sig1的基本频率与第2语音信号Sig2的基本频率相同、且(ii)第1语音信号Sig1中元音出现的定时及种类与第2语音信号Sig2中元音出现的定时及种类相同的情况下,判定为满足第1条件。另一方面,发言同一性判定部125在不满足上述的(i)及(ii)中的至少一方的情况下,判定为不满足第1条件。

[0055] 如上所述,在各语音信号中元音出现的定时能够根据各语音信号的频谱包络来检测。另外,第1语音信号Sig1经由通信线路,因此比第2语音信号Sig2延迟地由第1输入I/F10取得。因此,发言同一性判定部125考虑上述的延迟来对(ii)进行判定。

[0056] 这样,处理器12(发言同一性判定部125)基于第1语音信号Sig1中的与元音对应的分量和第2语音信号Sig2中的与元音对应的分量之间的相关性,来判定是否满足第1条件。具体而言,发言同一性判定部125在(i)计算第1语音信号Sig1的基本频率与第2语音信号Sig2的基本频率的差分,计算出的差分为阈值以下、且(ii)计算第1语音信号Sig1中元音出现的时间与第2语音信号Sig2中元音出现的时间的差分,计算出的差分为阈值以下、且第1语音信号Sig1中出现的元音的种类与第2语音信号Sig2中出现的元音的种类相同的情况

下,判定为满足第1条件。另一方面,发言同一性判定部125在(i)或(ii)中计算出的差分超过阈值的情况下,判定为不满足第1条件。此外,发言同一性判定部125也可以在(ii)的判定中,计算在第1语音信号Sig1中计算出的频谱包络与在第2语音信号Sig2中计算出的频谱包络的相关系数,判定计算出的相关系数是否为阈值以下。另外,发言同一性判定部125也可以在(ii)的判定中满足某一个条件的情况下,设为满足(ii)。

[0057] 此外,发言同一性判定部125可以仅根据是否满足(i)来判定是否满足第1条件,也可以仅根据是否满足(ii)来判定是否满足第1条件。另外,发言同一性判定部125也可以在满足(i)以及(ii)的至少一方的情况下判定为满足第1条件,在不满足(i)以及(ii)双方的情况下判定为不满足第1条件。

[0058] 另外,发言同一性判定部125也可以取代(ii),判定第1语音信号Sig1中元音连续出现的样式与第2语音信号Sig2中元音连续出现的样式是否相同。在该情况下,发言同一性判定部125也可以不考虑上述的延迟。

[0059] 在此,还考虑如下方法:基于第1语音信号Sig1的波形与第2语音信号Sig2的波形的相似度,如果波形的相似度为阈值以上则判定为满足第1条件,如果波形的相似度小于阈值则判定为不满足第1条件。这里所说的“波形”是信号的振幅的波形、即声压级的波形。然而,如上所述,第1语音信号Sig1是在会议系统100中执行修正处理后的语音信号,因此第1语音信号Sig1的波形与第2语音信号Sig2的波形会不同。因此,如上所述,发言同一性判定部125通过与基于波形的相似度的方法不同的方法来判定是否满足第1条件。

[0060] 此外,如果是在会议系统100中不对第1语音信号Sig1执行修正处理的情况,则发言同一性判定部125也可以基于波形的相似度来判定是否满足第1条件。例如,发言同一性判定部125也可以基于第1语音信号Sig1的声压级的变化与第2语音信号Sig2的声压级的变化是否大致一致,换言之,基于第1语音信号Sig1的振幅的包络线与第2语音信号Sig2的振幅的包络线的相关性,来判定是否满足第1条件。

[0061] 输出语音判定部126基于清晰度判定部122中的是否满足第2条件的判定结果、以及发言同一性判定部125中的是否满足第1条件的判定结果,判定处于第1状况以及第2状况中的哪一个状况。第1状况是用户与其他用户之间的距离比较近、用户容易直接听取其他用户发出的语音的状况。第2状况是第1状况以外的状况。第2状况例如包括用户与其他用户之间的距离比较远、用户难以直接听取其他用户发出的语音的状况。输出语音判定部126在满足第1条件及第2条件两者的情况下,判定为处于第1状况。另一方面,输出语音判定部126在不满足第1条件及第2条件中的至少一方的情况下,判定为处于第2状况。

[0062] 输出语音控制部127基于输出语音判定部126的判定结果,控制输出语音信号Sig3中包含的语音信号。具体而言,输出语音控制部127在输出语音判定部126中判定为处于第1状况的情况下,以降低从扬声器3输出的基于第1语音信号Sig1的语音的音量的方式进行控制。这里所说的“降低基于第1语音信号Sig1的语音的音量”是指使基于第1语音信号Sig1的语音的音量成为比第2状况下的基于第1语音信号Sig1的语音的音量(换言之,默认的音量)低的音量。另外,输出语音控制部127通过控制外部声音取入切换部128来开启外部声音取入功能,并且通过控制ANC控制部129来关闭噪声消除功能。

[0063] 即,处理器12(输出语音控制部127)在满足第1条件及第2条件两者的情况下,将降低了与第1语音信号Sig1对应的的分量的信号包含在输出语音信号Sig3中。此外,在实施方式

中,处理器12通过降低基于第1语音信号Sig1的语音的音量来减少与第1语音信号Sig1对应的分量,但不限于此。例如,处理器12也可以在满足第1条件以及第2条件两者的情况下,不将第1语音信号Sig1包含在输出语音信号Sig3中。此外,例如也可以是,处理器12在满足第1条件和第2条件两者的情况下,基于第2语音信号Sig2对第1语音信号Sig1执行抑制处理,并且将处理后的第1语音信号Sig1包含在输出语音信号Sig3中。

[0064] 另外,处理器12(输出语音控制部127)在满足第1条件以及第2条件两者的情况下,开启外部声音取入功能,即,将第2语音信号Sig2包含在输出语音信号Sig3中。另外,输出语音信号Sig3中包含的第2语音信号Sig2也可以是通过降噪处理或均衡处理等语音处理进行处理后的信号。

[0065] 另一方面,输出语音控制部127在输出语音判定部126中判定为处于第2状况的情况下,以将从扬声器3输出的基于第1语音信号Sig1的语音的音量设为默认的音量的方式进行控制。另外,输出语音控制部127通过控制外部声音取入切换部128来关闭外部声音取入功能,并且通过控制ANC控制部129来开启噪声消除功能。

[0066] 即,处理器12(输出语音控制部127)在不满足第1条件以及第2条件中的至少一方的情况下,将第1语音信号Sig1包含在输出语音信号Sig3中,并且关闭外部声音取入功能,即不将第2语音信号Sig2包含在输出语音信号Sig3中。另外,输出语音信号Sig3中包含的第1语音信号Sig1也可以是通过降噪处理或均衡处理等语音处理进行处理后的信号。

[0067] 另外,处理器12(输出语音控制部127)在不满足第1条件以及第2条件中的至少一方的情况下,开启噪声消除功能,即进一步将与第2语音信号Sig2相反相位的语音信号包含在输出语音信号Sig3中。

[0068] 外部声音取入切换部128通过被输出语音控制部127控制,来切换取入用户的周围的语音的外部声音取入功能的开启/关闭。在外部声音取入功能开启的情况下,扬声器3输出基于包含第2语音信号Sig2的输出语音信号Sig3的语音。另一方面,在外部声音取入功能关闭的情况下,扬声器3输出基于不包含第2语音信号Sig2的输出语音信号Sig3的语音。

[0069] ANC控制部129通过被输出语音控制部127控制来切换噪声消除功能的开启/关闭。在噪声消除功能开启的情况下,ANC控制部129生成与第2语音信号Sig2相反相位的语音信号,并将生成的语音信号包含在输出语音信号Sig3中。该情况下,扬声器3输出基于与第2语音信号Sig2相反相位的语音信号的语音。由此,在用户的耳边,作为第2语音信号Sig2的来源的语音、和基于与第2语音信号Sig2相反相位的语音信号的语音相互抵消,所以用户大致无法听到这些语音。另一方面,在噪声消除功能关闭的情况下,ANC控制部129不生成与第2语音信号Sig2相反相位的语音信号。

