

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 2 月 8 日 (08.02.2018)



(10) 国际公布号
WO 2018/024058 A1

(51) 国际专利分类号:
G01H 7/00 (2006.01) **G10K 15/08** (2006.01)

(21) 国际申请号: PCT/CN2017/090887

(22) 国际申请日: 2017 年 6 月 29 日 (29.06.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610626191.1 2016年8月2日 (02.08.2016) CN

(71) 申请人: 电信科学技术研究院
(CHINA ACADEMY OF TELECOMMUNICATIONS
TECHNOLOGY) [CN/CN]; 中国北京市海淀区
学院路40号, Beijing 100191 (CN)。

(72) 发明人: 赵亚利(ZHAO, Yali); 中国北京市海淀区
学院路40号, Beijing 100191 (CN)。 梁民(LIANG,
Min); 中国北京市海淀区学院路40号, Beijing
100191 (CN)。 沙永涛(SHA, Yongtao); 中国北京
市海淀区学院路40号, Beijing 100191 (CN)。

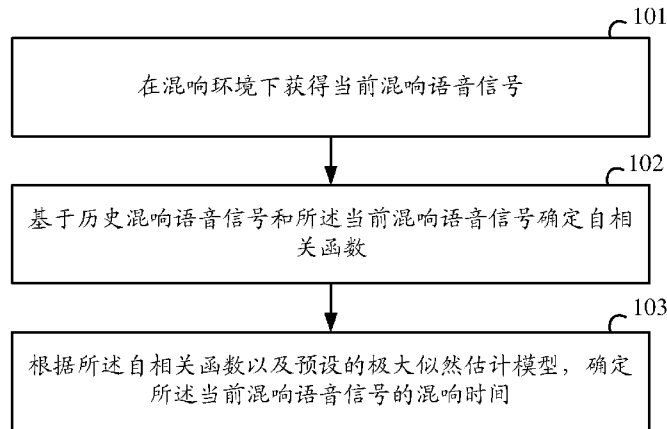
(74) 代理人: 北京同达信恒知识产权代理有限公司
(TDIP & PARTNERS); 中国北京市海淀区知春路7
号致真大厦A1304-05室, Beijing 100191 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家
保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,
BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,
CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,
JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,
LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) Title: REVERBERATION TIME ESTIMATION METHOD AND APPARATUS

(54) 发明名称: 一种混响时间估计方法及装置

图 1



101 IN REVERBERATION ENVIRONMENT, OBTAIN CURRENT REVERBERATION SPEECH SIGNAL
102 DETERMINE AUTOCORRELATION FUNCTION ON BASIS OF HISTORICAL REVERBERATION SPEECH
SIGNALS AND CURRENT REVERBERATION SPEECH SIGNAL
103 ACCORDING TO AUTOCORRELATION FUNCTION AND PRESET MAXIMUM-LIKELIHOOD ESTIMATION
MODEL, DETERMINE REVERBERATION TIME OF CURRENT REVERBERATION SPEECH SIGNAL

(57) Abstract: Provided are a reverberation time estimation method and apparatus, used for solving the problem, in existing reverberation time estimation methods based on speech models, of the reverberation time of reverberation speech for which the reverberation time is smaller than a preset threshold being overestimated. The reverberation time estimation method is: in a reverberation environment, obtaining a current reverberation speech signal (101); determining an autocorrelation function on the basis of historical reverberation speech signals and the current reverberation speech signal (102); according to the autocorrelation function and a preset maximum-likelihood estimation model, determining the reverberation time of the current reverberation speech signal (103).

(57) 摘要: 一种混响时间估计方法及装置, 用以解决现有的基于语音模型的混响时间估计方法中, 对混响时间小于预设门限的混响语音的混响时间产生过估计的问题。该混响时间估计方法为: 在混响环境下获得当前混响语音信号 (101); 基于历史混响语音信号和当前混响语音信号确定自相关函数 (102); 根据自相关函数以及预设的极大似然估计模型, 确定该当前混响语音信号的混响时间 (103)。

MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

一种混响时间估计方法及装置

本申请要求在2016年8月02日提交中国专利局、申请号为201610626191.1、发明名称为一种混响时间估计方法及装置的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及音频信号处理技术领域，尤其涉及一种混响时间估计方法及装置。

背景技术

混响时间（表示为 RT_{60} ）定义为：在特定房屋空间中从声音激励停止时起算，其残余声能经过多次反射后衰减至-60dB(相当于平均声能密度降为原来的 10^{-6})所需要的时间。混响时间是衡量特定房屋空间混响特性的一个重要指标，并且与解混响算法中后期混响（Late-Reverberation）功率的计算估计密切相关。

传统混响时间的估计方法是采用声音激励方法，主要思想为：在声音激励信号停止后测量声能衰减至低于起始观测时能量的 60dB 所经历的时间。该方法需要冲激特性的语音测试信号，以便在声音停止激励后对声能的测试不受语音拖尾成分的影响。显然，该方法不能满足实时处理应用的要求。

为此，人们探讨并提出用所接收的语音信号进行混响时间盲估计的方法，然而这种方法仅能在已检测出的语音间隙期间来应用，而且还需假设语音的结束时刻没有拖尾且语音中的间隙要足够长，因而也无法在实际中有效应用。

为改进和克服混响时间盲估计的缺陷，相关学者又提出了一种基于语音模型的更为鲁棒（Robust）的混响时间估计方法（以下简称基于语音模型的估计方法），该方法应用线性预测方法来处理所接收的语音信号，并获得该语音信号相应的残差信号，然后用该残差信号的自相关函数代入到混响时间的极大似然估计器（Maximum-Likelihood Estimator, MLE），从而获得该语音信号相应的混响时间估值。工作原理具体如下：

在混响环境下获得语音信号，假设该语音信号不存在噪声干扰，表示为公式 1：

$$x[n]=s[n]*h[n] \quad (\text{公式 1})$$

其中， $x[n]$ 表示接收语音信号， $s[n]$ 表示源语音信号， $h[n]$ 表示房屋冲激响应（Room Impulse Response, RIR），“*”表示线性卷积算子。根据统计声学理论，RIR 可以用 Polack 模型表示为如公式 2 所示的非平稳随机过程：

$$h[n]=w[n]a^n, \quad n>0 \quad (\text{公式 2})$$

其中，公式 2 中：

$$a = e^{-\delta} \quad (\text{公式 3})$$

$$\delta = \frac{3 \ln(10)}{RT_{60} f_s} \quad (\text{公式 4})$$

其中， f_s 表示采样频率， $w[n]$ 表示零均值的高斯白噪声。

由于 $s[n]$ 可以看作是激励信号 $e[n]$ 与声道滤波器 $v[n]$ 卷积的结果，因此公式 1 可以表示为公式 5 所示：

$$x[n] = e[n] * v[n] * h[n] \quad (\text{公式 5})$$

通过对 $x[n]$ 进行线性预测 (Linear Prediction, LP) 分析，获得残差信号，该残差信号可以看作是语音生成模型中激励信号的一种近似。考虑到声道滤波器可以表示成一个时变的全极点滤波器，而 RIR 通常被认为是全零点滤波器，故可以认为 $v[n]$ 与 $h[n]$ 彼此间互不相关，因此通过对 $x[n]$ 进行 LP 分析，可以近似地将声道滤波器的效应从混响语音数据中移去，从而获得如公式 6 所示的残差信号：

$$\tilde{x}[n] \approx e[n] * h[n] \quad (\text{公式 6})$$

残差信号的 N 点的自相关函数可以表示为公式 7 所示：

$$\begin{aligned} C_{\tilde{x}}[n] &= \sum_m \tilde{x}[m] \tilde{x}[m-n] = \tilde{x}[n] * \tilde{x}[-n] \\ &= (e * h)[n] * (e * h)[-n] \\ &= e[n] * e[-n] * h[n] * h[-n] \\ &= C_e[n] * C_h[n] \end{aligned} \quad (\text{公式 7})$$

其中， $C_e[n]$ 是 $e[n]$ 的自相关函数， $C_h[n]$ 是 $h[n]$ 的自相关函数。鉴于自相关函数 $C_e[n]$ 衰减速度比 $C_h[n]$ 快，可以认为：

$$C_{\tilde{x}}[n] \approx k C_h[n] \quad (\text{公式 8})$$

其中， k 表示 $C_e[n]$ 的均值。由上式可知残差信号的自相关与 RIR 的自相关有相同的统计特性，因此可以用残差信号的自相关代入 MLE 中。在工程实现中我们用时间平均代替统计平均，首先计算每一帧的自相关函数 $C_{\tilde{x}}[n]$ ，然后每隔 L 帧计算一次平均自相关函数 $\bar{C}_{\tilde{x}}[n]$ ，得到的平均自相关函数作为 MLE 估计器的输入。

计算 L 帧的自相关函数 $C_{\tilde{x}}[n]$ 的均值，表示为公式 9 所示：

$$\bar{C}_{\bar{x}}[n] = \frac{1}{L} \sum_{l=1}^L C_{l,\bar{x}}[n] \quad (\text{公式 9})$$

基于参数 a 和 k 的自相关函数 $C_{\bar{x}}[n]$ 的似然函数可表示为公式 10:

$$P(\bar{C}_{\bar{x}}; a, k) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp \left\{ -\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{C}_{\bar{x}}^2[n]}{ka^{2n}} \right\} \quad (\text{公式 10})$$

