



(12) 发明专利申请

(10) 申请公布号 CN 104981870 A

(43) 申请公布日 2015. 10. 14

(21) 申请号 201480008333. 7

(51) Int. Cl.

(22) 申请日 2014. 01. 15

G10L 21/0364(2006. 01)

(30) 优先权数据

2013-033347 2013. 02. 22 JP

(85) PCT国际申请进入国家阶段日

2015. 08. 11

(86) PCT国际申请的申请数据

PCT/JP2014/050573 2014. 01. 15

(87) PCT国际申请的公布数据

W02014/129233 JA 2014. 08. 28

(71) 申请人 三菱电机株式会社

地址 日本东京都

(72) 发明人 古田训 细谷耕佑

(74) 专利代理机构 北京三友知识产权代理有限公司

11127

代理人 李辉 马建军

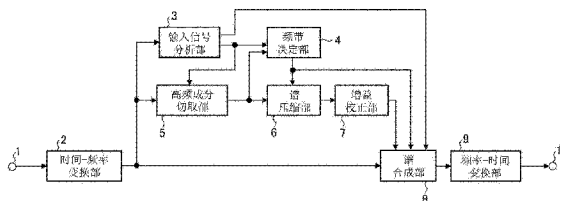
权利要求书1页 说明书11页 附图6页

(54) 发明名称

声音增强装置

(57) 摘要

输入信号分析部 (3) 根据输入信号的状态, 在不超过第 1 频率的范围内决定边界频率。谱压缩部 (6) 沿频率方向压缩比第 1 频率靠上的频带的频率的功率谱。增益校正部 (7) 进行被压缩后的功率谱的增益校正。谱合成部 (8) 将从增益校正部 (7) 输出的功率谱反映到由第 1 频率和边界频率决定的频带。频率 - 时间变换部 (9) 将谱合成部 (8) 的合成功率谱和输入信号的相位谱变换到时域进行输出。



1. 一种声音增强装置,其特征在于,该声音增强装置具有:
时间-频率变换部,其将时域的输入信号变换成作为频域信号的功率谱;
输入信号分析部,其根据所述功率谱分析所述输入信号的状态;
频带决定部,其根据所述输入信号的状态,在不超过预先设定的第1频率的范围内决定边界频率;
谱压缩部,其沿频率方向压缩比所述第1频率靠上的频带的频率的功率谱;
谱合成部,其将所述压缩后的功率谱反映到由所述第1频率和所述边界频率决定的频带中;以及
频率-时间变换部,其将从所述谱合成部输出的合成功率谱和所述输入信号的相位谱变换到时域而得到增强信号。

2. 根据权利要求1所述的声音增强装置,其特征在于,
所述声音增强装置具有增益校正部,该增益校正部以使由所述谱压缩部进行压缩的频带中的压缩前的功率谱的功率与压缩后的功率谱的功率相等的方式,对由该谱压缩部进行压缩的功率谱进行校正,或者通过乘以根据听觉上的因素而决定的规定的校正系数进行待压缩的所述功率谱的功率校正,

所述谱合成部反映由所述增益校正部校正后的功率谱。

3. 根据权利要求1所述的声音增强装置,其特征在于,
所述频带决定部在反映了所述压缩后的功率谱的情况下,将所述边界频率决定为与属于所述第1频率的功率谱之间的功率差最小的频率。

4. 根据权利要求1所述的声音增强装置,其特征在于,
所述频带决定部使用前一帧的边界频率进行当前帧的边界频率的时间方向的平滑化。

5. 根据权利要求1所述的声音增强装置,其特征在于,
所述谱合成部对所述输入信号的功率谱的SN比和所述压缩后的功率谱的SN比进行比较,选择SN比较高的功率谱生成合成功率谱。

6. 根据权利要求1所述的声音增强装置,其特征在于,
所述频带决定部在所述输入信号的SN比较高的情况下,设定较低的所述边界频率,并随着所述SN比降低而设定较高的所述边界频率。

7. 根据权利要求1所述的声音增强装置,其特征在于,
所述谱合成部对输入信号的功率谱和压缩后的功率谱进行加权平均来生成合成功率谱。

8. 根据权利要求1所述的声音增强装置,其特征在于,
在进行压缩时,在规定的频率以内生成功率谱的多个峰值的情况下,所述谱压缩部选择与该多个峰值相当的功率谱中的较大成分,或者在与所述多个峰值相当的功率谱的附近频率处不进行谱压缩。

9. 根据权利要求1所述的声音增强装置,其特征在于,
所述谱合成部在进行合成时,当在规定的频率以内生成功率谱的多个峰值的情况下,对与输入信号的功率谱和高频成分的功率谱各自的峰值相当的成分进行比较,选择较大的成分,或者所述谱压缩部在与所述多个峰值相当的功率谱的附近频率处不进行谱压缩。

声音增强装置

技术领域

[0001] 本发明涉及声音增强装置,用于被导入有声音通信、声音蓄积、声音合成、声音识别系统的车载导航仪、移动电话、内线电话系统(interphone)、收音装置等声音通信系统、免提通话系统、电视会议系统、监视系统、广播系统、声音合成系统等音质改善和声音识别系统的识别率的提高,并改善声音信号的质量及清晰度。

背景技术

[0002] 在模拟电话中,关于通过电话线路发送来的声音信号的频带,例如上限的频率被限制成 3400Hz 的狭窄频带。因此,现有的电话线路的音质不能说太良好。另外,在移动电话、业务无线等的数字声音通信中,由于比特率的严格限制,带宽与模拟线路同样受到限制,因而在这种情况下音质也不能说太良好。对于声音的子音成分、声音具有的“个人性(像某人的)”和“自然性”,时常还处于 3400Hz 以上的频带,但这些频带由于上述的带宽限制而丢失很多。

[0003] 另外,近年来随着声音压缩技术(声音编码技术)的发展,能够以低比特率无线传输宽带(例如上限的频率为 7000Hz)的声音信号。但是,需要发送侧终端和接收侧终端双方支持对应的宽带声音编码/解码方法,而且在双方的基站中也需要完备宽带编码用的网络,因而仅仅是在一部分业务通信系统中得到实际应用,当在公共电话通信网中实施时不仅带来很大的经济负担,而且普及需要很长的时间。

[0004] 为此,进行了压缩声音信号的频率,将其频谱整体控制在通过频带内的尝试。但是,在这种方法中,与原有的声音信号相比,包括含有声音的基本周期(音调(pitch))的低频带的信号在内都被压缩在较低的频率范围内,在该被压缩的信号不再次经过解压缩就被再现的情况下,所再现的声音成为不自然地低的音调,存在声音的个人性消失、质量明显下降的问题。针对此种情况,虽然通过在接收侧终端将压缩信号解压缩能够解决该问题,但是为此需要接收侧终端应对对在发送侧终端被压缩后的高频信号进行再次展开的作业。与应对宽带化同样,对于不具备将压缩信息与声音信号一起发送并再次将其展开的功能的通信终端,这种解决方案不实用。

[0005] 针对上述的问题,例如有专利文献 1 记载的技术。在该技术中,根据不改变声音的音调而仅将规定的高频压缩的方法,能够得到清晰的声音信号。

[0006] 现有技术文献

[0007] 专利文献

[0008] 专利文献 1:日本特开 2011-141551 号公报

发明内容

[0009] 发明要解决的问题

[0010] 但是,在上述专利文献 1 公开的现有技术中,虽然具有声音的个人性得到保留的优点,但是,由于仅仅是将压缩后的高频信号成分映射至低频的预先设定的固定频带中,因