[0070] 存储器13是存储处理器12执行的计算机程序以及实现各种功能所需的信息等的存储装置。存储器13例如由半导体存储器等实现。存储器13可以被实现为处理器12的内置存储器,而不是处理器12的外部存储器。

[0071] [2-3.清晰度的判定]

[0072] 以下,对由清晰度判定部122进行的第2语音信号Sig2是否清晰的判定动作进行详细说明。在实施方式中,清晰度判定部122执行第1判定动作以及第2判定动作,在任一判定动作中均判定为清晰的情况下,判定为第2语音信号Sig2清晰、即满足第2条件。另一方面,清晰度判定部122在第1判定动作以及第2判定动作中的至少一方中判定为不清晰的情况

下,判定为第2语音信号Sig2不清晰、即不满足第2条件。

[0073] 图3是用于判定第2语音信号Sig2的清晰度的第1判定动作的说明图。图3示出了第2语音信号Sig2的频谱对比度。在图3中,纵轴表示第2语音信号Sig2的频带,横轴表示时间(单位为“秒”)。另外,在图3中,明暗表示SNR的高低,越亮则表示SNR越高,越暗则表示SNR越低。

[0074] 在第1判定动作中,清晰度判定部122在语音区间(图3中的由矩形框包围的区间。例如,零点几秒的区间)中,将第2语音信号Sig2中的语音频带的SNR与阈值进行比较。并且,在第1判定动作中,若SNR高于阈值,则清晰度判定部122判定为第2语音信号Sig2清晰,若SNR低于阈值,则清晰度判定部122判定为第2语音信号Sig2不清晰。

[0075] 在此,语音信号中的语音频带的SNR例如能够作为语音信号中的语音频带所包含的各频带的SNR的代表值来计算。代表值例如是平均值、中央值、最大值或最频值等。另外,语音信号中的语音频带的SNR例如也能够作为语音频带所包含的各频带的SNR的代表值与语音频带以外的各频带的SNR的代表值之比来计算。在后者中,例如即使在由于换气扇的动作音大等而用户的周围比较嘈杂、在全部频带中SNR比较高的情况下,清晰度判定部122也容易判定第2语音信号Sig2是否清晰。

[0076] 图3的(a)表示在由矩形框包围的语音区间中,第2语音信号Sig2中的语音频带(图3的(a)中箭头所示的频带)的SNR比阈值低。因此,在图3的(a)所示的例子中,清晰度判定部122在第1判定动作中判定为第2语音信号Sig2不清晰。

[0077] 另一方面,图3的(b)表示在由矩形框包围的语音区间中,第2语音信号Sig2中的语音频带(图3的(b)中箭头所示的频带)的SNR比阈值高。因此,在图3的(b)所示的例子中,清晰度判定部122在第1判定动作中判定为第2语音信号Sig2清晰。

[0078] 图4是用于判定第2语音信号Sig2的清晰度的第2判定动作的说明图。图4表示上述语音区间中的第2语音信号Sig2的频谱。在图4中,纵轴表示第2语音信号Sig2的振幅值,横轴表示第2语音信号Sig2的频率。另外,在图4中,实线L1表示频谱包络,单点划线表示频谱包络的倾向。

[0079] 在第2判定动作中,清晰度判定部122在语音区间中计算第1频带B1、第2频带B2以及第3频带B3各自中的频谱包络的峰度,并将计算出的峰度与阈值进行比较。并且,在第2判定动作中,若在任一频带B1、B2、B3中峰度均高于阈值,则清晰度判定部122判定为第2语音信号Sig2清晰,若在至少一个频带中峰度低于阈值,则清晰度判定部122判定为第2语音信号Sig2不清晰。

[0080] 第1频带B1是与人的语音中的元音的第1共振峰对应的频带。第2频带B2是与人的语音中的元音的第2共振峰对应的频带。第3频带B3是与人的语音中的元音的第2共振峰以后的共振峰对应的频带。

[0081] 在此,各频带B1~B3是与日语中的元音的共振峰对应的频带。因而,例如在针对英语等日语以外的语言判定第2语音信号Sig2是否清晰的情况下,清晰度判定部122只要在与该语言的元音的共振峰对应的一个以上的频带中分别计算频谱包络的峰度,并将计算出的峰度与阈值进行比较即可。

[0082] 峰度是表示概率变量的概率密度函数或频度分布的锐度的指标。峰度越高,则成为与正态分布相比具有越尖锐的峰值和越长且越粗的下摆的分布(换言之,频谱包络的峰

值周围的变化越陡峭),峰度越低,则成为与正态分布相比具有越带有圆角的峰值和越短且越细的下摆的分布(换言之,频谱包络的变化越平缓)。

[0083] 在第2判定动作中,清晰度判定部122在如上述那样在各频带B1~B3中峰度比阈值高的情况下,判定为人的语音中的元音的特征显著显现,即人的语音清晰到能够听取元音的程度。

[0084] 图4的(a)及图4的(b)均表示在语音区间中人发出元音“o”的情况。图4的(a)如实线L1和单点划线所示,表示在各频带B1~B3中频谱包络平缓,即在各频带B1~B3中峰度比阈值低。因此,在图4的(a)所示的例子中,清晰度判定部122在第2判定动作中判定为第2语音信号Sig2不清晰。

[0085] 另一方面,图4的(b)如实线和单点划线所示,表示在各频带B1~B3中出现频谱包络的峰值,并且峰值周围的变化陡峭,即在各频带B1~B3中峰度比阈值高。因此,在图4的(b)所示的例子中,清晰度判定部122在第2判定动作中判定为第2语音信号Sig2清晰。

[0086] 这样,处理器12(清晰度判定部122)至少基于第2语音信号Sig2中的与元音对应的分量来判定是否满足第2条件。

[0087] 此外,在实施方式中,清晰度判定部122执行第1判定动作以及第2判定动作双方,但不限于此,例如也可以仅通过第2判定动作来判定是否满足第2条件。但是,若考虑空间中的混响等背景噪声的影响,则清晰度判定部122执行第1判定动作以及第2判定动作双方时能够更高精度地判定第2语音信号Sig2的清晰度。

[0088] [3.动作]

[0089] 以下,使用图5说明实施方式所涉及的语音通话装置4(语音处理系统1)的动作、即语音处理方法的一例。图5是表示实施方式所涉及的语音处理系统1的动作例的流程图。

[0090] 首先,若第2输入I/F11取得第2语音信号Sig2(S101:是),则处理器12将所取得的第2语音信号Sig2保持于缓冲器。以下,只要无特别说明,则“第2语音信号Sig2”相当于保持在缓冲器中的第2语音信号Sig2。然后,若第1输入I/F10取得第1语音信号Sig1(S102:是),则处理器12进行延迟时间的计算和更新(S103)。具体而言,处理器12通过计算第1输入I/F10取得第1语音信号Sig1的时间点与第2输入I/F11取得第2语音信号Sig2的时间点的差分,来计算延迟时间,并将之前的延迟时间更新为计算出的延迟时间。此外,如果计算出的延迟时间与之前的延迟时间相同,则处理器12不进行更新。

[0091] 接着,处理器12基于延迟时间,以使第1语音信号Sig1的开始时间点与第2语音信号Sig2的开始时间点一致的方式,进行第1语音信号Sig1与第2语音信号Sig2的时间偏差的修正(S104)。

[0092] 接着,处理器12基于第2语音信号Sig2,计算为了判定第2语音信号Sig2的清晰度而需要的参数(S105)。以下,使用图6对步骤S105进行详细说明。

[0093] 图6是表示为了判定第2语音信号Sig2的清晰度而需要的参数的计算例的流程图。首先,处理器12检测第2语音信号Sig2的语音区间(S201)。例如,处理器12将从第2语音信号Sig2的开始时间点起经过了规定时间的时间点作为起点来检测语音区间。语音区间例如是零点几秒的区间。

[0094] 接着,处理器12计算检测出的语音区间中的频谱对比度(S202)。并且,处理器12基于计算出的频谱对比度,计算第2语音信号Sig2中的语音频带的SNR(S203)。