最佳的参数 a 和 k 对应于公式 10 的自然对数极大值点, 为此:

$$\frac{\partial}{\partial a} \log P\{\bar{C}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{C}_{\bar{x}}^2[n] a^{-2n-1} = 0 \quad (\text{公式 11})$$

$$\frac{\partial}{\partial k} \log P\{\bar{C}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{C}_{\bar{x}}^2[n] a^{-2n} = 0 \quad (\text{公式 12})$$

其中, $\frac{\partial}{\partial a}$ 表示对参数 a 求偏导数, $\frac{\partial}{\partial k}$ 表示对参数 k 求偏导数。应用 Newton-Raphson

方法求解公式 11 和公式 12 组成的联立方程式计算出参数 a 和 k , 然后根据公式 3 和公式 4 得到混响时间的估计值。

基于语音模型的混响时间估计方法的作者们曾建议: 在实际应用中, $\bar{C}_{\bar{x}}[n]$ 对于语音数据采用无偏自相关的计算方法, 即表示为公式 13 所示, 其中 N 为一帧语音帧包含的样点数:

$$C_{\bar{x}}[n] = \frac{1}{N-n} \sum_{m=n}^{N-1} \tilde{x}[m] \tilde{x}[m-n] \quad (\text{公式 13})$$

尽管基于语音模型的混响时间估计方法不需要检测语音的时隙, 并且利用全部的语音数据进行估计, 但实际应用发现, 该方法通常对混响时间小于预设门限的混响语音的 RT_{60} 产生过估计。

发明内容

本发明实施例提供一种混响时间估计方法及装置, 用以解决现有的基于语音模型的混响时间估计方法中, 对混响时间小于预设门限的混响语音的混响时间产生过估计的问题。

本发明实施例提供的具体技术方案如下:

第一方面, 本发明实施例提供了一种混响时间估计方法, 包括:

在混响环境下获得当前混响语音信号;

基于历史混响语音信号和所述当前混响语音信号确定自相关函数;

根据所述自相关函数以及预设的极大似然估计模型，确定所述当前混响语音信号的混响时间。

可能的实施方式中，基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前，所述方法还包括：

对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

可能的实施方式中，基于历史混响语音信号和所述当前混响语音信号确定自相关函数，包括：

从所述当前混响语音信号中提取连续的 M 个混响语音帧，基于历史混响语音信号和

所述 M 个混响语音帧确定自相关函数为： $R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n]$ ，其中，

$R_{\tilde{x}}[n]$ 表示所述自相关函数， N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， $\tilde{x}$ 表示所述混响语音帧的残差信号， n 表示移位距离，所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到， M 为大于 1 的正整数。

可能的实施方式中，根据所述自相关函数以及预设的极大似然估计模型，确定所述当前混响语音信号的混响时间，包括：

若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍，在每获得连续的 L 个自相关函数后，根据所述连续的 L 个自相关函数确定平均自相关函数，根据所述平均自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间；

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍，对本次确定的所述自相关函数进行平滑处理后，根据平滑处理后的自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间。

可能的实施方式中，根据所述连续的 L 个自相关函数确定平均自相关函数，包括：

按照公式 $\bar{R}_{\tilde{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\tilde{x},l}[n]$ 确定平均自相关函数，其中， $R_{\tilde{x},l}[n]$ 表示连续 L 次

确定的自相关函数中的第 l 个自相关该函数， $\bar{R}_{\tilde{x}}[n]$ 表示平均自相关函数。

可能的实施方式中，对本次确定的所述自相关函数进行平滑处理，包括：

按照公式 $\bar{R}_{\tilde{x},l}[n] = \beta \cdot \bar{R}_{\tilde{x},l-1}[n] + (1-\beta) R_{\tilde{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑

处理，其中， $R_{\tilde{x}}[n]$ 表示本次确定的所述自相关函数， $\bar{R}_{\tilde{x},l}[n]$ 表示第 l 次平滑处理后的自

相关函数, $\bar{R}_{\bar{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数, β 表示预设的平滑系数, $0 < \beta < 1$ 。

可能的实施方式中, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间, 包括:

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial_a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial_k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示所述平均自相关函数, $\frac{\partial}{\partial_a}$ 表示对所述第一参

数求偏导数, $\frac{\partial}{\partial_k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\bar{x}}$ 、

所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 根据平滑处理后的自相关函数以及所述预设的极大似然估计模

型，确定所述当前混响语音信号的混响时间，包括：

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件，确定第一参数和第二参数；

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件，确定所述当前混响语音信号的混响时间的估计值；

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系，确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值；

其中，所述第一约束条件表示为：

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_x\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_x^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_x\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_x^2[n] a^{-2n} = 0 \end{cases}$$

其中，N表示一个混响语音帧中参与自相关函数计算的采样点的个数，a表示所述第一参数，k表示所述第二参数， $\bar{R}_x$ 表示平滑处理后的自相关函数， $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数， $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数， $P\{\bar{R}_x\}$ 表示所述平滑处理后的自相关函数 $\bar{R}_x$ 、所述第一参数以及所述第二参数的似然函数，表示为

$$P(\bar{R}_x) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_x^2[n]}{k a^{2n}}\right\};$$

其中，所述第二约束条件为： $RT_{60} = \frac{3 \log_e(10)}{(-\log_e a) f_s}$ ， RT_{60} 表示所述当前混响语音信号的混响时间的估计值， f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中，确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后，所述方法还包括：

按照公式 $\widehat{RT}_{60} = \begin{cases} \alpha \cdot \widehat{RT}_{60}' + (1-\alpha) \cdot \widehat{RT}_{60}, & \text{如果 } \widehat{RT}_{60} > \widehat{RT}_{60}' \\ \beta \cdot \widehat{RT}_{60} + (1-\beta) \cdot \widehat{RT}_{60}', & \text{否则} \end{cases}$ 对所述混响时间

的有效值进行平滑处理, 其中, $\widehat{RT}_{60}$ 表示所述混响时间的有效值, $\overline{RT}_{60}$ 表示本次平滑处理后的混响时间, $\overline{RT}_{60}^i$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

第二方面, 本发明实施例提供了一种混响时间估计装置, 包括:

第一处理模块, 用于在混响环境下获得当前混响语音信号;

第二处理模块, 用于基于历史混响语音信号和所述当前混响语音信号确定自相关函数;

第三处理模块, 用于根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

可能的实施方式中, 所述第二处理模块还用于:

基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前, 对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

可能的实施方式中, 所述第二处理模块具体用于:

从所述当前混响语音信号中提取连续的 M 个混响语音帧, 基于历史混响语音信号和所述 M 个混响语音帧确定自相关函数为: $R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n]$, 其中, $R_{\tilde{x}}[n]$ 表示所述自相关函数, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, $\tilde{x}$ 表示所述混响语音帧的残差信号, n 表示移位距离, 所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到, M 为大于 1 的正整数。

可能的实施方式中, 所述第三处理模块具体用于:

若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍, 在每获得连续的 L 个自相关函数后, 根据所述连续的 L 个自相关函数确定平均自相关函数, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间;

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍, 对本次确定的所述自相关函数进行平滑处理后, 根据平滑处理后的自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

可能的实施方式中, 所述第三处理模块具体用于:

按照公式 $\overline{R}_{\tilde{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\tilde{x},l}[n]$ 确定平均自相关函数, 其中, $R_{\tilde{x},l}[n]$ 表示连续 L 次

确定的自相关函数中的第 l 个自相关该函数, $\bar{R}_{\tilde{x}}[n]$ 表示平均自相关函数。

可能的实施方式中, 所述第三处理模块具体用于:

按照公式 $\bar{R}_{\tilde{x},l}[n] = \beta \cdot \bar{R}_{\tilde{x},l-1}[n] + (1-\beta)R_{\tilde{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑处理, 其中, $R_{\tilde{x}}[n]$ 表示本次确定的所述自相关函数, $\bar{R}_{\tilde{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数, $\bar{R}_{\tilde{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数, β 表示预设的平滑系数, $0 < \beta < 1$ 。

可能的实施方式中, 所述第三处理模块具体用于:

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\tilde{x}}$ 表示所述平均自相关函数, $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\tilde{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\tilde{x}}$ 、

所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\tilde{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\tilde{x}}^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号的

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 所述第三处理模块具体用于:

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial_a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} \\ \frac{\partial}{\partial_k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示平滑处理后的自相关函数, $\frac{\partial}{\partial_a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial_k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平滑处理后的自相

关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 所述第三处理模块具体用于:

确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 按照公

$$\text{式 } \overline{\text{RT}}_{60} = \begin{cases} \alpha \cdot \overline{\text{RT}}'_{60} + (1-\alpha) \cdot \widehat{\text{RT}}_{60}, & \text{如果 } \widehat{\text{RT}}_{60} > \overline{\text{RT}}'_{60} \\ \beta \cdot \overline{\text{RT}}'_{60} + (1-\beta) \cdot \widehat{\text{RT}}_{60}, & \text{否则} \end{cases}$$