而根据输入信号的声音状态,上述带宽有时不是最佳的,在这种情况下存在音质的劣化不能避免的问题。

[0011] 本发明正是为了解决上述问题而提出的,其目的在于,提供一种能够生成高质量的声音的声音增强装置。

[0012] 用于解决问题的手段

[0013] 本发明的声音增强装置具有:时间-频率变换部,其将时域的输入信号变换成作为频域信号的功率谱;输入信号分析部,其根据功率谱分析输入信号的状态;频带决定部,其根据输入信号的状态,在不超过预先设定的第1频率的范围内决定边界频率;谱压缩部,其沿频率方向压缩比第1频率靠上的频带的功率谱;谱合成部,其将压缩后的功率谱反映到由第1频率和边界频率决定的频带中;以及频率-时间变换部,其将从谱合成部输出的合成功率谱和输入信号的相位谱变换到时域而得到增强信号。

[0014] 发明效果

[0015] 本发明的声音增强装置根据输入信号的状态确定用于决定反映高频成分的频带的边界频率,因而能够防止谱合成造成的异常噪声,能够实现良好且清晰的声音增强处理。

附图说明

[0016] 图1是表示本发明的实施方式1的声音增强装置的结构图。

[0017] 图2是表示本发明的实施方式1的声音增强装置的输入信号分析部的详情的结构图。

[0018] 图3是示意地表示本发明的实施方式1的声音增强装置的处理流程的说明图。

[0019] 图4是表示本发明的实施方式1的声音增强装置的输入信号频谱的一例的说明图。

[0020] 图5是表示基于现有方法的声音增强处理的说明图。

[0021] 图6是表示本发明的实施方式1的声音增强装置的声音增强处理的说明图。

具体实施方式

[0022] 下面,为了更详细地说明本发明,参照附图说明用于实施本发明的方式。

[0023] 实施方式1

[0024] 图1是表示本发明的实施方式1的声音增强装置的结构图。

[0025] 本实施方式1的声音增强装置由输入端子1、时间-频率变换部2、输入信号分析部3、频带决定部4、高频成分切取部5、谱压缩部6、增益校正部7、谱合成部8、频率-时间变换部9、输出端子10构成。另外,图2是表示输入信号分析部3的内部结构的图,输入信号分析部3由自相关分析部11、噪声频谱估计部12、SN比计算部13、功率比分析部14、判定部15构成。

[0026] 输入端子1是被输入声音信号这样的信号作为输入信号的端子。时间-频率变换部2是将时域的输入信号变换成作为频域信号的功率谱的处理部。输入信号分析部3是根据从时间-频率变换部2输出的功率谱分析输入信号的状态,即分析输入信号是声音还是噪声以及在是声音时是母音还是子音的处理部。频带决定部4是根据由输入信号分析部3分析出的输入信号的状态,在不超过预先设定的第1频率的范围内决定边界频率的处理

部。高频成分切取部 5 是切取比第 1 频率靠上的频带的功率谱的处理部。谱压缩部 6 是沿频率方向压缩由高频成分切取部 5 切取出的频带的功率谱的处理部。增益校正部 7 是进行由谱压缩部 6 压缩后的增益校正的处理部。谱合成部 8 是将由增益校正部 7 进行增益校正后的压缩功率谱反映到由第 1 频率和边界频率决定的频带的处理部。频率 - 时间变换部 9 是用于将从谱合成部 8 输出的合成功率谱和输入信号的相位谱变换到时域而得到增强声音（增强信号）的处理部。输出端子 10 是用于将从频率 - 时间变换部 9 输出的增强声音输出到外部的端子。

[0027] 下面, 详细说明图 2 所示的输入信号分析部 3。

[0028] 自相关分析部 11 是根据输入信号的功率谱求出归一化自相关函数的处理部。噪声频谱估计部 12 是根据输入信号的功率谱求出估计噪声频谱的处理部。SN 比计算部 13 是根据时间 - 频率变换部 2 输出的功率谱和噪声频谱估计部 12 输出的估计噪声频谱计算原有帧的平均 SN 比的处理部。功率比分析部 14 是根据输入信号的功率谱求出高频 / 低频功率比的处理部。判定部 15 是输入时间 - 频率变换部 2 输出的功率谱、由自相关分析部 11 得到的归一化自相关函数的最大值、由功率比分析部 14 得到的高频 / 低频功率比以及噪声频谱估计部 12 输出的估计噪声频谱, 进行当前帧的输入信号是声音（母音 / 子音）还是噪声的判定的处理部。

[0029] 下面, 根据图 1 和图 2 说明本发明的声音增强装置的动作原理。

[0030] 首先, 在对通过传声器等获取的声音、音乐等进行 A/D (模拟 / 数字) 变换后, 按照规定的采样频率 (例如 16kHz) 进行采样, 并且分割成帧单位 (例如 10ms), 通过输入端子 1 输入到声音增强装置。另外, 在本实施方式中, 将最终得到的处理信号的频带设为 0Hz ~ 3400Hz (第 1 频率), 将 0Hz ~ 3400Hz 称作通过频带, 将 3400Hz ~ 8000Hz 称作高频频带进行说明。另外, 如果没有特殊指定, 则假定对包含 0Hz ~ 8000Hz 的通过频带和高频频带的信号进行处理。并且, 将高频频带的信号也称作高频成分进行说明。

[0031] 时间 - 频率变换部 2 在对分割成帧单位后的输入信号 $x(t)$ 进行例如汉宁窗处理后, 例如下式 (1) 所示进行 512 点的高速傅里叶变换 (Fast Fourier Transform : FFT), 从时域的信号变换成作为频域信号的频谱成分 $X(\lambda, k)$ 。

$$[0032] \quad X(\lambda, k) = \text{FT}[x(t)] \quad (1)$$

[0033] 其中, t 表示采样时间, λ 表示对输入信号进行帧分割时的帧编号, k 表示指定频谱的频带的频率成分的编号 (以下称作频谱编号), $\text{FT}[\cdot]$ 表示高速傅里叶变换处理。然后, 使用下式 (2) 从输入信号的频谱成分得到功率谱 $Y(\lambda, k)$, 并且得到相位谱 $\theta(\lambda, k)$ 。

$$[0034] \quad Y(\lambda, k) = \sqrt{\text{Re}\{X(\lambda, k)\}^2 + \text{Im}\{X(\lambda, k)\}^2} ; \quad 0 \leq k < 256 \quad (2)$$

[0035] 其中, $\text{Re}\{X(\lambda, k)\}$ 和 $\text{Im}\{X(\lambda, k)\}$ 分别表示傅里叶变换后的输入信号频谱的实数部和虚数部。

[0036] 输入信号分析部 3 进行当前帧的输入信号是声音还是噪声的判定。并且, 在是声音时, 还进行声音是母音还是子音的判定。首先, 在自相关分析部 11 中, 使用下式 (3) 从功率谱 $Y(\lambda, k)$ 求出归一化自相关函数 $\rho_N(\lambda, \tau)$ 。

$$\rho(\lambda, \tau) = FT[Y(\lambda, k)],$$

[0037]

$$\rho_N(\lambda, \tau) = \frac{\rho(\lambda, \tau)}{\rho(\lambda, 0)}, \quad (3)$$

[0038] 其中, τ 表示延迟时间, $FT[\cdot]$ 表示傅里叶变换处理, 例如可以与式 (1) 同样以点数 = 512 进行高速傅里叶变换。另外, 式 (3) 是维纳-辛钦 (Wiener-Khintchine) 定理, 因而省略说明。然后, 使用式 (4) 求出归一化自相关函数的最大值 $\rho_{\max}(\lambda)$ 。其中, 式 (4) 表示在 $32 \leq \tau \leq 192$ 的范围内检索 $\rho(\lambda, \tau)$ 的最大值。

$$\rho_{\max}(\lambda) = \max[\rho(\lambda, \tau)], 32 \leq \tau \leq 192 \quad (4)$$