[0095] 另外,处理器12与步骤S202、S203并行地或者在步骤S202、203的前后的任一方,计算所检测出的语音区间中的第2语音信号Sig2的特征量(S204)。在此,处理器12计算第2语音信号Sig2的基本频率和第2语音信号Sig2的频谱包络作为第2语音信号Sig2的特征量。接着,处理器12将计算出的第2语音信号Sig2的特征量保存在存储器13中(S205)。

[0096] 接着,处理器12计算检测出的语音区间中的第2语音信号Sig2的频谱包络的峰度(S206)。具体而言,处理器12在检测出的语音区间中,计算第1频带B1、第2频带B2和第3频带B3各自中的频谱包络的峰度。

[0097] 返回图5,处理器12判定第2语音信号Sig2的清晰度(S106)。具体而言,处理器12在检测出的语音区间中执行将第2语音信号Sig2的语音频带的SNR与阈值进行比较的第1判定动作。另外,处理器12在检测出的语音区间中执行将各频带B1~B3的频谱包络的峰度与阈值进行比较的第2判定动作。并且,处理器12在第1判定动作以及第2判定动作中的哪个判定动作中都判定为清晰的情况下,判定为第2语音信号Sig2清晰、即满足第2条件。另一方面,处理器12在第1判定动作以及第2判定动作中的至少一方的判定动作中判定为不清晰的情况下,判定为第2语音信号Sig2不清晰、即不满足第2条件。

[0098] 在判定为第2语音信号Sig2清晰、即满足第2条件的情况下(S106:是),处理器12接着基于第1语音信号Sig1以及第2语音信号Sig2,计算为了判定发言同一性而需要的参数(S107)。以下,使用图7对步骤S107进行详细说明。

[0099] 图7是表示为了判定发言同一性而需要的参数的计算例的流程图。首先,处理器12检测第1语音信号Sig1的语音区间(S301)。例如,处理器12将从第1语音信号Sig1的开始时间点起经过了规定时间的时间点作为起点,检测语音区间。检测出的语音区间是与第2语音信号Sig2的语音区间相同的区间。

[0100] 接着,处理器12读入保存在存储器13中的第2语音信号Sig2的特征量(S302)。另外,处理器12与步骤S302并行地或者在步骤S302的前后的任一方,计算所检测出的语音区间中的第1语音信号Sig1的特征量(S303)。在此,处理器12计算第1语音信号Sig1的基本频率和第1语音信号Sig1的频谱包络,作为第1语音信号Sig1的特征量。

[0101] 返回图5,处理器12判定发言同一性(S108)。具体而言,处理器12在(i)第1语音信号Sig1的基本频率与第2语音信号Sig2的基本频率相同、且(ii)第1语音信号Sig1中元音出现的定时与第2语音信号Sig2中元音出现的定时相同的情况下,判定为说话者相同,即满足第1条件。另一方面,处理器12在不满足上述的(i)及(ii)中的至少一方的情况下,判定为说话者不相同、即不满足第1条件。另外,在此,如果两个比较对象的差分为阈值以下,则处理器12判定为两个比较对象相同。

[0102] 在判定为说话者相同、即满足第1条件的情况下(S108:是),满足第1条件以及第2条件两者,因此处理器12判定为处于第1状况,降低从扬声器3输出的基于第1语音信号Sig1的语音(即,通信语音)的音量(S109)。另外,处理器12关闭噪声消除功能(S110),开启外部声音取入功能(S111)。此外,执行步骤S109~S111的顺序不限于该顺序。

[0103] 另一方面,处理器12在不满足第1条件以及第2条件中的至少一方的情况下,即在判定为第2语音信号Sig2不清晰的情况下(S106:否),或者在判定为说话者不相同的情况下(S108:否),判定为处于第2状况。并且,处理器12将通信语音的音量设定为默认的音量(S112)。另外,处理器12开启噪声消除功能(S113),关闭外部声音取入功能(S114)。此外,执

行步骤S112~S114的顺序不限于该顺序。

[0104] 另外,步骤S112~S114在第2输入I/F11未取得第2语音信号Sig2的情况下(S101:否)、或者第1输入I/F10未取得第1语音信号Sig1的情况下(S102:否)也执行。

[0105] 并且,处理器12在通话结束之前的期间(S115:否),重复上述的一系列处理。另一方面,若通话结束(S115:是),则处理器12结束动作。

[0106] 图8是实施方式所涉及的语音处理系统1的动作例的概要的说明图。图8表示在同一据点存在两个用户U1、U2的情况下的、用户U1佩戴的语音通话装置4(语音处理系统1)的一系列动作。

[0107] 如图8的(a)所示,若其他用户U2发出语音V2,则通过麦克风2将该语音转换为第2语音信号Sig2,由此第2输入I/F11取得第2语音信号Sig2。并且,处理器12检测语音区间,在检测出的语音区间中,计算作为第2语音信号Sig2的特征量的第2语音信号Sig2的基本频率以及MFCC(频谱包络)。另外,处理器12将计算出的第2语音信号Sig2的基本频率以及MFCC保存于存储器13。另外,其他用户U2发出的语音V2作为第1语音信号Sig1被发送到会议系统100。

[0108] 接着,如图8的(b)所示,处理器12计算所检测出的语音区间中的第2语音信号Sig2的语音频带的SNR和频谱包络的峰度。并且,处理器12使用计算出的第2语音信号Sig2的语音频带的SNR和频谱包络的峰度,判定第2语音信号Sig2是否清晰、即是否满足第2条件。

[0109] 另外,如图8的(c)所示,若第1输入I/F10取得从会议系统100发送的第1语音信号Sig1,则处理器12检测第1语音信号Sig1的语音区间,在检测出的语音区间中,计算作为第1语音信号Sig1的特征量的第1语音信号Sig1的基本频率以及MFCC(频谱包络)。并且,处理器12从存储器13读入第2语音信号Sig2的基本频率以及MFCC,进行与第1语音信号Sig1的基本频率以及MFCC的对照,由此判定说话者是否相同、即是否满足第1条件。

[0110] 在满足第1条件和第2条件两者的情况下,即在判定为处于第1状况的情况下,如图8的(d)所示,处理器12降低从扬声器3输出的通信语音(基于第1语音信号Sig1的语音)的音量,或者不从扬声器3再现通信语音。另外,处理器12关闭噪声消除功能,并且开启外部声音取入功能。由此,用户U1关于其他用户U2发出的语音V2,能够几乎听不到经由会议系统100的语音而主要听到来自其他用户U2的直接语音且清晰的语音。

[0111] 另一方面,在不满足第1条件以及第2条件中的至少一方的情况下,即在判定为处于第2状况的情况下,如图8的(e)所示,处理器12使通信语音(基于第1语音信号Sig1的语音)从扬声器3输出。另外,处理器12开启噪声消除功能,并且关闭外部声音取入功能。由此,用户U1关于其他用户U2发出的语音V2,能够几乎听不到来自其他用户U2的直接语音且不清晰的语音而主要听到经由会议系统100的语音。

[0112] [4.效果等]

[0113] 以下,使用图9说明实施方式所涉及的语音处理系统1的优点。图9是实施方式所涉及的语音处理系统1的优点的说明图。图9表示处于第1据点A1的两个用户U1、U2、处于第2据点A2的用户U3以及处于第3据点A3的用户U4使用会议系统100在线进行会议。

[0114] 图9的(a)表示处于第1据点A1的两个用户U1、U2存在于相互比较近的位置,用户U1容易直接听到其他用户U2发出的“你好”这样的语音V2的状况。在这样的状况的情况下,用户U1佩戴的语音通话装置4(语音处理系统1)判定为满足第1条件以及第2条件两者、即处于

第1状况,将降低了与第1语音信号Sig1对应的分量的信号包含在输出语音信号Sig3中。在此,语音处理系统1不使第1语音信号Sig1包含在输出语音信号Sig3中,即不从扬声器3再现基于第1语音信号Sig1的语音。