对所述混响时间的有效值进

行平滑处理, 其中, $\widehat{\text{RT}}_{60}$ 表示所述混响时间的有效值, $\overline{\text{RT}}_{60}$ 表示本次平滑处理后的混响时间, $\overline{\text{RT}}'_{60}$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

第三方面, 本发明实施例提供了一种设备, 该设备主要包括处理器和存储器, 其中, 存储器中保存有预设的程序, 处理器用于读取存储器中的程序, 按照该程序执行以下过程:

在混响环境下获得当前混响语音信号;

基于历史混响语音信号和所述当前混响语音信号确定自相关函数;

根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

可能的实施方式中, 处理器还用于:

在基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前, 对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

可能的实施方式中, 基于历史混响语音信号和所述当前混响语音信号确定自相关函数, 包括:

从所述当前混响语音信号中提取连续的 M 个混响语音帧, 基于历史混响语音信号和

所述 M 个混响语音帧确定自相关函数为: $R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n]$, 其中,

$R_{\tilde{x}}[n]$ 表示所述自相关函数, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, $\tilde{x}$ 表示所述混响语音帧的残差信号, n 表示移位距离, 所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到, M 为大于 1 的正整数。

可能的实施方式中, 处理器在根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间时, 具体用于:

若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍, 在每获得连续的 L 个自相关函数后, 根据所述连续的 L 个自相关函数确定平均自相关函数, 根据所述平均自

相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间；

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍，对本次确定的所述自相关函数进行平滑处理后，根据平滑处理后的自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间。

可能的实施方式中，处理器在根据所述连续的 L 个自相关函数确定平均自相关函数时，具体用于：

按照公式 $\bar{R}_{\tilde{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\tilde{x},l}[n]$ 确定平均自相关函数，其中， $R_{\tilde{x},l}[n]$ 表示连续 L 次

确定的自相关函数中的第 l 个自相关该函数， $\bar{R}_{\tilde{x}}[n]$ 表示平均自相关函数。

可能的实施方式中，处理器对本次确定的所述自相关函数进行平滑处理时，具体用于：

按照公式 $\bar{R}_{\tilde{x},l}[n] = \beta \cdot \bar{R}_{\tilde{x},l-1}[n] + (1-\beta)R_{\tilde{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑处理，其中， $R_{\tilde{x}}[n]$ 表示本次确定的所述自相关函数， $\bar{R}_{\tilde{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数， $\bar{R}_{\tilde{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数， β 表示预设的平滑系数， $0 < \beta < 1$ 。

可能的实施方式中，处理器根据所述平均自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间时，具体用于：

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件，确定第一参数和第二参数；

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件，确定所述当前混响语音信号的混响时间的估计值；

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系，确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值；

其中，所述第一约束条件表示为：

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中， N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， a 表示所述第一参数， k 表示所述第二参数， $\bar{R}_{\bar{x}}$ 表示所述平均自相关函数， $\frac{\partial}{\partial_a}$ 表示对所述第一参数求偏导数， $\frac{\partial}{\partial_k}$ 表示对所述第二参数求偏导数， $P\{\bar{R}_{\bar{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数，表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)}(2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中，所述第二约束条件为： $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$ ， RT_{60} 表示所述当前混响语音信号

的混响时间的估计值， f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中，处理器根据平滑处理后的自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间时，具体用于：

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件，确定第一参数和第二参数；

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件，确定所述当前混响语音信号的混响时间的估计值；

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系，确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值；

其中，所述第一约束条件表示为：

$$\begin{cases} \frac{\partial}{\partial_a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial_k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中， N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， a 表示所述第一参数， k 表示所述第二参数， $\bar{R}_{\bar{x}}$ 表示平滑处理后的自相关函数， $\frac{\partial}{\partial_a}$ 表示对所述第

一参数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_x\}$ 表示所述平滑处理后的自相关函数 $\bar{R}_x$ 、所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_x) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_x^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 处理器还用于:

在确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 按照公

$$\widehat{RT}_{60} = \begin{cases} \alpha \cdot \widehat{RT}_{60}' + (1-\alpha) \cdot \widehat{RT}_{60}, & \text{如果 } \widehat{RT}_{60} > \widehat{RT}_{60}' \\ \beta \cdot \widehat{RT}_{60}' + (1-\beta) \cdot \widehat{RT}_{60}, & \text{否则} \end{cases}$$

对所述混响时间的有效值进

行平滑处理, 其中, $\widehat{RT}_{60}$ 表示所述混响时间的有效值, $\widehat{RT}_{60}'$ 表示本次平滑处理后的混响时间, $\widehat{RT}_{60}'$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

基于以上技术方案, 本发明实施例中, 基于历史语音信号和当前混响语音信号确定自相关函数, 根据该自相关函数以及预设的极大似然估计模型, 确定混响语音信号的混响时间, 可以进一步提高混响时间的估计精度, 解决了现有的基于语音模型的混响时间估计方法中, 对混响时间小于预设门限的混响语音的混响时间产生过估计的问题。

附图说明

图1为本发明实施例中进行混响时间估计的方法流程示意图;

图2为本发明实施例中基于语音模型的混响时间估计过程示意图;

图3为本发明实施例中混响时间估计装置结构示意图;

图4为本发明实施例中设备结构示意图。

具体实施方式

为了使本发明实施例的目的、技术方案和优点更加清楚, 下面将结合附图对本发明实施例作进一步地详细描述, 显然, 所描述的实施例仅仅是本发明一部分实施例, 而不是全

部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例，都属于本发明保护的范围。

为了解决现有的基于语音模型的混响时间估计方法中，对混响时间小于预设门限的混响语音的混响时间产生过估计的问题，本发明实施例中对现有的基于语音模型的混响时间估计方法进行了改进。

本发明实施例中，将混响时间小于预设门限的混响语音成为小混响语音，例如，将混响时间小于 400 毫秒的混响语音称为小混响语音。

本发明实施例中，对现有的基于语音模型的混响时间估计方法进行改进后，进行混响时间估计的方法流程如图 1 所示，具体如下：

步骤 101：在混响环境下获得当前混响语音信号。

步骤 102：基于历史混响语音信号和所述当前混响语音信号确定自相关函数。

发明人发现，现有的基于语音模型的混响时间估计过程中，在公式 13 中，计算无偏自相关函数具有局部的窗口效应，为改善和提高自相关函数的计算精度，本发明实施例中提出了基于历史混响语音信号和当前混响时间信号确定自相关函数的方法。

相对公式 13 提出的无偏的自相关函数的定义，带历史数据的无偏自相关函数定义如下：

首先在帧长为 N 的数据前填充 M 长的历史数据，然后根据公式 14 求取：

$$R_{\tilde{x}}[n] = \frac{1}{N} \sum_{m=0}^{N-1} \tilde{x}[m] \tilde{x}[m-n] \quad n=0,1,\dots,M-1 \quad (\text{公式 14})$$

具体地，基于历史混响语音信号和所述当前混响语音信号确定自相关函数的过程具体如下：

从所述当前混响语音信号中提取连续的 M 个混响语音帧，基于历史混响语音信号和所述 M 个混响语音帧确定自相关函数为：

$$R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N * M - 1} \tilde{x}[m] \tilde{x}[m-n] \quad (\text{公式 15})$$

其中， $R_{\tilde{x}}[n]$ 表示所述自相关函数，N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， $\tilde{x}$ 表示所述混响语音帧的残差信号，n 表示移位距离（以样点为单位），所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到，M 为大于 1 的正整数。

较佳地，n 小于或等于 N 的二分之一。

根据实际应用中测试的结果可知, 采用 $R_{\bar{x}}[n]$ 代替 $C_{\bar{x}}[n]$ 代入到公式 11、公式 12、公式 3 和公式 4 中, 得到的 RT_{60} 比现有的基于语音模型的估计方法更为准确, 尤其是可以解决对小混响语音的过估计的问题。

步骤 103: 根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

本发明实施例中, 鉴于现有的基于语音模型的混响时间估计方法中, 需要对连续多次获得自相关函数进行平均, 这就为混响时间的估计造成较大的等待时延, 尤其是导致第一次计算得到混响时间的时间延迟太长。

例如, 假设一个混响语音帧的时长为 20 毫秒, 一个混响语音帧包含的采样点数为 $N=320$, 即混响语音信号的采样频率为 16KHz, 假设用于计算自相关函数的帧数 $M=12$, 用于计算平均自相关函数所需的自相关函数的个数 $L=20$, 则计算得到第一个混响时间估计值需要的延迟为: $20 \text{ 毫秒} \times 12 \times 20 = 4.8 \text{ 秒}$ 。以应用于去混响应用为例, 将直接影响前 4.8 秒的去混响性能。

为此, 本发明实施例中在开始混响时间估计的 4.8 秒之前, 与在开始混响时间估计的 4.8 之后, 采用不同的混响时间估计方式。具体地, 在开始混响时间估计的 4.8 秒之前采用确定混响时间的第二处理方式, 在开始混响时间估计的 4.8 之后采用确定混响时间的第一处理方式。

具体地, 根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间, 具体可以分为以下两种处理方式:

第一处理方式, 若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍, 在每获得连续的 L 个自相关函数后, 根据所述连续的 L 个自相关函数确定平均自相关函数, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

第一处理方式中, 根据所述连续的 L 个自相关函数确定平均自相关函数, 具体为: 按照公式 19 确定平均自相关函数,

$$\bar{R}_{\bar{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\bar{x},l}[n] \quad (\text{公式 16})$$