[0040] 然后, 在 SN 比计算部 13 中, 输入时间-频率变换部 2 输出的功率谱 $Y(\lambda, k)$ 和所述的噪声频谱估计部 12 输出的估计噪声频谱 $N(\lambda, k)$, 计算当前帧的平均 SN 比 $SNR_{AVE}(\lambda)$ 。

$$SNR_{AVE}(\lambda) = 10 \cdot \log_{10} \left(\frac{S_{pow}(\lambda)}{N_{pow}(\lambda)} \right) \quad (5)$$

[0041]

$$S_{pow}(\lambda) = \sum_{k=0}^{255} Y(\lambda, k), \quad N_{pow}(\lambda) = \sum_{k=0}^{255} N(\lambda, k)$$

[0042] 其中, 在式 (5) 中, $N(\lambda, k)$ 表示估计噪声频谱, S_{pow} 和 N_{pow} 分别表示输入信号的功率谱的总和、估计噪声频谱的总和。

[0043] 并且, 根据式 (6) 计算子带 SN 比 $SNR(\lambda, k)$ 作为每个频谱成分的 SN 比, 并输出到后述的谱合成部 8。

$$SNR(\lambda) = 10 \cdot \log_{10} (Y(\lambda)/N(\lambda)) \quad (6)$$

[0045] 功率比分析部 14 求出当前帧的输入信号的高频功率与低频功率之比 (高频/低频功率比)。该比近似从低频朝向高频的频谱倾斜的程度, 在高频功率大于低频功率的情况下, 子音的可能性较大, 通过使用该值, 能够判定输入信号的状态是母音还是子音。具体而言, 例如按照式 (7) 所示, 求出 125Hz ~ 3000Hz 的频带与 3000Hz ~ 6000Hz 的频带的功率比。

$$P_{Ratio}(\lambda) = \frac{P_{S_High}(\lambda)}{P_{S_Low}(\lambda)} \quad (7)$$

[0047] 其中, $P_{S_Low}(\lambda)$ 表示输入信号的 125Hz ~ 3000Hz 的频带功率, $P_{S_High}(\lambda)$ 表示输入信号的 3000Hz ~ 6000Hz 的频带功率, 按照以下所述计算。

$$P_{S_Low}(\lambda) = \frac{1}{(96-4+1)} \sum_{k=4}^{96} Y(\lambda, k)$$

[0048]

$$P_{S_High}(\lambda) = \frac{1}{(192-97+1)} \sum_{k=97}^{192} Y(\lambda, k) \quad (8)$$

[0049] 将以上得到的当前帧的高频/低频功率比 $P_{Ratio}(\lambda)$ 输出到判定部 15。

[0050] 判定部 15 输入时间-频率变换部 2 输出的功率谱 $Y(\lambda, k)$ 、由自相关分析部 11 得到的归一化自相关函数的最大值 $\rho_{\max}(\lambda)$ 、由功率比分析部 14 得到的高频/低频功率比 $P_{Ratio}(\lambda)$ 以及后述的噪声频谱估计部 12 输出的估计噪声频谱 $N(\lambda, k)$, 进行当前帧的输入

信号是声音（母音 / 子音）还是噪声的判定, 将其结果作为判定标志进行输出。关于声音 / 噪声区间的判定方法, 例如在满足下式 (9) 的条件(的情况下, 视为声音而将判定标志 Vflag 设定为“1(声音)”进行输出, 在除此以外的情况下, 视为噪声而将判定标志 Vflag 设定为“0(噪声)”进行输出。

$$[0051] \quad Vflag = \begin{cases} 1, & \text{if } SNR_{AVE}(\lambda) > TH_{FR_SN} \text{ or } \rho_{\max}(\lambda) > TH_{ACF} \\ 0, & \text{Otherwise} \end{cases} \quad (9)$$

[0052] 另外, 当在式 (9) 中判定为声音的情况下, 使用下式 (10), 在功率比分析部 14 输出的高频 / 低频功率比 $P_{Ratio}(\lambda)$ 超过规定的阈值的的情况下, 判定为子音, 将判定标志 Vflag 设定为“2(子音)”进行输出。

$$[0053] \quad Vflag = \begin{cases} 2, & \text{if } P_{Ratio}(\lambda) > TH_{POW_Ratio} \text{ and } Vflag = 1 \\ 1, & \text{Otherwise} \end{cases} \quad (10)$$

[0054] 其中, TH_{FR_SN} 、 TH_{ACF} 以及 TH_{POW_Ratio} 表示判定用的规定的常数阈值, 作为优选的示例, $TH_{FR_SN} = 3.0$, $TH_{ACF} = 0.3$, $TH_{POW_Ratio} = 1.25$, 但也能够根据输入信号的状态和噪声电平进行适当变更。

[0055] 另外, 在本实施方式中, 作为输入声音分析方法采用了自相关函数法、输入信号的平均 SN 比以及高频 / 低频功率比, 但不限于这种方法, 也可以采取使用开普勒分析等公知手法等的其它方法。并且, 按照本领域技术人员的自由裁量, 通过组合各种公知的方法例如零交叉数等, 还能够提高判定精度。

[0056] 噪声频谱估计部 12 输入时间 - 频率变换部 2 输出的功率谱 $Y(\lambda, k)$ 和判定部 15 输出的判定标志 Vflag, 按照下式 (11) 和判定标志 Vflag 进行噪声频谱的估计和更新, 并输出估计噪声频谱 $N(\lambda, k)$ 。

[0057]

$$N(\lambda, k) = \begin{cases} (1-\alpha) \cdot N(\lambda-1, k) + \alpha \cdot |Y(\lambda, k)|^2 & \text{if } Vflag = 0 \\ N(\lambda-1, k) & \text{if } Vflag = 1 \text{ or } Vflag = 2 \end{cases} ; 0 \leq k < 256 \quad (11)$$

[0058] 其中, $N(\lambda-1, k)$ 表示前一帧的估计噪声频谱, 保存在噪声频谱估计部 12 内的例如 RAM(Random Access Memory) 等存储单元中。在式 (11) 中, 在判定标志 $Vflag = 0$ 的情况下, 判定当前帧的输入信号是噪声, 因而使用输入信号的功率谱 $Y(\lambda, k)$ 和更新系数 α , 进行前一帧的估计噪声频谱 $N(\lambda-1, k)$ 的更新。另外, 更新系数 α 是 $0 < \alpha < 1$ 的范围内的规定常数, 作为优选的示例, $\alpha = 0.95$, 但也能够根据输入信号的状态和噪声电平进行适当变更。

[0059] 另一方面, 在判定标志 $Vflag = 1$ 或者判定标志 $Vflag = 2$ 的情况下, 当前帧的输入信号是声音, 将前一帧的估计噪声频谱 $N(\lambda-1, k)$ 直接作为当前帧的估计噪声频谱 $N(\lambda, k)$ 进行输出。

[0060] 频带决定部 4 输入输入信号分析部 3 输出的判定标志 Vflag, 对于通过频带决定表示合成 (或者卷积) 后述高频成分的带宽的边界频率 $B(\lambda)$ 。边界频率 $B(\lambda)$ 例如能够使用式 (12) 和式 (13) 决定。在式 (12) 中, 在判定当前帧是噪声区间的的情况下, 将带宽设为最小, 即将边界频率设为接近通过频带的上限频率 3400Hz 的值 (3300Hz)。另外, 在当前帧

是声音区间而且是母音的情况下,声音的峰段(formant)往往集中在 3000Hz 以下的频率,为了保持 3000Hz 以下的峰段并反映高频成分,将设定的边界频率设为 3000Hz,并将高频成分卷积到 3000Hz ~ 3400Hz 的带宽。在是子音的情况下,子音成分往往存在于 3400Hz 以上的高频,因而为了将该成分更多地反映到通过频带而扩宽频带,即将边界频率设为比母音的边界频率低的值(2500Hz),将高频成分卷积到 2500Hz ~ 3400Hz 的频带。

$$[0061] \quad B_F = \begin{cases} 3300(\text{Hz}), & Vflag = 0 \\ 2500(\text{Hz}), & Vflag = 2 \\ 3000(\text{Hz}), & Vflag = 1 \end{cases} \quad (12)$$