[0115] 因此,用户U1能够直接听到其他用户U2发出的“你好”这样的清晰的语音V2,而不会听到经由会议系统100发送的其他用户U2发出的语音V1。即,关于其他用户U2发出的语音,用户U1不会听到来自其他用户U2的直接语音和经由会议系统100(通信线路)的语音双方,因此容易听到其他用户U2发出的语音。因此,在语音处理系统1中,具有即使在相同的据点存在多个用户的环境中也不易损害会话的舒适性的优点。

[0116] 另外,在实施方式中,语音处理系统1在判定为处于第1状况的情况下,开启外部声音取入功能,即将第2语音信号Sig2包含在输出语音信号Sig3中,因此具有如下优点:通过取入用户U1的周围的语音,更容易听到来自其他用户U2的直接语音。

[0117] 图9的(b)表示处于第1据点A1的两个用户U1、U2存在于相互比较远离的位置,用户U1难以直接听到其他用户U2发出的“你好”这样的语音V2的状况。在这样的状况的情况下,用户U1佩戴的语音通话装置4(语音处理系统1)判定为至少不满足第2条件、即处于第2状况,将第1语音信号Sig1包含在输出语音信号Sig3中,并且不将第2语音信号Sig2包含在输出语音信号Sig3中。

[0118] 因此,用户U1能够听到经由会议系统100发送的其他用户U2发出的“你好”这样的语音V1,而几乎不会听到来自其他用户U2的直接语音且不清晰的语音V2。即,关于其他用户U2发出的语音,用户U1不会听到来自其他用户U2的直接语音和经由会议系统100(通信线路)的语音双方,因此容易听到其他用户U2发出的语音。因此,在语音处理系统1中,具有即使在相同的据点存在多个用户的环境中也不易损害会话的舒适性的优点。

[0119] 另外,在实施方式中,语音处理系统1在判定为处于第2状况的情况下,开启噪声消除功能,即将与第2语音信号Sig2相反相位的语音信号包含在输出语音信号Sig3中。因此,通过去除包括来自其他用户U2的直接语音的用户U1周围的噪声,具有用户U1更容易听到来自其他用户U2的经由会议系统100(通信线路)的语音的优点。

[0120] 另外,在多个用户(在此为用户U1~U4)同时发出语音的情况下,语音处理系统1判定为第1条件和第2条件均不满足、即处于第2状况。在该情况下,从扬声器3输出经由会议系统100的来自其他用户U2~U4的语音。在这样的情况下,各用户U1~U4的会话暂时停止,因此不会妨碍语音处理系统1的优点。采用语音处理系统1的用户U1只要至少在各用户U1~U4交替发出语音的状况下能够享受上述优点即可。

[0121] [5.其他实施方式]

[0122] 以上,对实施方式进行了说明,但本公开并不限定于上述实施方式。

[0123] 在上述实施方式中,处理器12也可以不具备外部声音取入切换部128及ANC控制部129。在该情况下,语音处理系统1也可以不执行图5所示的流程图中的步骤S110、S111、S113、S114。更具体而言,在语音通话装置4为开放型的语音通话装置的情况下,处理器12可以具备外部声音取得切换部128及ANC控制部129,但也可以不具备。此外,在语音通话装置4为开放型的语音通话装置的情况下,处理器12也可以具备ANC控制部129。在该情况下,语音处理系统1也可以不执行图5所示的流程图中的步骤S111、S114。另外,在语音通话装置4是密闭型的语音通话装置的情况下,处理器12最好具备外部声音取入切换部128,但例如如果

是多少可听见外部声音的情况,则也可以不具备外部声音取入切换部128。另外,在语音通话装置4为密闭型的语音通话装置的情况下,处理器12最好具备ANC控制部129,但如果是因用户的耳孔被堵塞而使外部声音降低一定程度的情况,则也可以不具备ANC控制部129。

[0124] 另外,在上述实施方式中,例如在一边在室内移动一边实施的VR会议、或者室内的环境噪声变化的状态等清晰度的特征量容易变化的状况下,输出语音控制部127、外部声音取入切换部128以及ANC控制部129各自始终被控制,但它们也可以在一定时间内不被控制。更具体而言,语音处理系统1也可以将图5所示的流程图中的步骤S112~S114以及S109~S111不执行一定时间(例如,几毫秒)。在该情况下,输出语音控制部127、外部声音取入切换部128以及ANC控制部129各自可以每隔一定时间被控制,防止被高频率地控制。此外,对输出语音控制部127进行控制的时间、对外部声音取得切换部128进行控制的时间以及对ANC控制部129进行控制的时间也可以分别不同。

[0125] 另外,在上述实施方式中,语音处理系统1通过单一的装置(语音通话装置4)来实现,但也可以作为多个装置来实现。在语音处理系统1由多个装置实现的情况下,语音处理系统1所具备的功能性的构成要素可以任意地分配给多个装置。另外,例如,语音处理系统1也可以通过具备第1输入I/F10、第2输入I/F11以及处理器12的服务器来实现。在该情况下,语音处理系统1通过与具备麦克风2以及扬声器3的设备之间进行通信,能够从麦克风2取得第2语音信号Sig2,或者从扬声器3输出基于输出语音信号Sig3的语音。

[0126] 另外,上述实施方式中的装置间的通信方法没有特别限定。在上述实施方式中,在两个装置进行通信的情况下,也可以在两个装置之间存在未图示的中继装置。

[0127] 另外,在上述实施方式中说明的处理的顺序是一例。多个处理的顺序也可以变更,也可以并行执行多个处理。另外,特定的处理部所执行的处理也可以由其他处理部执行。另外,在上述实施方式中说明的数字信号处理的一部分也可以通过模拟信号处理来实现。

[0128] 另外,在上述实施方式中,各构成要素也可以通过执行适合于各构成要素的软件程序来实现。各构成要素也可以通过CPU或处理器等程序执行部读出并执行记录于硬盘或半导体存储器等记录介质的软件程序来实现。

[0129] 另外,各构成要素也可以通过硬件来实现。例如,各构成要素也可以是电路(或集成电路)。这些电路既可以作为整体构成一个电路,也可以分别是不同的电路。另外,这些电路分别可以是通用的电路,也可以是专用的电路。

[0130] 另外,本公开的整体或具体的技术方案也可以通过系统、装置、方法、集成电路、计算机程序或计算机可读的CD-ROM等记录介质来实现。另外,也可以通过系统、装置、方法、集成电路、计算机程序以及记录介质的任意组合来实现。例如,本公开既可以作为计算机执行的语音处理方法来执行,也可以作为用于使计算机执行这样的语音处理方法的程序来实现。另外,本公开也可以作为记录有这样的程序的计算机可读的非暂时性的记录介质来实现。此外,这里的程序包括用于使通用的信息终端作为上述实施方式的语音处理系统发挥功能的应用程序。

[0131] 此外,对各实施方式实施本领域技术人员想到的各种变形而得到的方式、或者通过在不脱离本公开的主旨的范围内将各实施方式中的构成要素以及功能任意地组合而实现的方式也包含于本公开。

[0132] (总结)

[0133] 如上所述,第1方式所涉及的语音处理系统1具备第1输入I/F10、第2输入I/F11以及处理器12。处理器12是信号处理电路的一例。第1输入I/F10取得经由通信线路的第1语音信号Sig1。第2输入I/F11取得基于由麦克风2收集到的语音的第2语音信号Sig2。处理器12将基于第1语音信号Sig1及第2语音信号Sig2的输出语音信号Sig3输出到扬声器3。处理器12在满足第1条件以及第2条件两者的情况下,将降低了与第1语音信号Sig1对应的分量的信号包含在输出语音信号Sig3中,第1条件为第1语音信号Sig1以及第2语音信号Sig2均包含基于同一人物发出的语音的语音信号,第2条件为第2语音信号Sig2清晰。

[0134] 由此,在相同的据点存在多个用户的环境中,用户在来自其他用户的直接语音和经由通信线路的语音中的直接语音清晰的情况下,主要听到直接语音,因此容易听取其他用户发出的语音。即,具有即使在相同的据点存在多个用户的环境中也不易损害会话的舒适性的优点。