其中, $R_{\bar{x},l}[n]$ 表示连续 L 次确定的自相关函数中的第 l 个自相关该函数, $\bar{R}_{\bar{x}}[n]$ 表示平均自相关函数。

该第一处理方式中, 根据所述平均自相关函数以及所述预设的极大似然估计模型,

确定所述当前混响语音信号的混响时间，具体过程如下：根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件，确定第一参数和第二参数；根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件，确定所述当前混响语音信号的混响时间的估计值；其中，所述第一约束条件表示为：

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases} \quad (\text{公式 17})$$

其中， N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， a 表示所述第一参数， k 表示所述第二参数， $\bar{R}_{\bar{x}}$ 表示所述平均自相关函数， $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数， $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数， $P\{\bar{R}_{\bar{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数，表示为：

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\} \quad (\text{公式 18})$$

其中，所述第二约束条件为：

$$RT_{60} = \frac{3 \log_e(10)}{(-\log_e a) f_s} \quad (\text{公式 19})$$

其中， RT_{60} 表示所述当前混响语音信号的混响时间的估计值， f_s 表示确定所述混响时间采用的混响语音帧的采样率。

该第一处理方式中，每连续获得 L 个自相关函数后计算平均自相关函数，基于平均自相关函数确定混响时间的估计值，可以在不影响混响时间估计的准确性的情况下，进一步降低运算量。

第二处理方式，若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍，对本次确定的所述自相关函数进行平滑处理后，根据平滑处理后的自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间。

例如，在混响时间估计开始后的 4.8 秒之前，每 240 毫秒得到一个自相关函数后就确定一次混响时间，为了避免减少由于自相关函数的波动，采用递归平滑的方式对自相关

函数进行平滑处理。

具体地，该第二处理方式中，对本次确定的所述自相关函数进行平滑处理，具体为：按照公式 23 对所述混响语音帧的自相关函数进行平滑处理，

$$\bar{R}_{\tilde{x},l}[n] = \beta \cdot \bar{R}_{\tilde{x},l-1}[n] + (1-\beta)R_{\tilde{x}}[n] \quad (\text{公式 20})$$

其中， $R_{\tilde{x}}[n]$ 表示本次确定的所述自相关函数， $\bar{R}_{\tilde{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数， $\bar{R}_{\tilde{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数， β 表示预设的平滑系数， $0 < \beta < 1$ 。

该第二处理方式中，根据平滑处理后的自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间，具体为：根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件，确定第一参数和第二参数；根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件，确定所述当前混响语音信号的混响时间的估计值；其中，所述第一约束条件表示为：

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n} = 0 \end{cases} \quad (\text{公式 21})$$

其中， N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， a 表示所述第一参数， k 表示所述第二参数， $\bar{R}_{\tilde{x}}$ 表示平滑处理后的自相关函数， $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数， $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数， $P\{\bar{R}_{\tilde{x}}\}$ 表示所述平滑处理后的自相

关函数 $\bar{R}_{\tilde{x}}$ 、所述第一参数以及所述第二参数的似然函数，表示为：

$$P(\bar{R}_{\tilde{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\tilde{x}}^2[n]}{ka^{2n}}\right\} \quad (\text{公式 22})$$

其中，所述第二约束条件为：

$$RT_{60} = \frac{3 \log_e(10)}{(-\log_e a) f_s} \quad (\text{公式 23})$$

其中, RT_{60} 表示所述当前混响语音信号的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

优选地, 基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前, 对所述历史混响语音信号和所述当前混响语音进行降采样处理, 基于降采样后的历史混响语音信号和降采样后的当前混响语音信号确定自相关函数。

鉴于现有的基于语音模型的混响时间估计方法中, 假设一帧语音帧包含的采样点数为 N , 计算自相关的帧数为 M , 且 $N \times M$ 需要足够大才能够使得语音模型成立, 计算自相关的运算量很大。例如, 假设 $N \times M = 3840$, 即对应 240 毫秒内以 16KHz 的采样速率进行采样获得数据个数, 如果只计算 $N \times M/2 = 1920$ 个自相关值, 则计算复杂度为: 乘法次数为 $N \times M \times N \times M/2 = 7372800$, 加法次数为 $(N \times M - 1) \times N \times M/2 = 7370800$ 。

通过降采样可以有效降低计算量, 具体地, 假设输入的语音信号的采样频率为 16KHz, 假设将采样频率降为 4KHz, 则自相关函数的计算运算量将变为原来的 1/16, 即降采样后的计算复杂度为: 乘法次数为 $7372800/16 = 460800$, 加法次数为 $7370800/16 = 460680$ 。同时极大似然估计过程的计算复杂度也会降为原来的 1/16。

优选地, 如果采用降采样后的历史语音信号和降采样后的当前混响语音信号计算自相关函数以及计算混响语音信号的混响时间的估计值, 则在采用第一或第二处理方式确定当前混响语音信号的混响时间的估计值之后, 根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值。

具体地, 采用降采样后的历史语音信号和降采样后的当前混响语音信号计算得到的自相关函数后, 将该自相关函数代入极大似然估计模型中得到的为混响时间的估计值。采用未进行降采样处理的历史语音信号和当前混响语音信号计算得到的自相关函数, 将该自相关函数代入极大似然估计模型中得到的为混响时间的有效值。混响时间的估计值与混响时间的有效值之间存在映射关系, 该映射关系可采用数学映射函数 $\Gamma(\cdot)$ 表示为:

$$\widehat{RT}_{60} = \Gamma\left(\widehat{RT}_{60} \Big|_{\text{降采样}}\right) \quad (\text{公式 24})$$

其中, $\widehat{RT}_{60} \Big|_{\text{降采样}}$ 表示混响时间的估计值, $\widehat{RT}_{60}$ 表示混响时间的有效值。

优选地, 对于第二处理方式, 为了减小估计出的混响时间的有效值的波动性, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 按照公式 25 对所

述混响时间的有效值进行平滑处理,

$$\overline{RT}_{60} = \begin{cases} \alpha \cdot \overline{RT}_{60}' + (1-\alpha) \cdot \widehat{RT}_{60}, & \text{如果 } \widehat{RT}_{60} > \overline{RT}_{60}' \\ \beta \cdot \overline{RT}_{60}' + (1-\beta) \cdot \widehat{RT}_{60}, & \text{否则} \end{cases} \quad \tilde{x} \quad (\text{公式 25})$$

其中, $\widehat{RT}_{60}$ 表示所述混响时间的有效值, $\overline{RT}_{60}$ 表示本次平滑处理后的混响时间, $\overline{RT}_{60}'$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

该第二处理方式中对混响时间的有效值进行平滑处理的基本思想是以“快上升慢下降”为准则的, 即当瞬时的混响时间的有效值大于上一次平滑处理后的混响时间时, 使用较小的平滑因子 α , 否则, 使用较大的平滑因子 β 。

具体应用中, 在混响时间估计开始后的第一个 4.8 秒之后, 按照第一处理方式, 每 4.8 秒进行一次 MLE 来更新混响时间的估计值, 从而进一步降低运算量, 因为同一空间内混响时间是相对固定的。

以下通过一个具体实施例对本发明实施例所提供的基于语音模型的混响时间估计方法进行完整说明。

如图 2 所示为该具体实施例中基于语音模型的混响时间估计过程示意图, 具体过程如下:

步骤 201: 在初始化过程中配置以下参数: 一个语音帧中包含的样点数为 N; 配置 M、L、平滑因子 α 和 β , 设置帧计数器 Cnt=0, 设置计数器 m=0, 以及设置计数器 h=0。

步骤 202: 读取一个混响语音帧, 更新 Cnt=Cnt+1 以及更新 m=m+1。

步骤 203: 对混响语音帧进行降采样处理。

步骤 204: 对将采样处理后的混响语音帧进行 LPC 分析以及 LP 滤波。

步骤 205: 判断是否满足 m>M, 若是, 执行步骤 206, 否则, 执行步骤 202。

步骤 206: 按照公式 15 计算 M 个混响语音帧的残差信号的自相关函数, 并更新 m=0, 以及更新 h=h+1。

步骤 207: 判断是否满足 Cnt>M×L, 若不满足, 执行步骤 208, 否则执行步骤 209;

步骤 208: 按照公式 20 对自相关函数进行平滑处理后, 执行步骤 211。

步骤 209: 按照公式 16 计算连续 L 个自相关函数的平均自相关函数后, 执行步骤 210。

步骤 210: 判断是否满足 h>L, 若不满足, 转去执行步骤 202, 否则, 执行步骤 211。

步骤 211: 计算混响时间, 具体为: 根据平滑处理后的自相关函数或者平均自相关函

数, 采用 Newton-Raphson 方法求解公式 21, 得到参数 a 和 k, 根据公式 23 得到混响时间的估计值, 根据公式 24 得到该混响时间的估计值对应的混响时间的有效值, 更新 $h=0$ 。

步骤 212: 判断是否仍在接收混响语音信号, 若是, 转去执行步骤 202, 否则, 结束。

基于同一发明构思, 本发明实施例中提供了一种混响时间估计装置, 该装置的具体实施可参见方法实施例部分的描述, 重复之处不再赘述, 如图 3 所示, 该装置主要包括:

第一处理模块 301, 用于在混响环境下获得当前混响语音信号;

第二处理模块 302, 用于基于历史混响语音信号和所述当前混响语音信号确定自相关函数;

第三处理模块 303, 用于根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

可能的实施方式中, 所述第二处理模块还用于:

基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前, 对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

可能的实施方式中, 所述第二处理模块具体用于:

从所述当前混响语音信号中提取连续的 M 个混响语音帧, 基于历史混响语音信号和所述 M 个混响语音帧确定自相关函数为: $R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n]$, 其中,

$R_{\tilde{x}}[n]$ 表示所述自相关函数, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, $\tilde{x}$ 表示所述混响语音帧的残差信号, n 表示移位距离, 所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到, M 为大于 1 的正整数。

可能的实施方式中, 所述第三处理模块具体用于:

若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍, 在每获得连续的 L 个自相关函数后, 根据所述连续的 L 个自相关函数确定平均自相关函数, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间;

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍, 对本次确定的所述自相关函数进行平滑处理后, 根据平滑处理后的自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

可能的实施方式中, 所述第三处理模块具体用于:

按照公式 $\bar{R}_{\tilde{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\tilde{x},l}[n]$ 确定平均自相关函数, 其中, $R_{\tilde{x},l}[n]$ 表示连续 L 次

确定的自相关函数中的第 l 个自相关该函数, $\bar{R}_{\tilde{x}}[n]$ 表示平均自相关函数。

可能的实施方式中, 所述第三处理模块具体用于:

按照公式 $\bar{R}_{\tilde{x},l}[n] = \beta \cdot \bar{R}_{\tilde{x},l-1}[n] + (1-\beta)R_{\tilde{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑处理, 其中, $R_{\tilde{x}}[n]$ 表示本次确定的所述自相关函数, $\bar{R}_{\tilde{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数, $\bar{R}_{\tilde{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数, β 表示预设的平滑系数, $0 < \beta < 1$ 。

可能的实施方式中, 所述第三处理模块具体用于:

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\tilde{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\tilde{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\tilde{x}}$ 表示所述平均自相关函数, $\frac{\partial}{\partial a}$ 表示对所述第一参

数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\tilde{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\tilde{x}}$ 、

所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\tilde{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\tilde{x}}^2[n]}{k a^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号的

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 所述第三处理模块具体用于:

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial_a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} \\ \frac{\partial}{\partial_k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示平滑处理后的自相关函数, $\frac{\partial}{\partial_a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial_k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平滑处理后的自相

关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 所述第三处理模块具体用于:

确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 按照公

$$\text{式 } \overline{\widehat{\text{RT}}}_{60} = \begin{cases} \alpha \cdot \overline{\widehat{\text{RT}}}_{60}' + (1-\alpha) \cdot \widehat{\text{RT}}_{60}, & \text{如果 } \widehat{\text{RT}}_{60} > \overline{\widehat{\text{RT}}}_{60}' \\ \beta \cdot \overline{\widehat{\text{RT}}}_{60}' + (1-\beta) \cdot \widehat{\text{RT}}_{60}, & \text{否则} \end{cases} \quad \text{对所述混响时间的有效值进}$$

行平滑处理, 其中, $\widehat{\text{RT}}_{60}$ 表示所述混响时间的有效值, $\overline{\widehat{\text{RT}}}_{60}$ 表示本次平滑处理后的混响时间, $\overline{\widehat{\text{RT}}}_{60}'$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

基于同一发明构思, 本发明实施例还提供了一种设备, 该设备的具体实施可参见方法实施例的相关描述, 如图 4 所示, 该设备主要包括处理器 401 和存储器 402, 其中, 存储器 402 中保存有预设的程序, 处理器 401 用于读取存储器 402 中的程序, 按照该程序执行以下过程:

在混响环境下获得当前混响语音信号;

基于历史混响语音信号和所述当前混响语音信号确定自相关函数;

根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

可能的实施方式中, 处理器 401 还用于: 基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前, 对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

可能的实施方式中, 处理器 401 基于历史混响语音信号和所述当前混响语音信号确定自相关函数时, 具体用于: 从所述当前混响语音信号中提取连续的 M 个混响语音帧, 基于历史混响语音信号和所述 M 个混响语音帧确定自相关函数为:

$$R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n], \quad \text{其中, } R_{\tilde{x}}[n] \text{ 表示所述自相关函数, } N \text{ 表示一个混响语音帧中参与自相关函数计算的采样点的个数, } \tilde{x} \text{ 表示所述混响语音帧的残差信号, } n \text{ 表示移位距离, 所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到, } M \text{ 为大于 1 的正整数。}$$

可能的实施方式中, 处理器 401 根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间时, 具体用于: 若确定混响时间估计过程中提取混

响语音帧的总数超过 M 的 L 倍, 在每获得连续的 L 个自相关函数后, 根据所述连续的 L 个自相关函数确定平均自相关函数, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间;

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍, 对本次确定的所述自相关函数进行平滑处理后, 根据平滑处理后的自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

可能的实施方式中, 处理器 401 根据所述连续的 L 个自相关函数确定平均自相关函数时, 具体用于: 按照公式 $\bar{R}_{\bar{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\bar{x},l}[n]$ 确定平均自相关函数, 其中, $R_{\bar{x},l}[n]$ 表示连续 L 次确定的自相关函数中的第 l 个自相关该函数, $\bar{R}_{\bar{x}}[n]$ 表示平均自相关函数。

可能的实施方式中, 处理器 401 对本次确定的所述自相关函数进行平滑处理时, 具体用于:

按照公式 $\bar{R}_{\bar{x},l}[n] = \beta \cdot \bar{R}_{\bar{x},l-1}[n] + (1-\beta)R_{\bar{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑处理, 其中, $R_{\bar{x}}[n]$ 表示本次确定的所述自相关函数, $\bar{R}_{\bar{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数, $\bar{R}_{\bar{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数, β 表示预设的平滑系数, $0 < \beta < 1$ 。

可能的实施方式中, 处理器 401 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间时, 具体用于:

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示所述平均自相关函数, $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)}} (2\pi k)^N} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{k a^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 处理器 401 根据平滑处理后的自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间时, 具体用于:

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_x$ 表示平滑处理后的自相关函数, $\frac{\partial}{\partial_a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial_k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_x\}$ 表示所述平滑处理后的自相关函数 $\bar{R}_x$ 、所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_x) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_x^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号的

混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

可能的实施方式中, 处理器 401 还用于:

在确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 按照

$$\text{公式 } \overline{\overline{RT}}_{60} = \begin{cases} \alpha \cdot \overline{\overline{RT}}'_{60} + (1-\alpha) \cdot \widehat{RT}_{60}, & \text{如果 } \widehat{RT}_{60} > \overline{\overline{RT}}'_{60} \\ \beta \cdot \overline{\overline{RT}}'_{60} + (1-\beta) \cdot \widehat{RT}_{60}, & \text{否则} \end{cases} \quad \text{对所述混响时间的有效值}$$

进行平滑处理, 其中, $\widehat{RT}_{60}$ 表示所述混响时间的有效值, $\overline{\overline{RT}}_{60}$ 表示本次平滑处理后的混响时间, $\overline{\overline{RT}}'_{60}$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

其中, 处理器和存储器之间通过总线连接, 总线架构可以包括任意数量的互联的总线和桥, 具体由处理器代表的一个或多个处理器和存储器代表的存储器的各种电路链接在一起。总线架构还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起, 这些都是本领域所公知的, 因此, 本文不再对其进行进一步描述。总线接口提供接口。处理器负责管理总线架构和通常的处理, 存储器可以存储处理器在执行操作时所使用的数据。

基于以上技术方案, 本发明实施例中, 基于历史语音信号和当前混响语音信号确定自相关函数, 根据该自相关函数以及预设的极大似然估计模型, 确定混响语音信号的混响时间, 可以进一步提高混响时间的估计精度, 解决了现有的基于语音模型的混响时间估计方

法中，对混响时间小于预设门限的混响语音的混响时间产生过估计的问题。

并且，本发明实施例中，在对历史语音信号以及当前混响语音信号进行降采样处理，基于降采样后的历史语音信号以及降采样后的当前混响语音信号计算自相关函数，可以进一步降低计算复杂度，提高混响时间估计的实时性。

另外，本发明实施例中，在混响时间估计开始的设定时长内，在每次计算得到自相关函数后，根据该自相关函数确定混响时间，从而可以解决现有的基于语音模型的混响时间估计过程中，首次获得混响时间的等待时延长的的问题，缩短了工程实现中首次获得混响时间的等待时长，使得更适合实际应用。

本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器和光学存储器等）上实施的计算机程序产品的形式。

本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

权 利 要 求

1、一种混响时间估计方法，其特征在于，包括：

在混响环境下获得当前混响语音信号；

基于历史混响语音信号和所述当前混响语音信号确定自相关函数；

根据所述自相关函数以及预设的极大似然估计模型，确定所述当前混响语音信号的混响时间。

2、如权利要求 1 所述的方法，其特征在于，基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前，所述方法还包括：