[0062] 当由上述的式(12)决定了基本的边界频率后,参照要切取的高频频率的最低频率(3400Hz)下的频谱成分的功率,使边界频率上下移动而调整成最佳的频率,以便减轻因后述的频率合成时的功率差异(功率谱的阶梯差)而产生的异常噪声。具体而言,例如当在式(12)中选择了 $B_F = 3000\text{Hz}$ 的情况下,在其上下 100Hz 的范围(2900Hz ~ 3100Hz)内,将假定的边界频率的功率谱与 3400Hz 的功率谱进行比较,在功率差最小的假定的边界频率是 2980Hz 的情况下,修正成 $B_F = 2980\text{Hz}$ 。

[0063] 然后,对于修正后的边界频率 B_F ,如式(13)所示使用前一帧的边界频率 $B(\lambda - 1)$ 进行边界频率的时间方向的平滑化,以便抑制由于帧间带宽的急剧变化而引起的异常噪声的产生。

$$[0064] \quad \begin{cases} B(\lambda) = (1 - W_B) \cdot B(\lambda - 1) + W_B \cdot B_F, & \lambda > 0 \\ B(\lambda) = 3300, & \lambda = 0 \end{cases} \quad (13)$$

[0065] 其中, W_B 表示时间方向平滑化用的规定常数,作为优选的示例, $W_B = 0.8$, 但能够进行适当变更,以便按照输入信号的类型等良好地进行动作。

[0066] 高频成分切取部 5 根据输入信号分析部 3 输出的判定标志进行高频成分的切取。在本实施方式中,高频成分的频率范围是 3400Hz ~ 8000Hz,因而该范围的功率谱 $Y(\lambda, k)$ 被切取出来,作为高频成分的功率谱 $Y_h(\lambda, k)$ 进行输出。

[0067] 另外,在判定标志是声音区间的情况下进行切取,而在判定标志是噪声区间(或者无音区间)的情况下也可以不进行切取。在这种情况下,不进行后述的谱压缩处理和谱合成处理,因而能够削减处理量。将切取出的高频成分发送到后级的谱压缩部 6。

[0068] 谱压缩部 6 以能够将按照高频成分切取部 5 切取出的高频成分的功率谱 $Y_h(\lambda, k)$ 与通过频带上的频谱进行合成的方式,进行频谱的频率方向的频带压缩,并使其与通过频带的带宽一致。

[0069] 将边界频率 $B(\lambda)$ 和通过频带的上限频率 $B_{NL} = 3400\text{Hz}$ 呈现出的带宽,即,在设向通过频带反映高频成分的带宽为 $B_{WC}(\lambda)$, 设高频频带的上限频率 $B_{NH} = 8000\text{Hz}$ 与通过频带的上限频率 $B_{NL} = 3400\text{Hz}$ 之间的频带即高频成分的带宽为 B_{WH} 时,频谱频带的线性压缩例如能够用下式(14)表示。

[0070]

$$Y_m(\lambda, k + B(\lambda)) = Y(\lambda, B_{NL} + \lfloor k \cdot B_{WH} / B_{WC} \rfloor), \quad k = 0, \dots, B_{WC}(\lambda) \quad (14)$$

[0071] 其中, $Y_m(\lambda, k)$ 是被实施频率压缩后的高频成分的功率谱。另外, $\lfloor x \rfloor$ 是取整函数,

给出不超过 x 的最大整数。作为具体例,例如在 $B(\lambda) = 3050\text{Hz}$ 的情况下,以在通过频带上更狭窄的 $3050\text{Hz} \sim 3400\text{Hz}$ 的范围内反映作为高频成分的 $3400\text{Hz} \sim 8000\text{Hz}$ 的频谱的方式,进行频率方向的压缩。

[0072] 增益校正部 7 进行被实施频率压缩后的高频成分的功率谱 $Y_m(\lambda, k)$ 的功率校正。基本上是以使压缩前的频带即压缩前的 $3400\text{Hz} \sim 8000\text{Hz}$ 的频带的功率与压缩后的频谱成分的功率相等的方式进行校正,也可以乘以校正系数以使听觉上更好。将被实施频率压缩以及增益校正后的高频成分的功率谱 $\hat{Y}_m(\lambda, k)$ 输出到谱合成部 8。

[0073] 谱合成部 8 使用被实施频率压缩以及增益校正后的高频成分的功率谱 $\hat{Y}_m(\lambda, k)$ 和通过频带的功率谱 $Y(\lambda, k)$, 按照下式将谱合成而得到增强声音的功率谱 $Y_{EMP}(\lambda, k)$ 。

[0074] 在谱合成中,如式 (15) 所示使用输入信号分析部 3 输出的子带 SN 比,对通过频带的频谱成分的子带 SN 比 $SNR(\lambda, k)$ 和被实施频率压缩后的高频的频谱成分的子带 SN 比 $SNR_m(\lambda, k)$ 进行比较,通过选择子带 SN 比较大的一方而得到增强声音的功率谱。

$$Y_{EMP}(\lambda, k) = \begin{cases} \hat{Y}_m(\lambda, k), & \text{if } SNR_m(\lambda, k) > SNR(\lambda, k) \\ Y(\lambda, k), & \text{if } SNR_m(\lambda, k) \leq SNR(\lambda, k) \end{cases},$$

[0075]

$$k = B(\lambda), \dots, B_{w_c}(\lambda) + B(\lambda) \quad (15)$$

[0076] 其中, $B_{w_c}(\lambda)$ 与在谱压缩部 6 中叙述的同样,是边界频率 $B(\lambda)$ 和通过频带的上限频率 $B_{NL} = 3400\text{Hz}$ 呈现出的带宽。然后,在谱合成后,在边界频率附近的频带中进行频率间的平滑化。在平滑化中使用例如移动平均滤波器等公知的方法,以在边界频率的上下 100Hz (200Hz 的宽度) 内变平滑的方式进行平滑化。通过平滑化,能够进一步缓解谱合成时的功率谱的功率差(阶梯差),具有抑制异常噪声产生的效果。

[0077] 将以上进行平滑化后的增强声音的功率谱 $\hat{Y}_{EMP}(\lambda, k)$ 输出到频率-时间变换部 9。

[0078] 频率-时间变换部 9 输入进行平滑化后的增强声音的功率谱 $\hat{Y}_{EMP}(\lambda, k)$ 和输入声音的相位谱 $\theta(\lambda, k)$, 在实施了逆 FFT 后,实施从 16kHz 变换成 8kHz 的采样信号的降频采样处理,通过低通滤波器生成 $0\text{Hz} \sim 3400\text{Hz}$ 频带的时间信号,从输出端子 10 输出。

[0079] 图 3 是为了更易于理解地说明而示意地图示出本实施方式 1 的一系列的动作原理的图。(a) 表示被输入的声音信号的功率谱,(b) 表示高频成分切取部 5 的输出即高频成分的功率谱,(c) 表示谱压缩部 6 的输出即被实施频率压缩后的高频成分的功率谱,(d) 表示增益校正部 7 的输出即被实施频率压缩以及增益校正后的高频成分的功率谱。另外,箭头表示处理的顺序。

[0080] 图 4 是声音区间的输入信号频谱的一例的图。(a) 是子音区间的声音信号,(b) 是虽然是在子音区间,但是 3400Hz 以上的高频成分的功率较小且缺乏子音特征的信号。图 5 是对图 4 所示的输入信号实施现有方法的处理时的一例,图 6 是对该图 4 所示的输入信号实施本实施方式 1 的处理时的一例。在图 5 的现有方法中,如 (a) 的箭头 500a、(b) 的箭头 500b 所示,带宽是固定的,因而对于没有子音特征的 (b) 的声音信号,在处理声音中产生异常噪声(如虚线部分 501 所示,在频谱中产生峰值成分而产生异常噪声),而在图 6 的本发明的方法中,如 (a) 的箭头 600a、(b) 的箭头 600b 所示,带宽能够根据输入信号的状态而变