[0135] 另外,在第2方式所涉及的语音处理系统1中,在第1方式中,处理器12在不满足第1条件及第2条件中的至少一方的情况下,将第1语音信号Sig1包含在输出语音信号Sig3中,且不将第2语音信号Sig2包含在输出语音信号Sig3中。

[0136] 由此,在相同的据点存在多个用户的环境中,用户主要听到来自其他用户的直接语音和经由通信线路的语音中的、比直接语音容易听取的经由通信线路的语音,因此容易听取其他用户发出的语音。即,具有即使在相同的据点存在多个用户的环境中也不易损害会话的舒适性的优点。

[0137] 另外,在第3方式所涉及的语音处理系统1中,在第1或第2方式中,处理器12根据第1语音信号Sig1中的与元音对应的分量和第2语音信号Sig2中的与元音对应的分量之间的相关性,判定是否满足第1条件。

[0138] 由此,与基于第1语音信号Sig1的波形和第2语音信号Sig2的波形的相似度的方法相比,具有容易判定是否包含基于同一人物发出的语音的语音信号的优点。

[0139] 另外,在第4方式所涉及的语音处理系统1中,在第1~第3方式中的任一方式中,处理器12基于第2语音信号Sig2中的与元音对应的分量来判定是否满足第2条件。

[0140] 由此,具有如下优点:基于作为人的语音的听取容易度的指标的与元音对应的分量来进行判定,因此容易判定第2语音信号Sig2是否清晰的优点。

[0141] 另外,在第5方式所涉及的语音处理系统1中,在第1~第4方式中的任一方式中,处理器12在满足第1条件以及第2条件两者的情况下,将第2语音信号Sig2包含在输出语音信号Sig3中。

[0142] 由此,具有如下优点:通过取入用户周围的语音,用户更容易听到来自其他用户的直接语音。

[0143] 另外,在第6方式所涉及的语音处理系统1中,在第2方式中,处理器12在不满足第1条件以及第2条件中的至少一方的情况下,进一步将与第2语音信号Sig2相反相位的语音信号包含在输出语音信号Sig3中。

[0144] 由此,具有如下优点:通过去除包含来自其他用户的直接语音的用户周围的噪声,用户更容易听取来自其他用户的经由通信线路的语音。

[0145] 另外,在第7方式的语音处理方法中,取得经由通信线路的第1语音信号Sig1(S102:是),取得基于由麦克风2收集到的语音的第2语音信号Sig2(S101:是),在满足第1条

件以及第2条件两者的情况下(S106:是、S108:是),将减少了与第1语音信号Sig1对应的分量的信号包含在输出语音信号Sig3中并输出到扬声器3(S109),第1条件为第1语音信号Sig1以及第2语音信号Sig2均包含基于同一人物发出的语音的语音信号,第2条件为第2语音信号Sig2清晰。

[0146] 由此,在相同的据点存在多个用户的环境中,用户在来自其他用户的直接语音和经由通信线路的语音中的直接语音清晰的情况下,主要听到直接语音,因此容易听取其他用户发出的语音。即,具有即使在相同的据点存在多个用户的环境中也不易损害会话的舒适性的优点。

[0147] 并且,第8方式所涉及的程序使1个以上的处理器执行第7方式所涉及的语音处理方法。

[0148] 由此,在相同的据点存在多个用户的环境中,用户在来自其他用户的直接语音和经由通信线路的语音中的直接语音清晰的情况下,主要听到直接语音,因此容易听到其他用户发出的语音。即,具有即使在相同的据点存在多个用户的环境中也不易损害会话的舒适性的优点。