对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

3、如权利要求 2 所述的方法，其特征在于，基于历史混响语音信号和所述当前混响语音信号确定自相关函数，包括：

从所述当前混响语音信号中提取连续的 M 个混响语音帧，基于历史混响语音信号和

所述 M 个混响语音帧确定自相关函数为： $R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n]$ ，其中，

$R_{\tilde{x}}[n]$ 表示所述自相关函数， N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， $\tilde{x}$ 表示所述混响语音帧的残差信号， n 表示移位距离，所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到， M 为大于 1 的正整数。

4、如权利要求 2 所述的方法，其特征在于，根据所述自相关函数以及预设的极大似然估计模型，确定所述当前混响语音信号的混响时间，包括：

若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍，在每获得连续的 L 个自相关函数后，根据所述连续的 L 个自相关函数确定平均自相关函数，根据所述平均自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间；

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍，对本次确定的所述自相关函数进行平滑处理后，根据平滑处理后的自相关函数以及所述预设的极大似然估计模型，确定所述当前混响语音信号的混响时间。

5、如权利要求 4 所述的方法，其特征在于，根据所述连续的 L 个自相关函数确定平均自相关函数，包括：

按照公式 $\overline{R_{\tilde{x}}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\tilde{x},l}[n]$ 确定平均自相关函数，其中， $R_{\tilde{x},l}[n]$ 表示连续 L 次

确定的自相关函数中的第 l 个自相关该函数, $\bar{R}_{\bar{x}}[n]$ 表示平均自相关函数。

6、如权利要求 4 所述的方法, 其特征在于, 对本次确定的所述自相关函数进行平滑处理, 包括:

按照公式 $\bar{R}_{\bar{x},l}[n] = \beta \cdot \bar{R}_{\bar{x},l-1}[n] + (1-\beta)R_{\bar{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑处理, 其中, $R_{\bar{x}}[n]$ 表示本次确定的所述自相关函数, $\bar{R}_{\bar{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数, $\bar{R}_{\bar{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数, β 表示预设的平滑系数, $0 < \beta < 1$ 。

7、如权利要求 5 所述的方法, 其特征在于, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间, 包括:

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示所述平均自相关函数, $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\bar{x}}$ 、

所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)}} (2\pi k)^N} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{k a^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号的

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

8、如权利要求 6 所述的方法, 其特征在于, 根据平滑处理后的自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间, 包括:

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示平滑处理后的自相关函数, $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平滑处理后的自相

关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数, 表示为

表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

9、如权利要求 8 所述的方法, 其特征在于, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 所述方法还包括:

$$\text{按照公式 } \overline{\widehat{RT}}_{60} = \begin{cases} \alpha \cdot \overline{\widehat{RT}}'_{60} + (1-\alpha) \cdot \widehat{RT}_{60}, & \text{如果 } \widehat{RT}_{60} > \overline{\widehat{RT}}'_{60} \\ \beta \cdot \overline{\widehat{RT}}'_{60} + (1-\beta) \cdot \widehat{RT}_{60}, & \text{否则} \end{cases} \quad \text{对所述混响时间}$$

的有效值进行平滑处理, 其中, $\widehat{RT}_{60}$ 表示所述混响时间的有效值, $\overline{\widehat{RT}}_{60}$ 表示本次平滑处理后的混响时间, $\overline{\widehat{RT}}'_{60}$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

10、一种混响时间估计装置, 其特征在于, 包括:

第一处理模块, 用于在混响环境下获得当前混响语音信号;

第二处理模块, 用于基于历史混响语音信号和所述当前混响语音信号确定自相关函数;

第三处理模块, 用于根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

11、如权利要求 10 所述的装置, 其特征在于, 所述第二处理模块还用于:

基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前, 对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

12、如权利要求 11 所述的装置, 其特征在于, 所述第二处理模块具体用于:

从所述当前混响语音信号中提取连续的 M 个混响语音帧, 基于历史混响语音信号和所述 M 个混响语音帧确定自相关函数为: $R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n]$, 其中, $R_{\tilde{x}}[n]$ 表示所述自相关函数, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, $\tilde{x}$ 表示所述混响语音帧的残差信号, n 表示移位距离, 所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语音帧进行线性预测滤波后得到, M 为大于 1 的正整数。

13、如权利要求 11 所述的装置, 其特征在于, 所述第三处理模块具体用于:

若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍, 在每获得连续的 L 个自相关函数后, 根据所述连续的 L 个自相关函数确定平均自相关函数, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间;

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍,对本次确定的所述自相关函数进行平滑处理后,根据平滑处理后的自相关函数以及所述预设的极大似然估计模型,确定所述当前混响语音信号的混响时间。

14、如权利要求 13 所述的装置,其特征在于,所述第三处理模块具体用于:

按照公式 $\bar{R}_{\bar{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\bar{x},l}[n]$ 确定平均自相关函数,其中, $R_{\bar{x},l}[n]$ 表示连续 L 次

确定的自相关函数中的第 l 个自相关该函数, $\bar{R}_{\bar{x}}[n]$ 表示平均自相关函数。

15、如权利要求 13 所述的装置,其特征在于,所述第三处理模块具体用于:

按照公式 $\bar{R}_{\bar{x},l}[n] = \beta \cdot \bar{R}_{\bar{x},l-1}[n] + (1-\beta) R_{\bar{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑处理,其中, $R_{\bar{x}}[n]$ 表示本次确定的所述自相关函数, $\bar{R}_{\bar{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数, $\bar{R}_{\bar{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数, β 表示预设的平滑系数, $0 < \beta < 1$ 。

16、如权利要求 14 所述的装置,其特征在于,所述第三处理模块具体用于:

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件,确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件,确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系,确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中,所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示所述平均自相关函数, $\frac{\partial}{\partial a}$ 表示对所述第一参

数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平均自相关函数 $\bar{R}_{\bar{x}}$ 、

所述第一参数以及所述第二参数的似然函数，表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中，所述第二约束条件为： $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$ ， RT_{60} 表示所述当前混响语音信号

的混响时间的估计值， f_s 表示确定所述混响时间采用的混响语音帧的采样率。

17、如权利要求 15 所述的装置，其特征在于，所述第三处理模块具体用于：

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件，确定第一参数和第二参数；

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件，确定所述当前混响语音信号的混响时间的估计值；

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系，确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值；

其中，所述第一约束条件表示为：

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中， N 表示一个混响语音帧中参与自相关函数计算的采样点的个数， a 表示所述第一参数， k 表示所述第二参数， $\bar{R}_{\bar{x}}$ 表示平滑处理后的自相关函数， $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数， $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数， $P\{\bar{R}_{\bar{x}}\}$ 表示所述平滑处理后的自相

关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数，表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号的

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

18、如权利要求 17 所述的装置, 其特征在于, 所述第三处理模块具体用于:

确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 按照公

$$\text{式 } \overline{\overline{RT}}_{60} = \begin{cases} \alpha \cdot \overline{RT}_{60}' + (1-\alpha) \cdot \widehat{RT}_{60}, & \text{如果 } \widehat{RT}_{60} > \overline{RT}_{60}' \\ \beta \cdot \overline{RT}_{60}' + (1-\beta) \cdot \widehat{RT}_{60}, & \text{否则} \end{cases}$$

对所述混响时间的有效值进

行平滑处理, 其中, $\widehat{RT}_{60}$ 表示所述混响时间的有效值, $\overline{\overline{RT}}_{60}$ 表示本次平滑处理后的混响时间, $\overline{RT}_{60}'$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

19、一种设备, 其特征在于, 该设备包括处理器和存储器, 其中, 所述存储器用于存储计算机可读程序, 所述处理器用于读取所述存储器中的程序, 按照该程序执行以下过程:

在混响环境下获得当前混响语音信号;

基于历史混响语音信号和所述当前混响语音信号确定自相关函数;

根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

20、如权利要求 19 所述的设备, 其特征在于, 所述处理器还用于:

在基于历史混响语音信号和所述当前混响语音信号确定自相关函数之前, 对所述历史混响语音信号和所述当前混响语音信号进行降采样处理。

21、如权利要求 20 所述的设备, 其特征在于, 所述处理器基于历史混响语音信号和所述当前混响语音信号确定自相关函数时, 具体用于:

从所述当前混响语音信号中提取连续的 M 个混响语音帧, 基于历史混响语音信号和所

$$\text{述 } M \text{ 个混响语音帧确定自相关函数为: } R_{\tilde{x}}[n] = \frac{1}{N * M} \sum_{m=0}^{N*M-1} \tilde{x}[m] \tilde{x}[m-n], \text{ 其中, } R_{\tilde{x}}[n]$$

表示所述自相关函数, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, $\tilde{x}$ 表示所述混响语音帧的残差信号, n 表示移位距离, 所述混响语音帧对应的残差信号为对所述混响语音帧进行线性预测得到线性预测系数并基于所述线性预测系数对所述混响语

音帧进行线性预测滤波后得到, M 为大于 1 的正整数。

22、如权利要求 20 所述的设备, 其特征在于, 所述处理器根据所述自相关函数以及预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间时, 具体用于:

若确定混响时间估计过程中提取混响语音帧的总数超过 M 的 L 倍, 在每获得连续的 L 个自相关函数后, 根据所述连续的 L 个自相关函数确定平均自相关函数, 根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间;

若确定混响时间估计过程中提取混响语音帧的总数未超过 M 的 L 倍, 对本次确定的所述自相关函数进行平滑处理后, 根据平滑处理后的自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间。