化,因而 (a) 的声音信号不仅将子音特征进一步反映到通过频带 (如范围 601a 所示,高频成分被反映到更宽范围的低频部分),而且即使在 (b) 的没有子音特征的情况下,也能够实现不产生异常噪声的良好的声音增强处理 (如范围 601b 所示,峰值成分的产生受到抑制,不产生异常噪声)。

[0081] 以上,根据本实施方式 1 的声音增强装置,具有:时间-频率变换部,其将时域的输入信号转换成作为频域信号的功率谱;输入信号分析部,其根据功率谱分析输入信号的状态;频带决定部,其根据输入信号的状态,在不超出预先设定的第 1 频率的范围内决定边界频率;谱压缩部,其沿频率方向压缩比第 1 频率靠上的频带的功率谱;谱合成部,其将压缩后的功率谱反映到由第 1 频率和边界频率决定的频带中;以及频率-时间变换部,其将从谱合成部输出的合成功率谱和输入信号的相位谱变换到时域而得到增强信。因而能够按照输入信号的状态反映通过频带以上的高频信号而进行输入信号的增强,因此,能够将高频的特征恰当地反映到通过频带,能够实现没有谱合成造成的异常噪声的、良好且清晰的声音增强处理。

[0082] 另外,根据本实施方式 1 的声音增强装置,具有增益校正部,该增益校正部以使由谱压缩部进行压缩的频带中的压缩前的功率谱的功率与压缩后的功率谱的功率相等的方式,对由谱压缩部进行压缩的功率谱进行校正,或者通过乘以根据听觉上的因素而决定的规定的校正系数进行待压缩的功率谱的功率校正,谱合成部反映由增益校正部校正后的功率谱。因而能够抑制谱合成时发生异常噪声,实现良好的声音增强处理。

[0083] 另外,根据本实施方式 1 的声音增强装置,频带决定部在反映了压缩后的功率谱的情况下,将边界频率决定为与属于第 1 频率的功率谱之间的功率差最小的频率,因而能够抑制谱合成时发生异常噪声,实现良好的声音增强处理。

[0084] 另外,根据本实施方式 1 的声音增强装置,频带决定部使用前一帧的边界频率进行当前帧的边界频率的时间方向的平滑化,因而能够抑制谱合成时发生异常噪声,实现良好的声音增强处理。

[0085] 另外,根据本实施方式 1 的声音增强装置,谱合成部对输入信号的功率谱的 SN 比和压缩后的功率谱的 SN 比进行比较,选择 SN 比较高的功率谱生成合成功率谱。因而在与高频成分的谱合成时,能够防止在高频频带信号的子带 SN 比较低的情况下通过频带不反映该频谱成分,即防止通过频带反映劣化的高频成分,因此,能够在抑制声音的劣化感增大的同时,实现良好的声音增强处理。

[0086] 实施方式 2

[0087] 在实施方式 1 中,是在频带决定部 4 决定边界频率时,通过使用输入信号分析结果即判定标志进行与输入信号的状态对应的控制,但不限于此,例如也能够使用 SN 比计算部 13 输出的平均 SN 比和功率比分析部 14 输出的高频/低频功率比进行控制。此时,例如也可以按照两个 SN 比的值获取 2500Hz ~ 3400Hz 之间的连续值,而不是按照式 (11) 所示用 3 个离散值表述边界频率。

[0088] 具体而言,如果平均 SN 比 $SNR_{AVE}(\lambda)$ 大,则输入信号是声音的可能性大,因而降低边界频率 $B(\lambda)$,扩大反映高频成分的带宽,并且,如果高频/低频功率比 $P_{Ratio}(\lambda)$ 大,则是子音的可能性大,因而降低边界频率 $B(\lambda)$ 。相反,如果平均 SN 比 $SNR_{AVE}(\lambda)$ 低,则提高边界频率 $B(\lambda)$,缩小反映高频成分的带宽。

[0089] 根据本实施方式 2,能够按照输入信号的 SN 比连续地控制边界频率,因而在能够设定成与输入信号的 SN 比对应的最佳带宽后,输入信号的平均 SN 比低的情况下,通过缩小频带,能够抑制将多余的高频成分反映到通过频带,因而能够防止声音的劣化感增大,实现更加良好的声音增强处理。

[0090] 如以上说明的那样,根据实施方式 2 的声音增强装置,频带决定部在输入信号的 SN 比高的情况下,设定较低的边界频率,并随着 SN 比降低而设定较高的边界频率,因而能够防止声音的劣化感增大,实现更加良好的声音增强处理。

[0091] 实施方式 3

[0092] 在实施方式 1 中,是在谱合成部 8 中对高频成分的子带 SN 比和通过频带的子带 SN 比进行比较,按照子带 SN 比进行谱合成,但也可以不使用子带 SN 比选择频谱成分,而是如式 (16) 所示,对于每个频谱成分获取输入信号的功率谱 $Y(\lambda, k)$ 与被实施频率压缩及增益校正后的高频成分的功率谱 $Y_m(\lambda, k)$ 的加权平均进行合成,求出增强信号的功率谱。

$$[0093] \quad Y_{EMP}(\lambda, k) = (1 - W_s(k)) \cdot Y(\lambda, k) + W_s(k) \cdot \hat{Y}_m(\lambda, k) \quad (16)$$

[0094] 其中, $W_s(k)$ 是加权用的规定的常数,作为优选的示例,能够如式 (17) 所示随着频率提高而增大高频成分的功率谱的权重,但也能够按照输入信号的类型等进行适当变更,以便良好地进行动作。

$$[0095] \quad W_s(k) = 0.5 + 0.3 \cdot (k/B_{w_c})、k = 0, \dots, B_{w_c}(\lambda) \quad (17)$$

[0096] 其中, $B_{w_c}(\lambda)$ 与在谱压缩部 6 中叙述的同样,是边界频率 $B(\lambda)$ 和通过频带的上限频率 $B_{NL} = 3400\text{Hz}$ 呈现的带宽,并且,决定 $W_s(k)$ 的常数值能够按照输入信号的状态预先调整成合适的值。

[0097] 如上所述,根据实施方式 3 的声音增强装置,谱合成部对输入信号的功率谱和压缩后的功率谱进行加权平均来生成合成功率谱,因而具有能够缓解频率方向的频谱的功率不连续的效果。另外,例如通过随着频率提高而增大高频成分的功率谱的权重,能够在高频反映更多的高频成分,能够实现更加自然且清晰的声音增强处理。

[0098] 实施方式 4

[0099] 在上述的实施方式 1 中,是在谱压缩部 6 沿频率方向压缩谱时,在接近以峰段频率为代表的频谱的峰值(频谱的山谷构造的“山脊”部分)的情况下,有时增强声音产生回响感。为了抑制这种回响感,例如对与峰值相当的功率谱进行比较而选择成分较大者,或者在与该峰值相当的功率谱的附近频率处不进行谱压缩,由此分开配置峰值。

[0100] 如以上说明的那样,根据实施方式 4 的声音增强装置,在进行压缩时,在规定的频率以内生成功率谱的多个峰值的情况下,谱压缩部选择与多个峰值相当的功率谱中的较大成分,或者在与多个峰值相当的功率谱的附近频率处不进行谱压缩,因而能够将在谱压缩时有可能成为回响感原因的频谱的峰值分开配置,因此,能够抑制增强声音的异常噪声感,实现良好的声音增强处理。