[0149] 工业实用性

[0150] 本公开的语音处理系统等能够应用于对扬声器发出的语音进行处理的系统等。

[0151] 附图标记说明

[0152] 1 语音处理系统

[0153] 10 第1输入I/F

[0154] 11 第2输入I/F

[0155] 12 处理器

[0156] 121 清晰度计算部

[0157] 122 清晰度判定部

[0158] 123 第1特征量计算部

[0159] 124 第2特征量计算部

[0160] 125 发言同一性判定部

[0161] 126 输出语音判定部

[0162] 127 输出语音控制部

[0163] 128 外部声音取入切换部

[0164] 129 ANC控制部

[0165] 13 存储器

[0166] 2 麦克风

[0167] 3 扬声器

[0168] 100 会议系统

[0169] Sig1 第1语音信号

[0170] Sig2 第2语音信号

[0171] Sig3 输出语音信号

[0172] V1、V2 语音

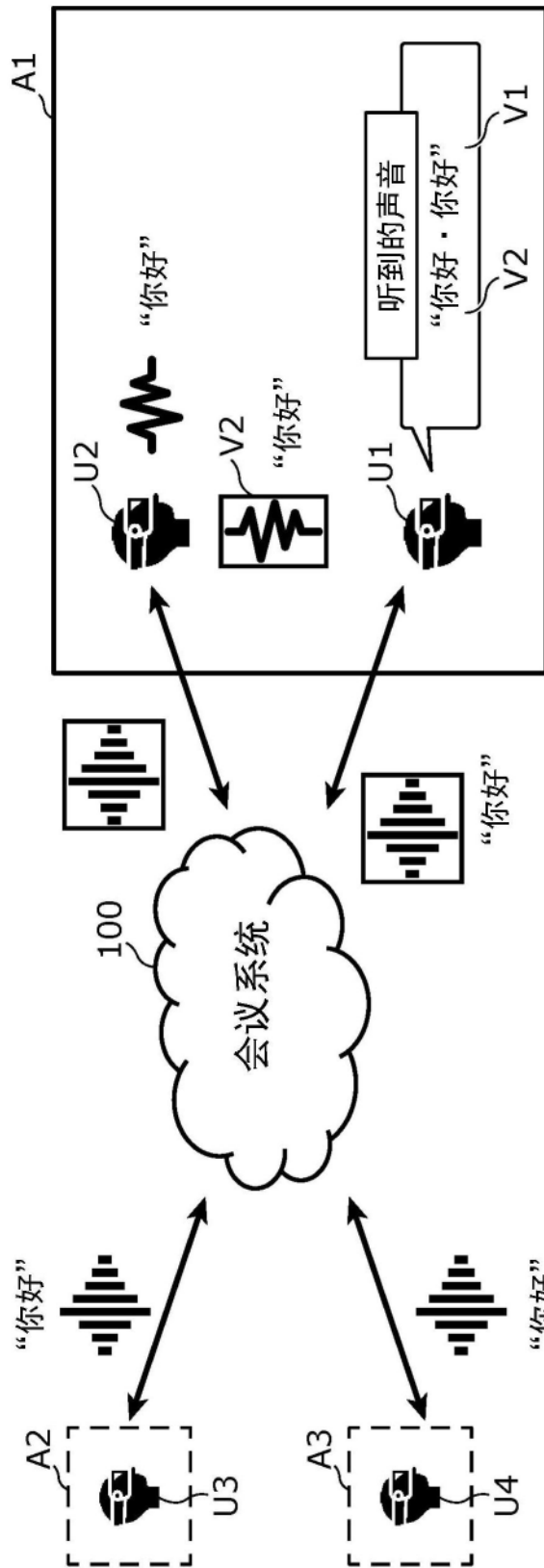


图1

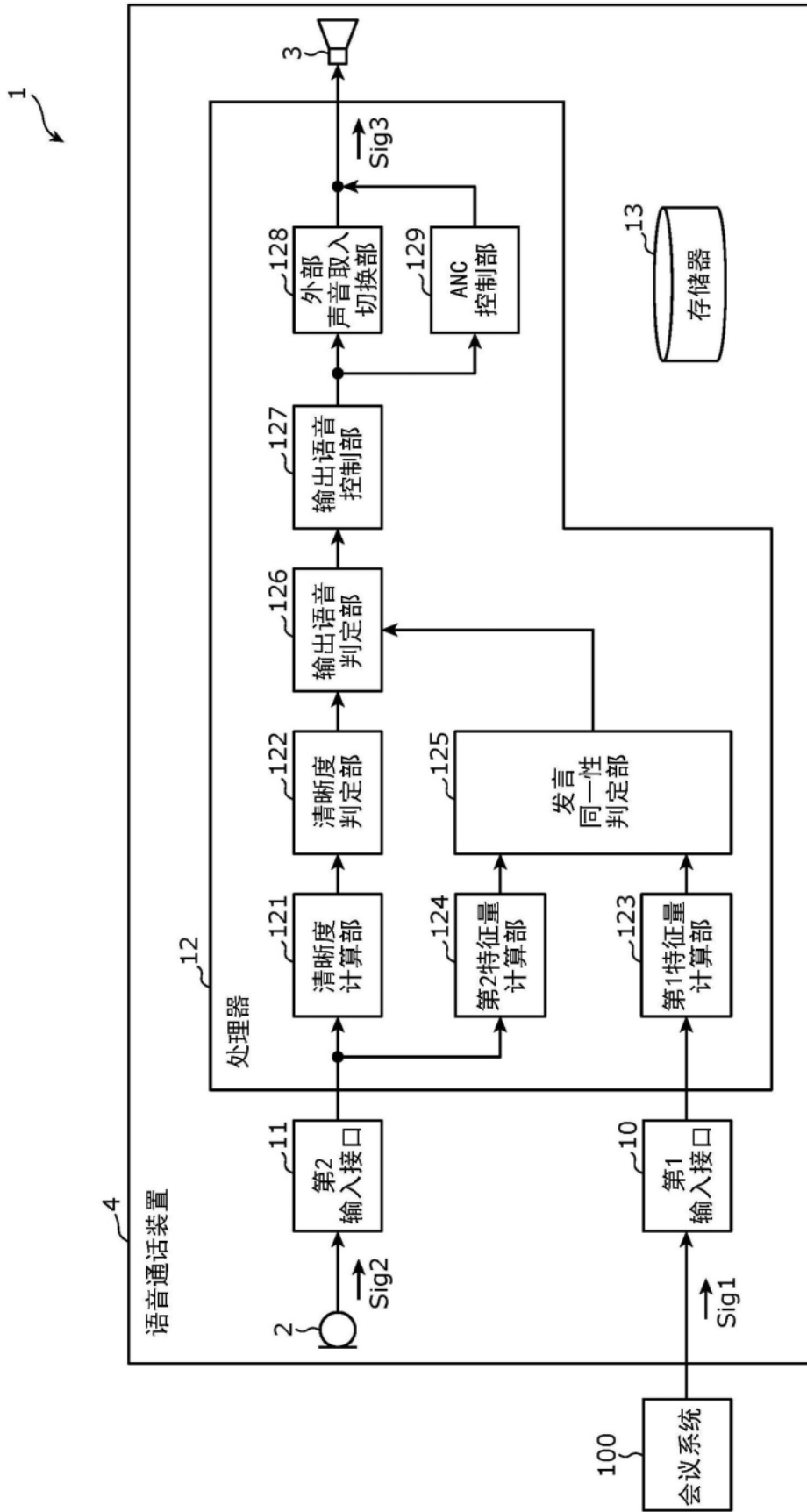


图2