23、如权利要求 22 所述的设备, 其特征在于, 所述处理器根据所述连续的 L 个自相关函数确定平均自相关函数时, 具体用于:

按照公式 $\bar{R}_{\bar{x}}[n] = \frac{1}{L} \sum_{l=0}^{L-1} R_{\bar{x},l}[n]$ 确定平均自相关函数, 其中, $R_{\bar{x},l}[n]$ 表示连续 L 次

确定的自相关函数中的第 l 个自相关该函数, $\bar{R}_{\bar{x}}[n]$ 表示平均自相关函数。

24、如权利要求 22 所述的设备, 其特征在于, 所述处理器对本次确定的所述自相关函数进行平滑处理时, 具体用于:

按照公式 $\bar{R}_{\bar{x},l}[n] = \beta \cdot \bar{R}_{\bar{x},l-1}[n] + (1-\beta)R_{\bar{x}}[n]$ 对所述混响语音帧的自相关函数进行平滑处理, 其中, $R_{\bar{x}}[n]$ 表示本次确定的所述自相关函数, $\bar{R}_{\bar{x},l}[n]$ 表示第 l 次平滑处理后的自相关函数, $\bar{R}_{\bar{x},l-1}[n]$ 表示第 $l-1$ 次平滑处理后的自相关函数, β 表示预设的平滑系数, $0 < \beta < 1$ 。

25、如权利要求 23 所述的设备, 其特征在于, 所述处理器根据所述平均自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间时, 具体用于:

根据所述平均自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial a} \log P\{\bar{R}_x\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_x^2[n] a^{-2n-1} = 0 \\ \frac{\partial}{\partial k} \log P\{\bar{R}_x\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_x^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_x$ 表示所述平均自相关函数, $\frac{\partial}{\partial a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_x\}$ 表示所述平均自相关函数 $\bar{R}_x$ 、所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_x) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_x^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3\log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号

的混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

26、如权利要求 24 所述的设备, 其特征在于, 所述处理器根据平滑处理后的自相关函数以及所述预设的极大似然估计模型, 确定所述当前混响语音信号的混响时间时, 具体用于:

根据所述平滑处理后的自相关函数以及所述预设的极大似然估计模型中定义的第一约束条件, 确定第一参数和第二参数;

根据所述第一参数以及所述预设的极大似然估计模型中定义的第二约束条件, 确定所述当前混响语音信号的混响时间的估计值;

根据预设的采用降采样后的混响语音信号帧确定的混响时间的估计值与混响时间的有效值之间的映射关系, 确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值;

其中, 所述第一约束条件表示为:

$$\begin{cases} \frac{\partial}{\partial_a} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N(N-1)}{2a} + \frac{1}{k} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n-1} \\ \frac{\partial}{\partial_k} \log P\{\bar{R}_{\bar{x}}\} = -\frac{N}{2k} + \frac{1}{2k^2} \sum_{n=0}^{N-1} \bar{R}_{\bar{x}}^2[n] a^{-2n} = 0 \end{cases}$$

其中, N 表示一个混响语音帧中参与自相关函数计算的采样点的个数, a 表示所述第一参数, k 表示所述第二参数, $\bar{R}_{\bar{x}}$ 表示平滑处理后的自相关函数, $\frac{\partial}{\partial_a}$ 表示对所述第一参数求偏导数, $\frac{\partial}{\partial_k}$ 表示对所述第二参数求偏导数, $P\{\bar{R}_{\bar{x}}\}$ 表示所述平滑处理后的自相关函数 $\bar{R}_{\bar{x}}$ 、所述第一参数以及所述第二参数的似然函数, 表示为

$$P(\bar{R}_{\bar{x}}) = \frac{1}{\sqrt{a^{N(N-1)} (2\pi k)^N}} \exp\left\{-\frac{1}{2} \sum_{n=0}^{N-1} \frac{\bar{R}_{\bar{x}}^2[n]}{ka^{2n}}\right\};$$

其中, 所述第二约束条件为: $RT_{60} = \frac{3 \log_e(10)}{(-\log_e a) f_s}$, RT_{60} 表示所述当前混响语音信号的

混响时间的估计值, f_s 表示确定所述混响时间采用的混响语音帧的采样率。

27、如权利要求 26 所述的设备, 其特征在于, 所述处理器还用于:

在确定所述混响语音信号的混响时间的估计值对应的混响时间的有效值之后, 按照公

$$\text{式 } \overline{\overline{RT}}_{60} = \begin{cases} \alpha \cdot \overline{\overline{RT}}'_{60} + (1-\alpha) \cdot \widehat{RT}_{60}, & \text{如果 } \widehat{RT}_{60} > \overline{\overline{RT}}'_{60} \\ \beta \cdot \overline{\overline{RT}}'_{60} + (1-\beta) \cdot \widehat{RT}_{60}, & \text{否则} \end{cases} \quad \text{对所述混响时间的有效值进}$$

行平滑处理, 其中, $\widehat{RT}_{60}$ 表示所述混响时间的有效值, $\overline{\overline{RT}}_{60}$ 表示本次平滑处理后的混响时间, $\overline{\overline{RT}}'_{60}$ 表示上一次平滑处理后的混响时间, α 表示第一平滑因子, β 表示第二平滑因子, α 小于 β , α 大于零且小于 1, β 大于零且小于 1。