[0101] 实施方式 5

[0102] 作为实施方式 4 的变形,下面将针对在谱合成部 8 的谱合成时生成的功率谱的峰值的处理作为实施方式 5 进行说明。

[0103] 在谱合成部 8 中,在将输入信号的功率谱和被实施谱压缩后的高频成分的功率谱

合成的结果是接近以峰段频率为代表的频谱的峰值的情况下,有时增强声音产生回响感。为了抑制这种回响感,例如进行如下的处理:对与输入信号的功率谱和高频成分的功率谱各自的峰值相当的成分进行比较,选择成分较大者,或者在与该峰值相当的功率谱的附近频率处不进行谱压缩而将峰值分开合成。因此,能够将在谱压缩时有可能成为回响感原因的频谱的峰值分开配置。

[0104] 如以上说明的那样,根据实施方式 5 的声音增强装置,谱合成部在进行合成时,当在规定的频率以内生成功率谱的多个峰值的情况下,对与输入信号的功率谱和高频成分的功率谱各自的峰值相当的成分进行比较,选择较大的成分,或者谱压缩部在与多个峰值相当的功率谱的附近频率处不进行谱压缩。因而能够将在谱压缩时有可能成为回响感原因的频谱的峰值分开配置,因此,能够抑制增强声音的异常噪声感,能够实现良好的声音增强处理。

[0105] 另外,在上述各实施方式中,以通过频带为 3400Hz 进行了说明,但不限于此,例如也能够实现 7000Hz 的宽带传输。在这种情况下,例如能够通过输入 11kHz 频带的输入信号并进行分析来实现。

[0106] 本发明实现超过通过频带的宽带感和清晰性,并且不需要在接收侧进行高频成分的解压缩处理,即不需要接收终端侧的追加处理,因而具有存储量和处理量不会增加且无论何种类型的接收终端都能够提高质量的效果。

[0107] 另外,本发明不是仅以声音为对象,也能够适用于音乐等非声音信号。此时,将输入信号分析部 3 更换成适应于乐音的部件,使用与相当于声音的子音/母音的乐音对应的公知的分析单元进行判定即可。

[0108] 进而,本发明不仅作为针对无线通信传输时的带宽限制的音质提高对策是有效的,而且在扬声器的高频再现能力不足的情况下或在扩声广播等中高频成分衰减的情况下也是有效的,能够将高频成分的特征反映到扬声器能够再现的频带或不易衰减的低频频带,因此,具有能够再现清晰声音的效果。并且,在以电视的节目表的声音读取、玩具的合成声音输出等为代表的成本要求严格的家电、玩具等中,在由于蓄积装置的存储容量限制或 D/A(数字/模拟)变换的限制而限制了采样频率的情况下,也具有能够再现在听觉上超过采样频率的清晰的声音信号的效果。

[0109] 另外,在上述各个实施方式中,是将被实施增强处理后的输出信号以数字数据形式发送到声音编码装置、声音识别装置、声音蓄积装置、免提通话装置等各种声音音响处理装置,但各个实施方式的声音增强装置也能够单独实现或者与上述其它装置一起利用 DSP(数字信号处理用处理器)实现,还可以作为软件程序执行来实现。程序可以存储在执行软件程序的计算机装置的存储装置中,也可以是通过 CD-ROM 等存储介质发布的形式。另外,也能够通过网络提供程序。并且,除了发送到各种声音音响处理装置外,还可以在进行了 D/A 变换后通过放大装置进行放大,作为声音信号直接从扬声器等进行输出。

[0110] 另外,本发明能够在本发明的范围内对各个实施方式进行自由组合,或者对各个实施方式的任意构成要素进行变形,或者在各个实施方式中省略任意构成要素。

[0111] 产业上的可利用性

[0112] 本发明的声音增强装置具有频带决定部,该频带决定部根据输入信号的状态,在不超过预先设定的第 1 频率的范围内决定反映高频成分的频带,能够防止谱合成造成的异

常噪声,能够实现良好且清晰的声音增强处理,因而适合用于导入有声音通信、声音蓄积、声音合成、声音识别系统的车载导航仪、移动电话、内线电话系统、收音装置等。

[0113] 标号说明

[0114] 1 输入端子 ;2 时间-频率变换部 ;3 输入信号分析部 ;4 频带决定部 ;5 高频成分切取部 ;6 谱压缩部 ;7 增益校正部 ;8 谱合成部 ;9 频率-时间变换部 ;10 输出端子 ;11 自相关分析部 ;12 噪声频谱估计部 ;13 SN 比计算部 ;14 功率比分析部 ;15 判定部。