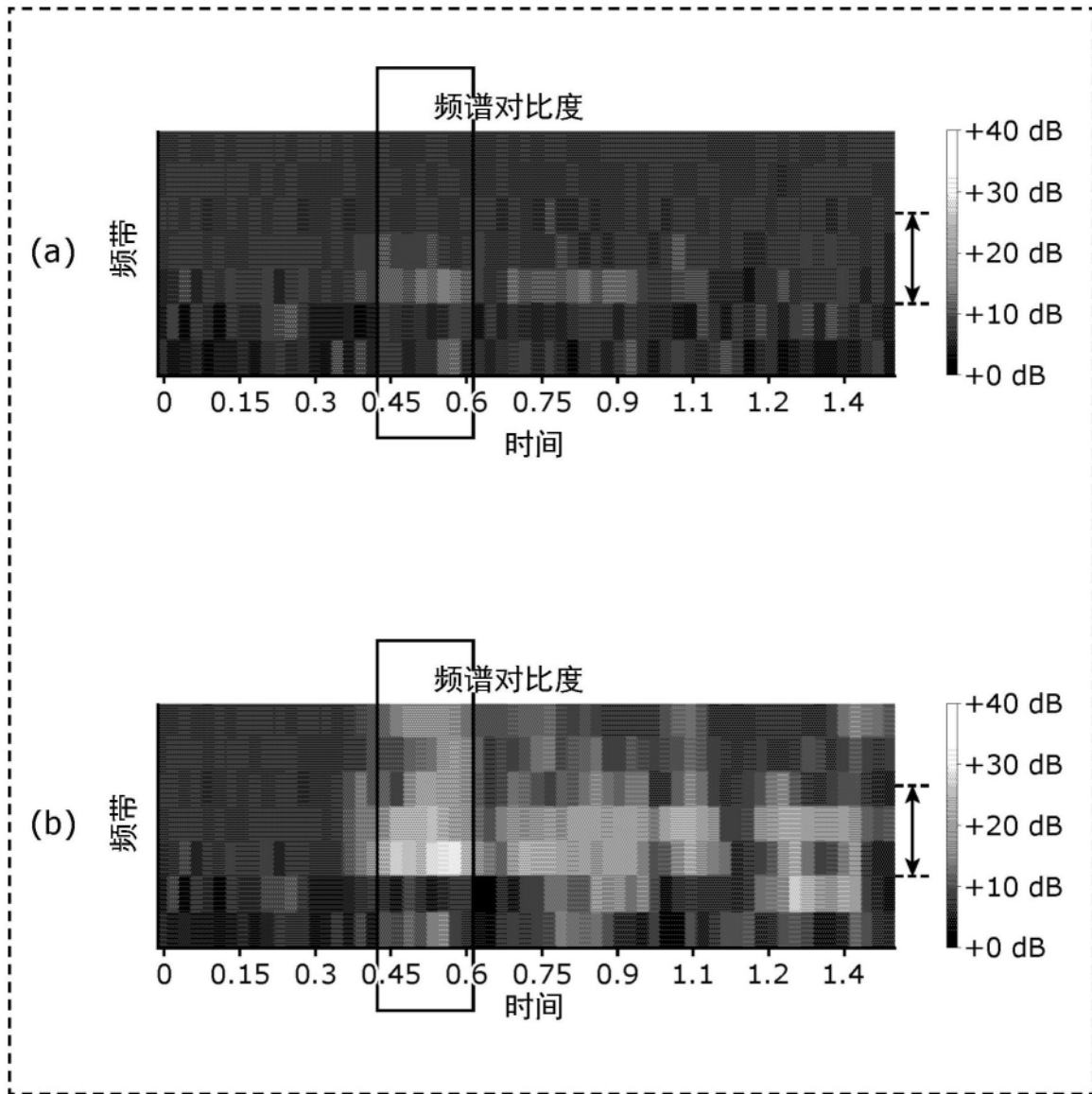


图3

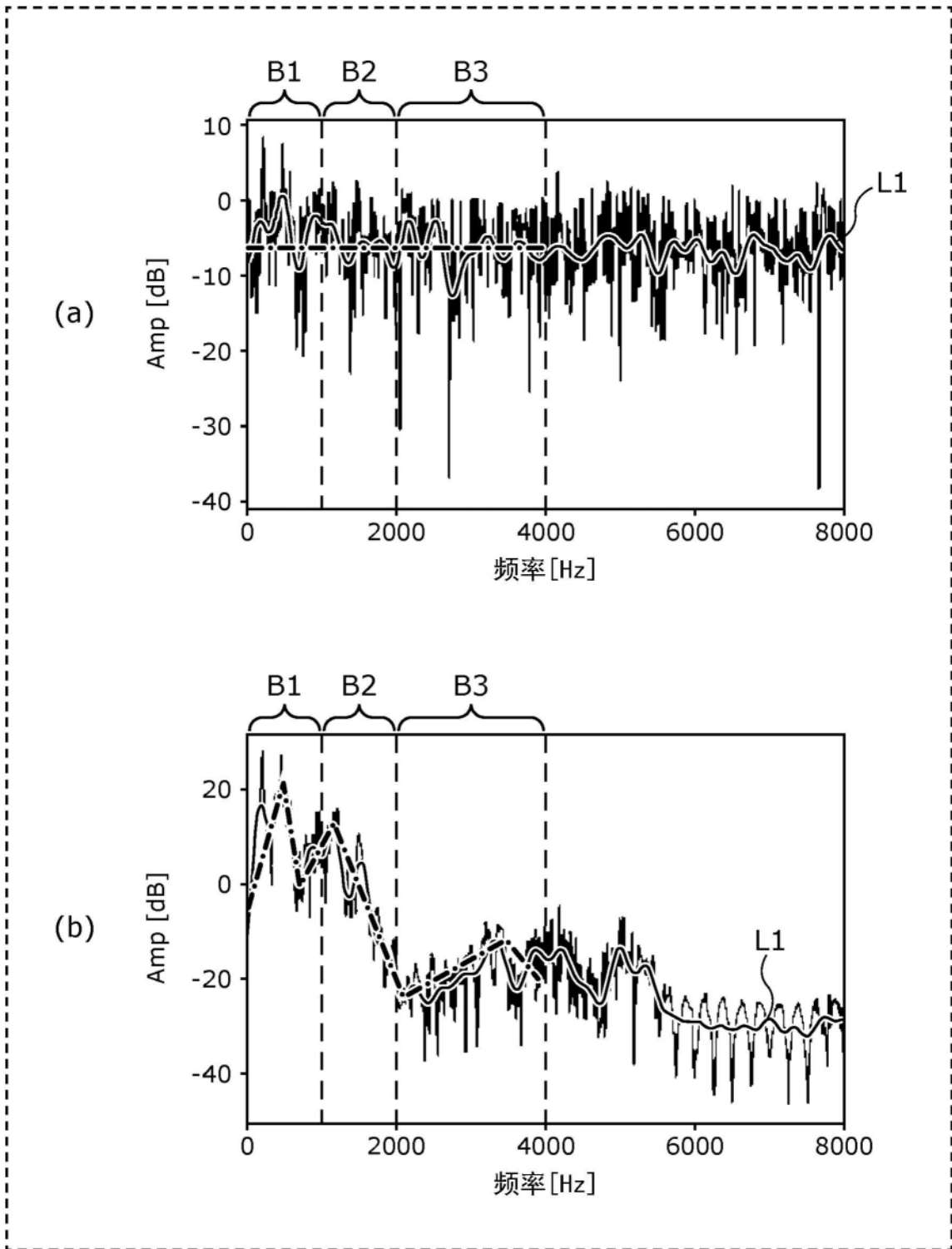


图4

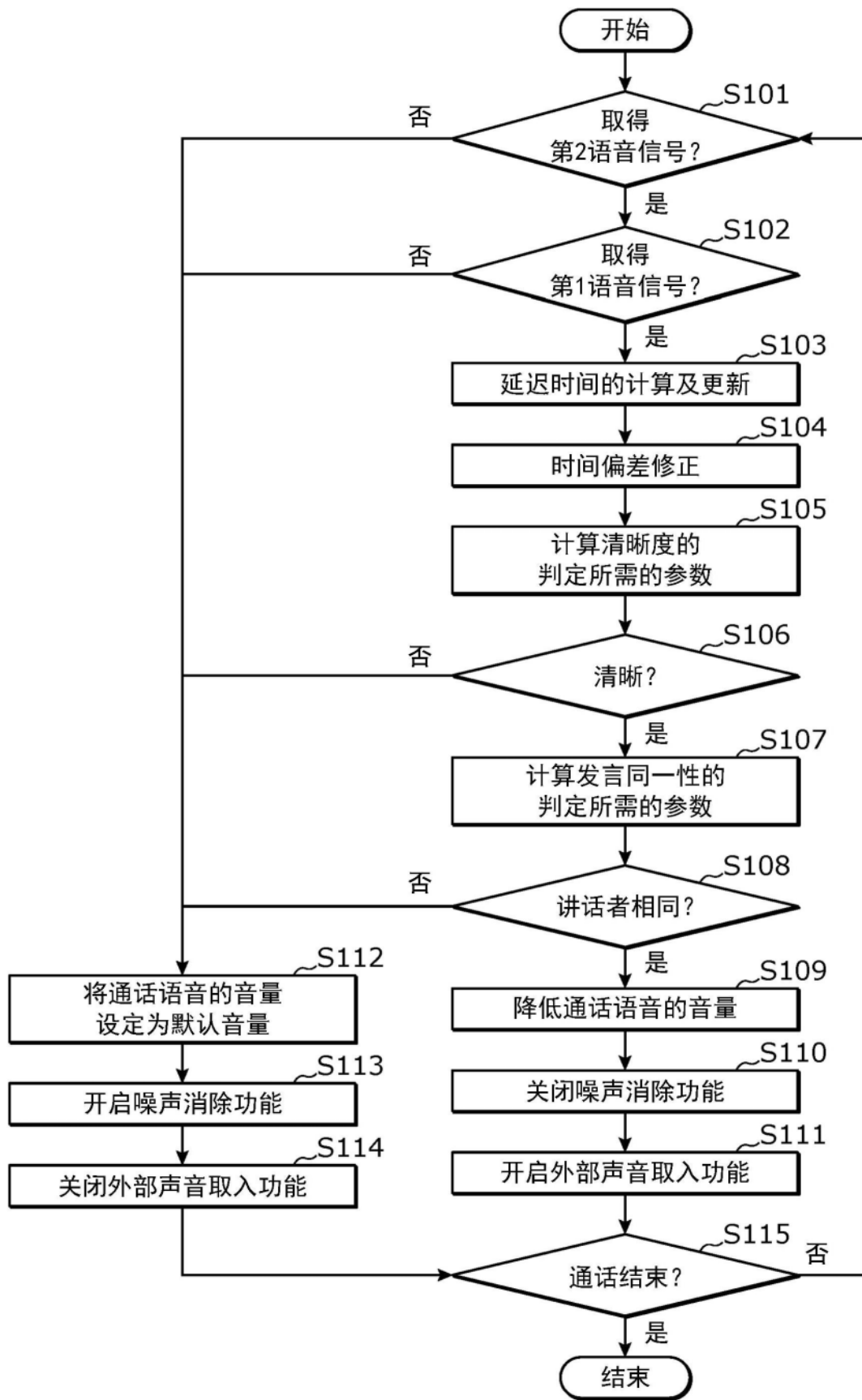


图5

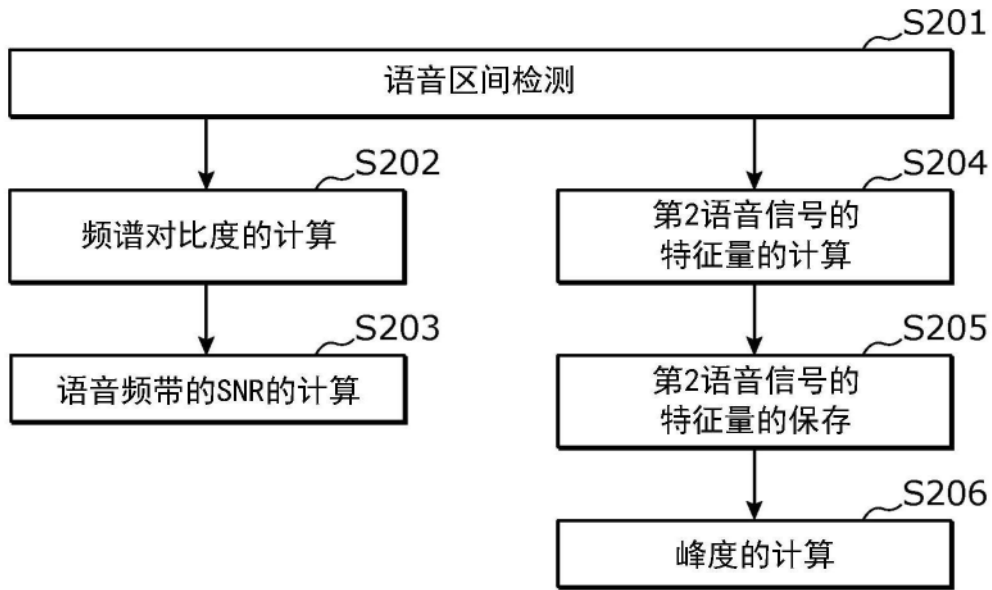


图6



图7

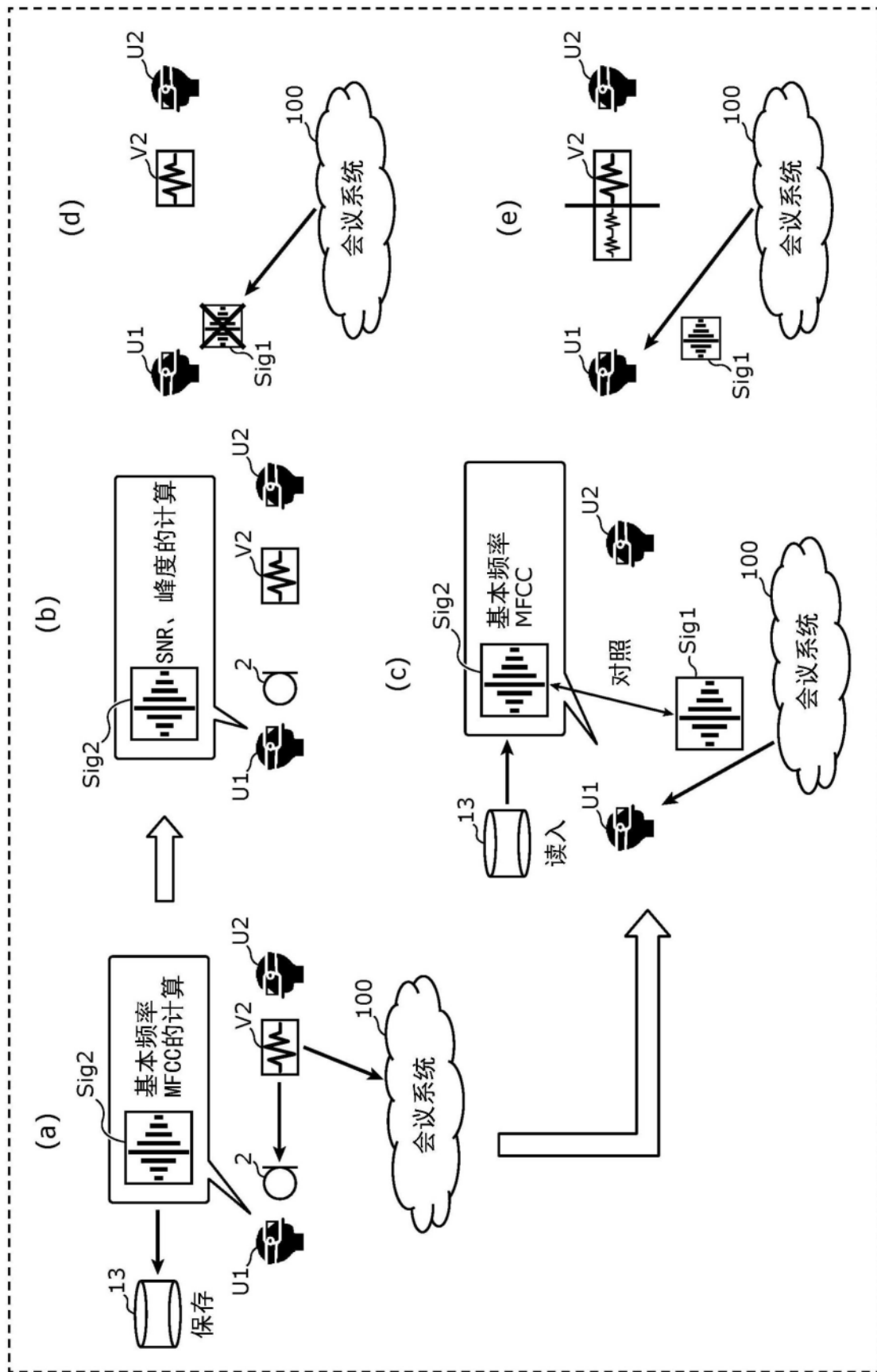


图8

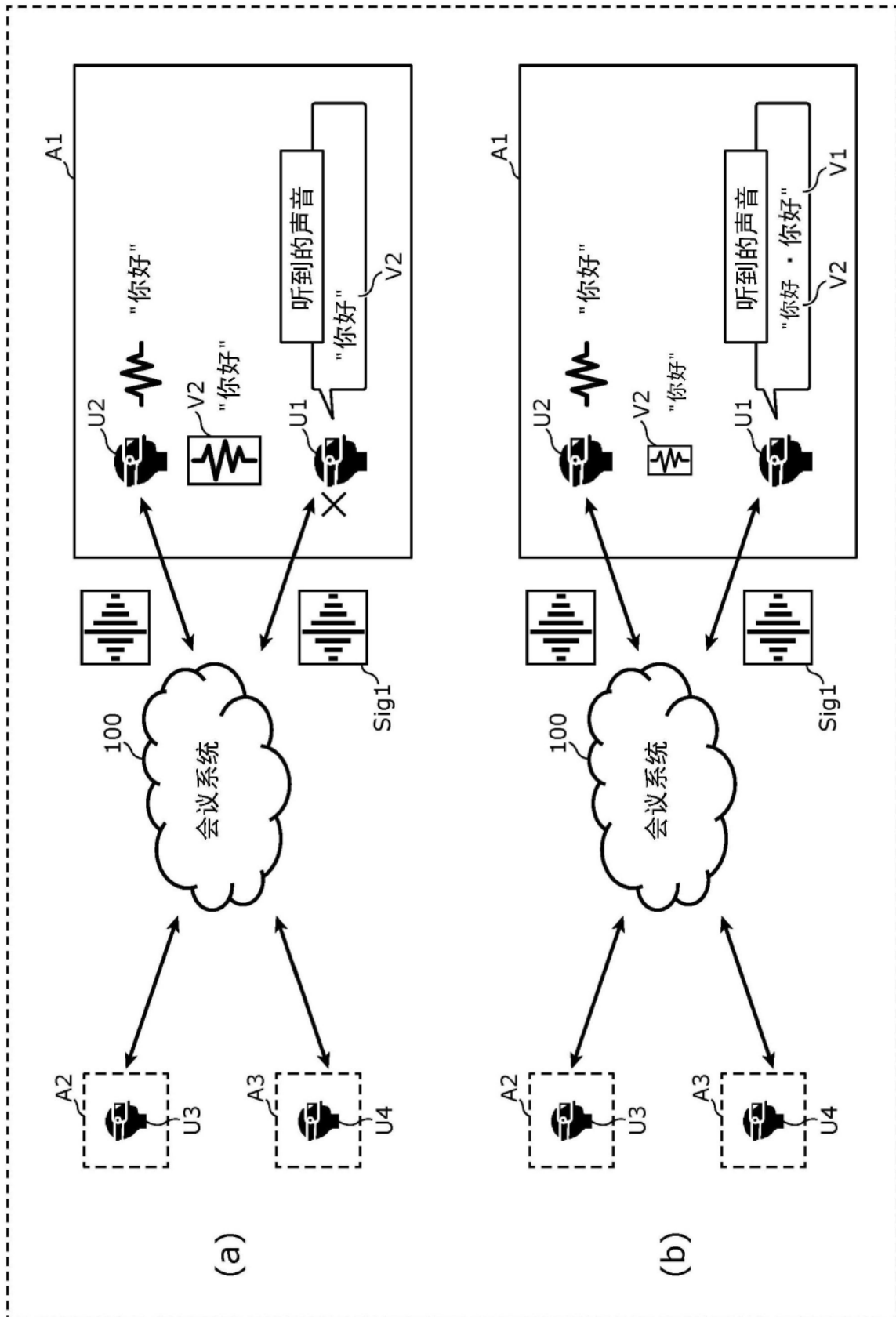


图9