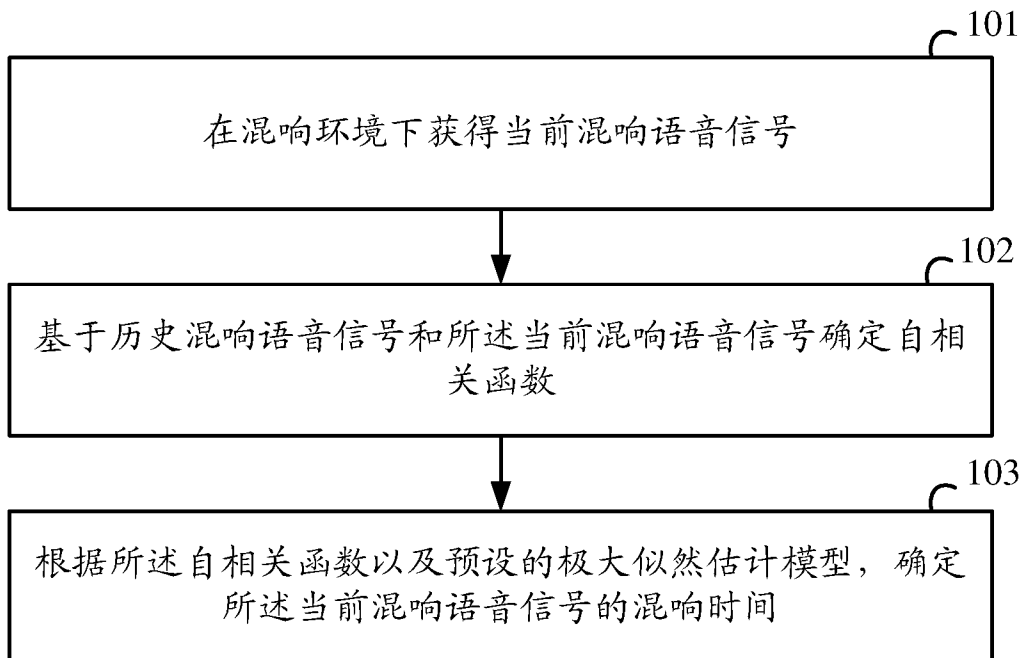


图 1

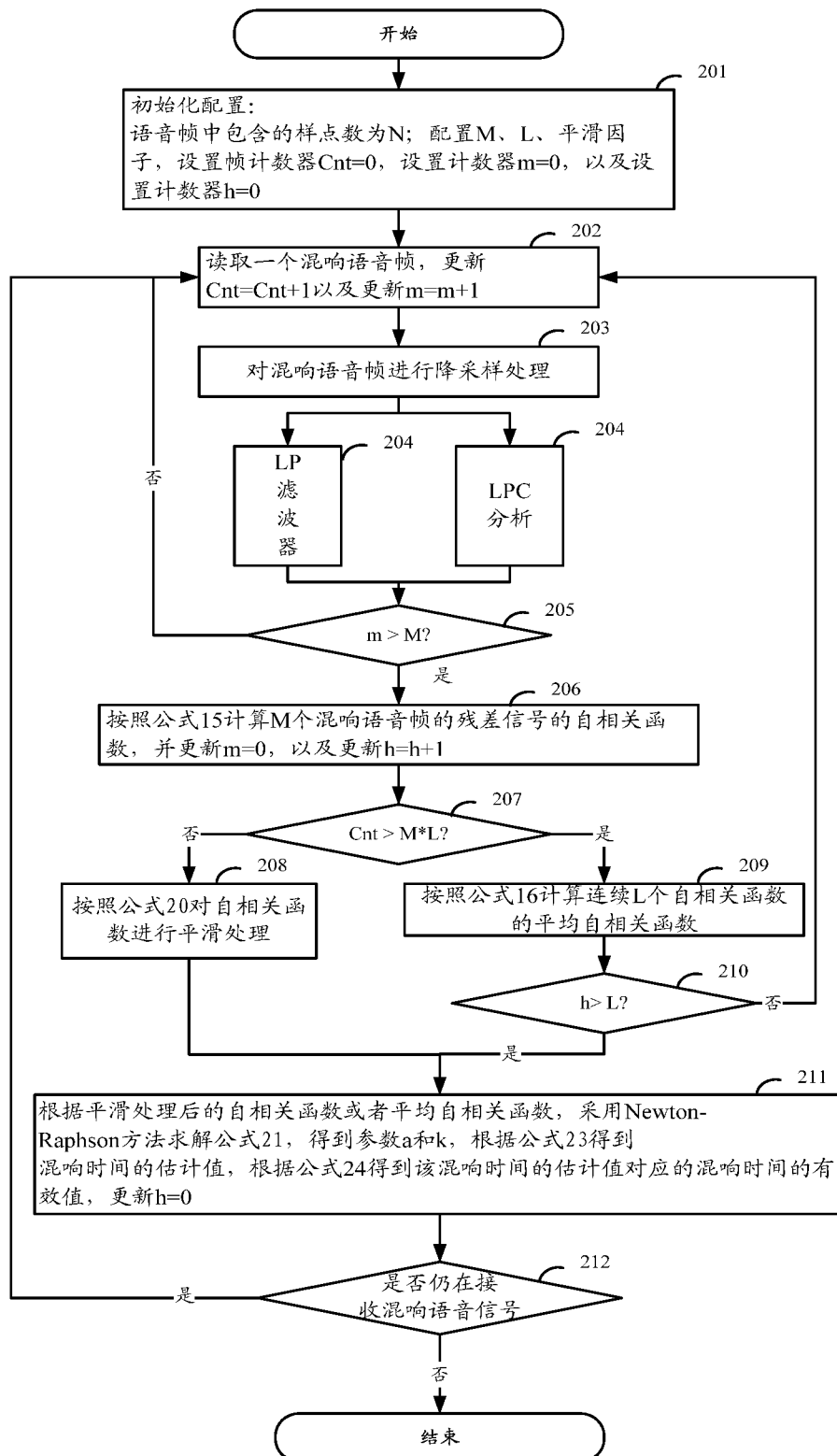


图 2

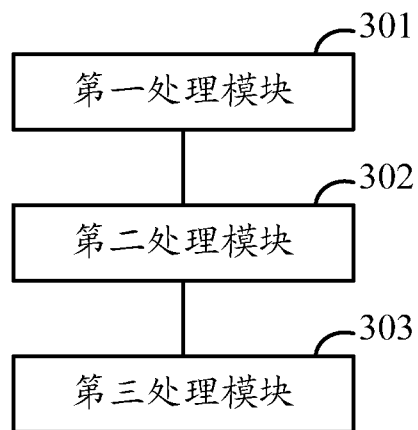


图 3

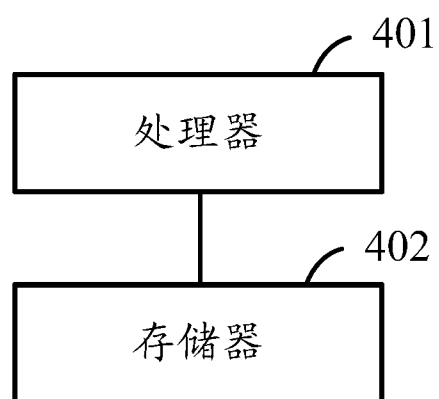


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/090887

A. CLASSIFICATION OF SUBJECT MATTER

G01H 7/00 (2006.01) i; G10K 15/08 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10; G01H 7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKI, BAIDU SCHOLAR, VEN: residual error, reverberation, time, histor+, past, forgone, old, residual, autocorrelat+, maximum likelihood

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KESHAVARZ, A. et al., "Speech-Model Based Accurate Blind Reverberation Time Estimation Using an LPC Filter[J]", IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING, vol. 20, no. 6, 30 August 2012 (30.08.2012), pages 1884-1893	1-27
A	JAYASHREE, R. et al., "Accurate Estimation of Reverberation Time and Drr Using Maximum Likelihood Estimator", INTERNATIONAL JOURNAL OF EMERGING TECHNOLOGY AND ADVANCED ENGINEERING, vol. 3, special issue 1, 30 January 2013 (30.01.2013), ISSN: 2250-2459, page 14	1-27
A	WU, Lifu et al., "An Improved Algorithm for Blind Estimation of Reverberation Time Based on Maximum Likelihood", JOURNAL OF APPLIED ACOUSTICS, vol. 35, no. 4, 30 July 2016 (30.07.2016), ISSN: 1000-310X, pages 288-293	1-27

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
27 September 2017 (27.09.2017)

Date of mailing of the international search report
10 October 2017 (10.10.2017)

Name and mailing address of the ISA/CN:
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No.: (86-10) 62019451

Authorized officer
YANG, Shilin
Telephone No.: (86-10) **62085717**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/090887

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MEI, Tiemin et al., "A New Method of Reverberation Time Estimation", Journal of Shenyang Ligong University, vol. 31, no. 6, 30 December 2012 (30.12.2012), ISSN: 1003-1251, pages 66-69	1-27
A	CN 105628170 A (GUANGZHOU AUTOMOBILE GROUP CO., LTD.), 01 June 2016 (01.06.2016), the whole document	1-27
A	JP 2009211021 A (UNIV. HOKURIKU), 17 September 2009 (17.09.2009), the whole document	1-27

International application No.
PCT/CN2017/090887

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/090887

A. 主题的分类 G01H 7/00(2006.01)i; G10K 15/08(2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类														
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G10; G01H 7 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS, CNTXT, CNKI, 百度学术, VEN: 混响, 时间, 历史, 过去, 残差, 自相关, 极大似然, 最大似然, reverberation, time, histor+, past, forgone, old, residual, autocorrelat+, maximum likelihood														
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>Keshavarz A. 等. "Speech-Model Based Accurate Blind Reverberation Time Estimation Using an LPC Filter[J]" IEEE Transactions on Audio Speech and Language Processing, 第20卷, 第6期, 2012年 8月 30日 (2012 - 08 - 30), 第1884-1893页</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>R. Jayashree 等. "Accurate Estimation Of Reverberation Time And Drr Using Maximum Likelihood Estimator" International Journal of Emerging Technology and Advanced Engineering, 第3卷, 第特刊1期, 2013年 1月 30日 (2013 - 01 - 30), ISSN: 2250-2459, 第14页</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>吴礼福 等. "一种基于最大似然的混响时间盲估计方法" 应用声学, 第35卷, 第4期, 2016年 7月 30日 (2016 - 07 - 30), ISSN: 1000-310X, 第288-293页</td> <td>1-27</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	Keshavarz A. 等. "Speech-Model Based Accurate Blind Reverberation Time Estimation Using an LPC Filter[J]" IEEE Transactions on Audio Speech and Language Processing, 第20卷, 第6期, 2012年 8月 30日 (2012 - 08 - 30), 第1884-1893页	1-27	A	R. Jayashree 等. "Accurate Estimation Of Reverberation Time And Drr Using Maximum Likelihood Estimator" International Journal of Emerging Technology and Advanced Engineering, 第3卷, 第特刊1期, 2013年 1月 30日 (2013 - 01 - 30), ISSN: 2250-2459, 第14页	1-27	A	吴礼福 等. "一种基于最大似然的混响时间盲估计方法" 应用声学, 第35卷, 第4期, 2016年 7月 30日 (2016 - 07 - 30), ISSN: 1000-310X, 第288-293页	1-27
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求												
A	Keshavarz A. 等. "Speech-Model Based Accurate Blind Reverberation Time Estimation Using an LPC Filter[J]" IEEE Transactions on Audio Speech and Language Processing, 第20卷, 第6期, 2012年 8月 30日 (2012 - 08 - 30), 第1884-1893页	1-27												
A	R. Jayashree 等. "Accurate Estimation Of Reverberation Time And Drr Using Maximum Likelihood Estimator" International Journal of Emerging Technology and Advanced Engineering, 第3卷, 第特刊1期, 2013年 1月 30日 (2013 - 01 - 30), ISSN: 2250-2459, 第14页	1-27												
A	吴礼福 等. "一种基于最大似然的混响时间盲估计方法" 应用声学, 第35卷, 第4期, 2016年 7月 30日 (2016 - 07 - 30), ISSN: 1000-310X, 第288-293页	1-27												
☒ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。														
* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件														
国际检索实际完成的日期 2017年 9月 27日		国际检索报告邮寄日期 2017年 10月 10日												
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 杨士林 电话号码 (86-10)62085717												

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	梅铁民 等. “一种新的混响时间计算方法” 沈阳理工大学学报, 第31卷, 第6期, 2012年 12月 30日 (2012 - 12 - 30), ISSN: 1003-1251, 第66-69页	1-27
A	CN 105628170 A (广州汽车集团股份有限公司) 2016年 6月 1日 (2016 - 06 - 01) 全文	1-27
A	JP 2009211021 A (UNIV. HOKURIKU) 2009年 9月 17日 (2009 - 09 - 17) 全文	1-27

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/090887

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	105628170	A	2016年 6月 1日	无	
JP	2009211021	A	2009年 9月 17日	JP 5077847 B2	2012年 11月 21日