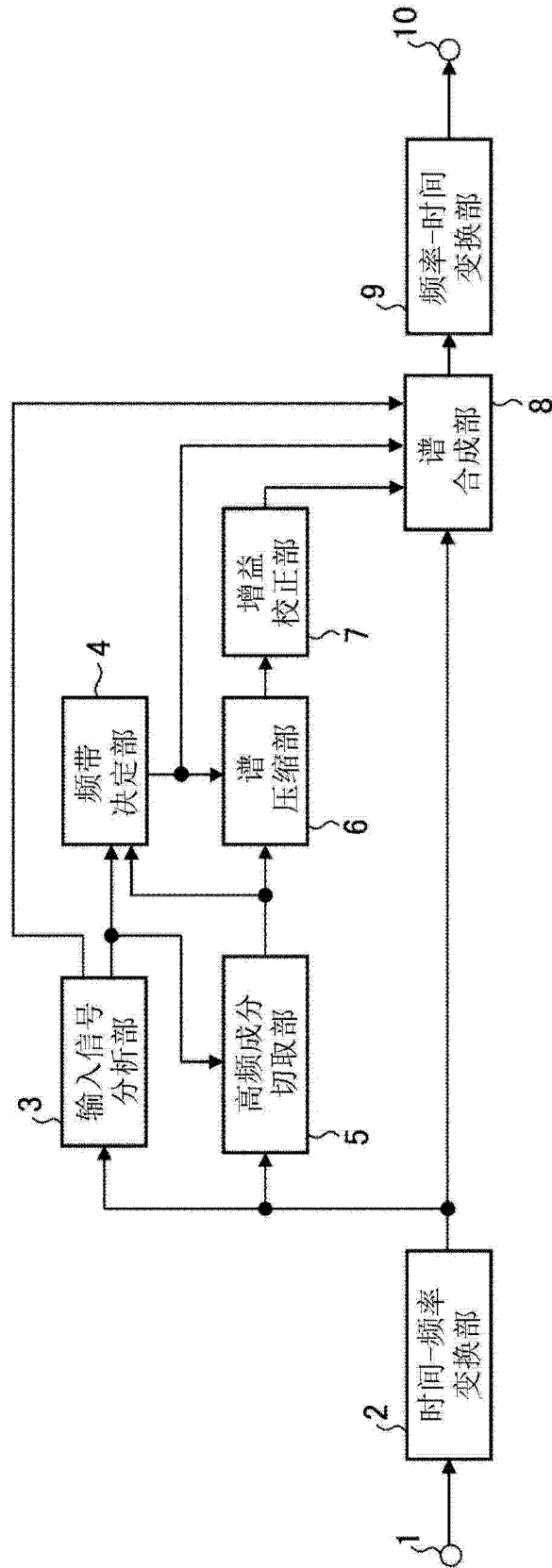


图 1

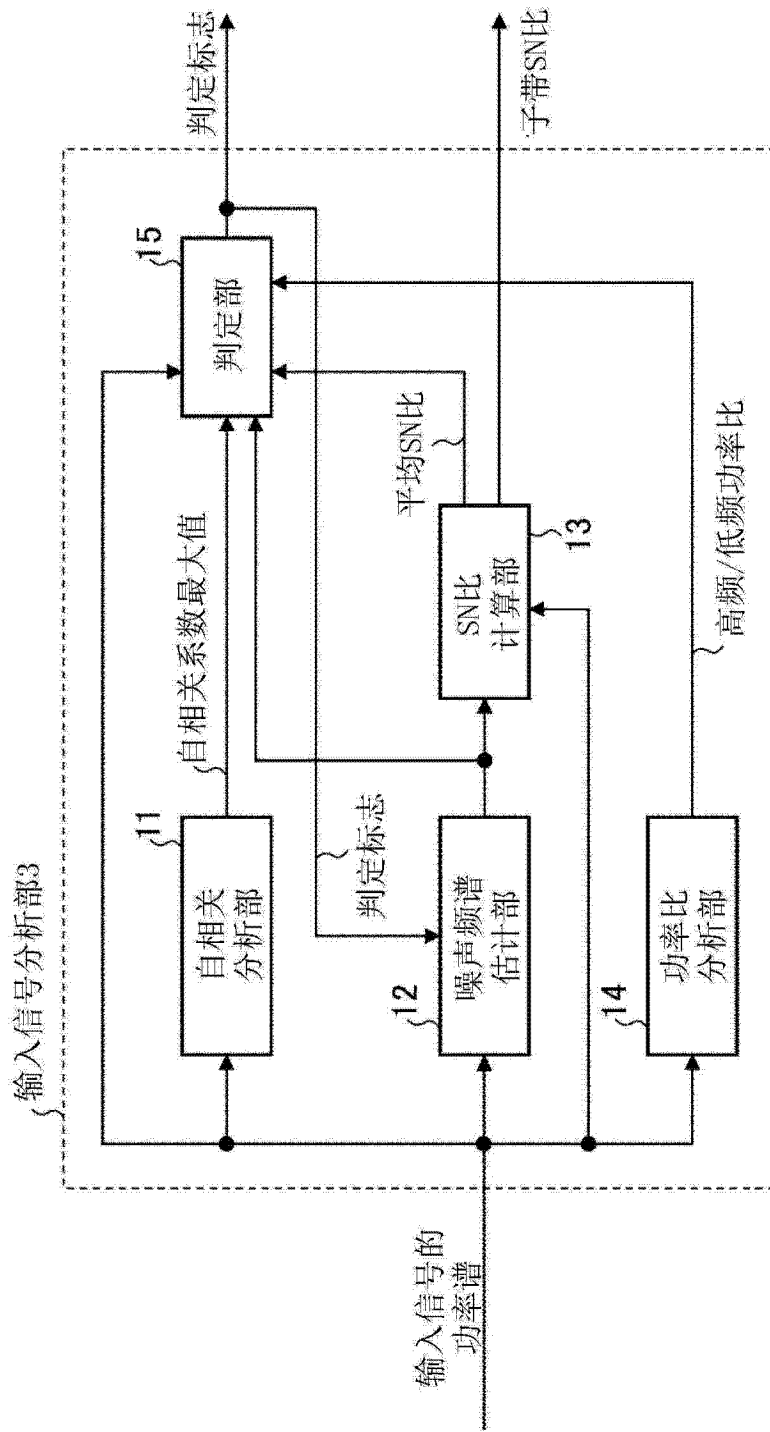


图 2

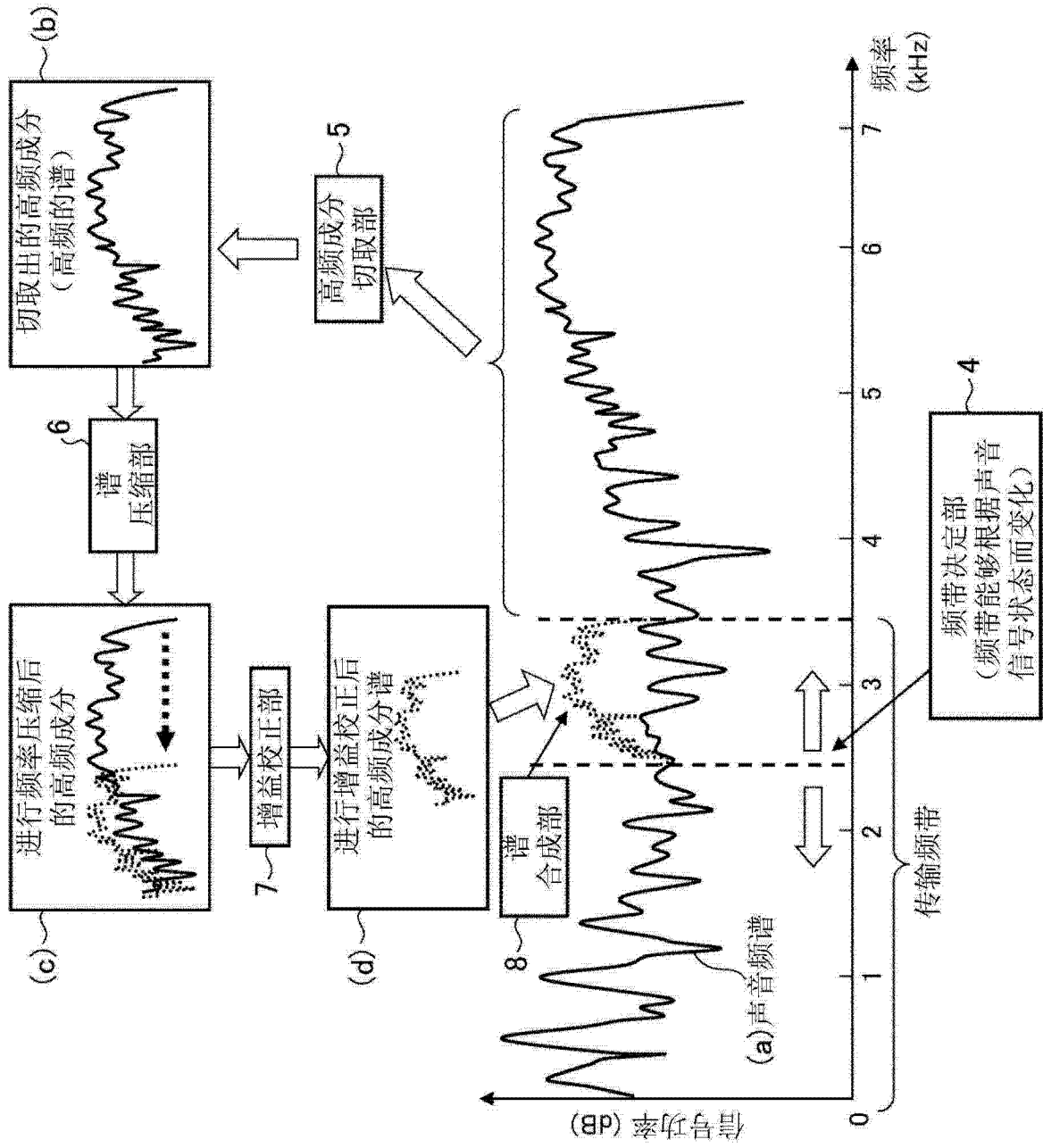


图 3

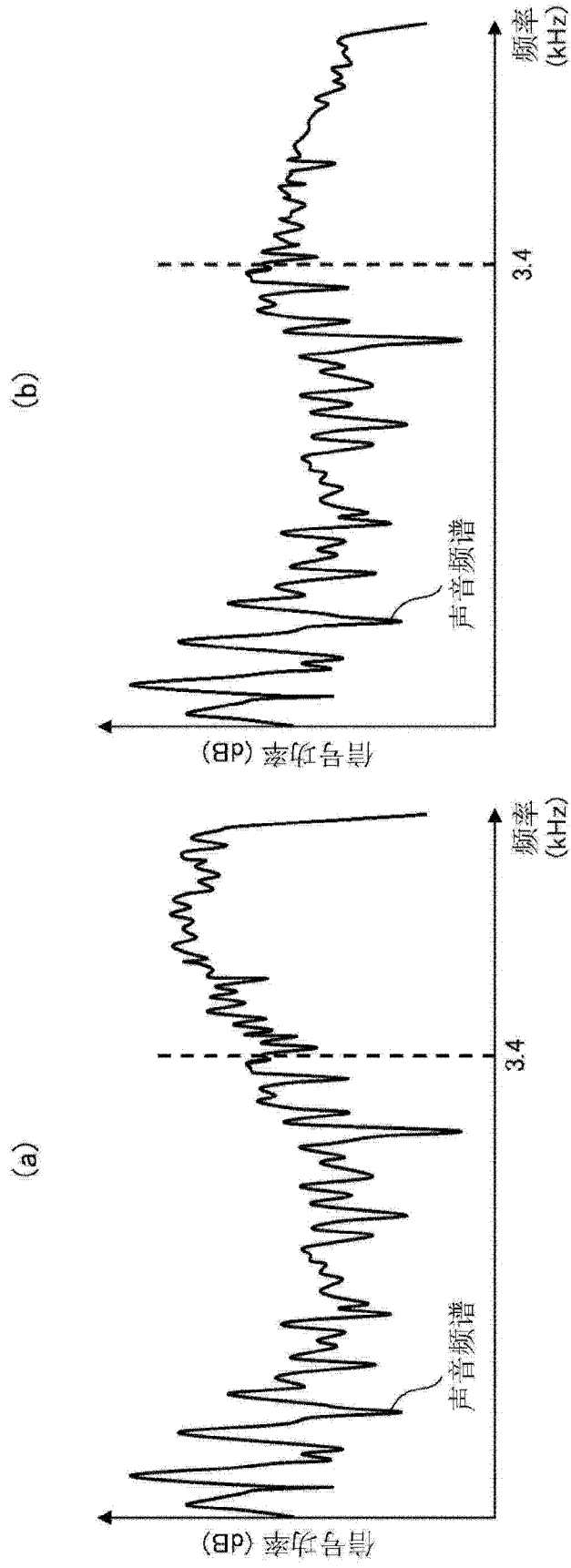


图 4

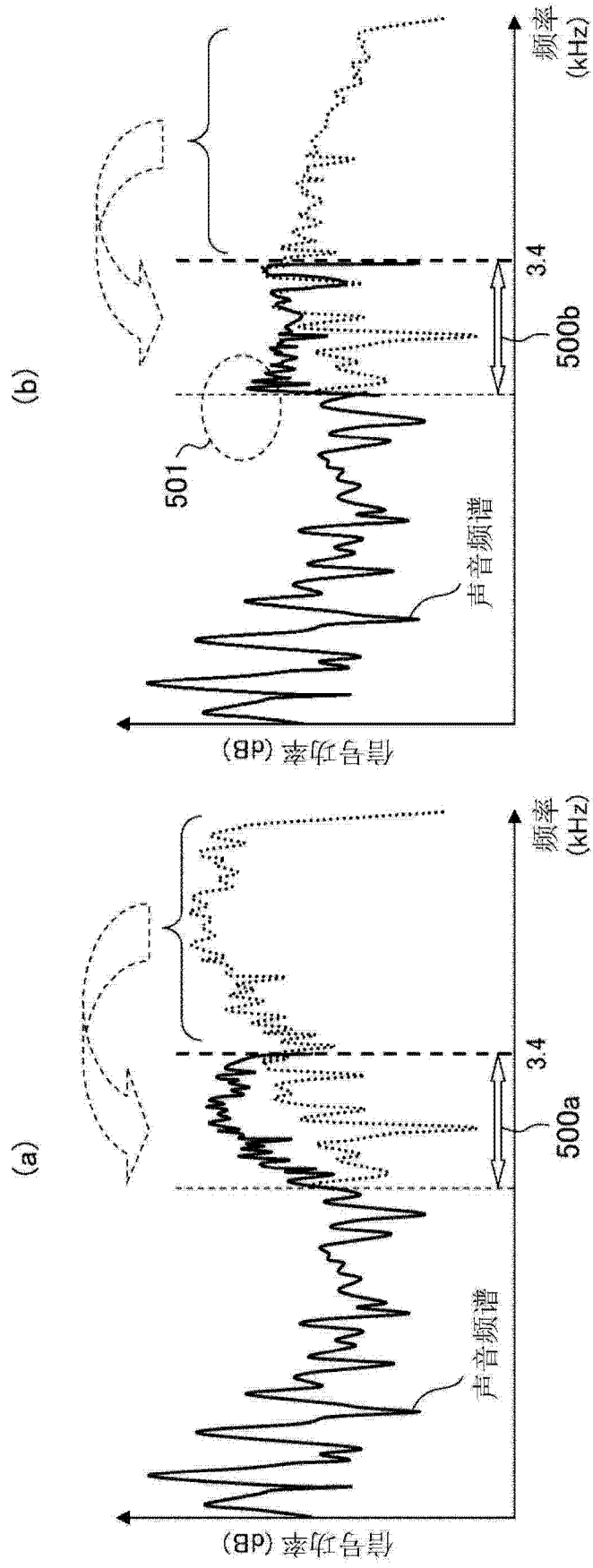


图 5

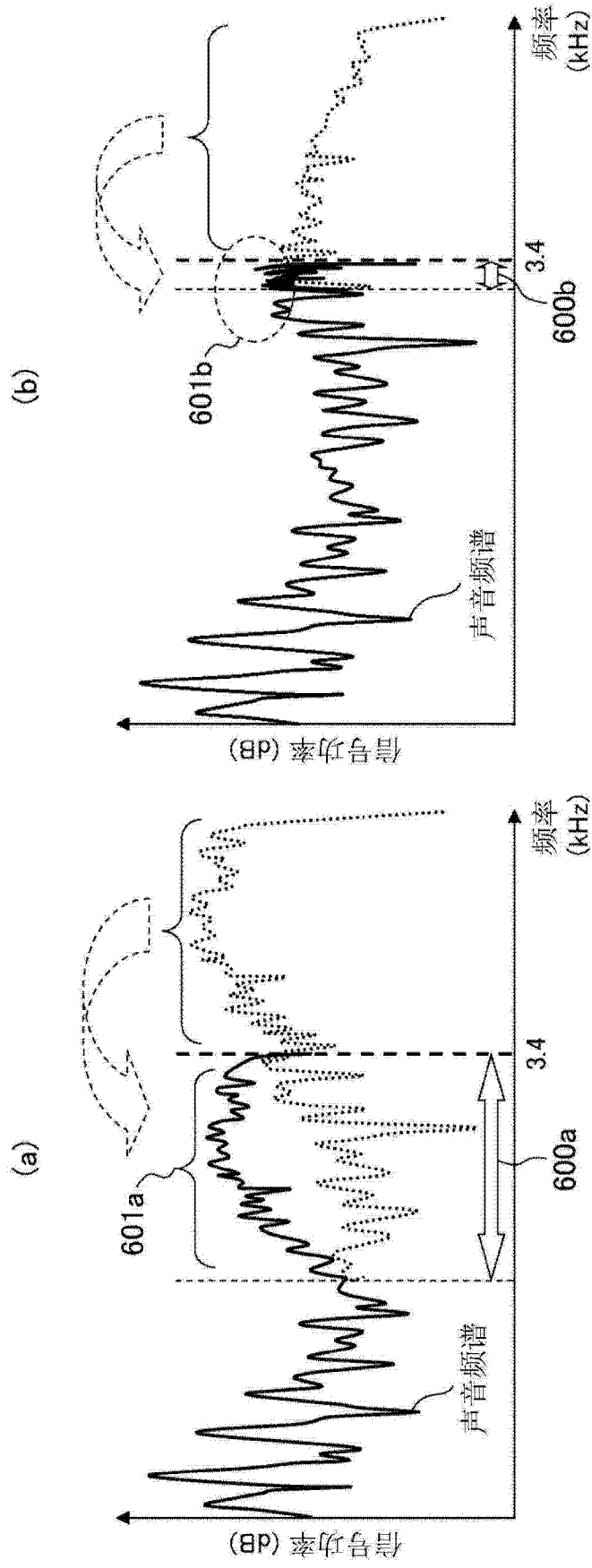


图 6