

(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
2020 年 7 月 16 日 (16.07.2020)



(10) 国际公布号
WO 2020/143670 A1

(51) 国际专利分类号:
G10L 21/0208 (2013.01)

(21) 国际申请号: PCT/CN2020/070922

(22) 国际申请日: 2020 年 1 月 8 日 (08.01.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910016620.7 2019 年 1 月 8 日 (08.01.2019) CN

(71) 申请人: 电信科学技术研究院有限公司
(**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: 梁民 (**LIANG, Min**); 中国北京市海淀区
学院路 40 号, Beijing 100191 (CN)。 徐红艳 (**XU, Hongyan**); 中国北京市海淀区学院路 40 号, Beijing
100191 (CN)。 毕海 (**BI, Hai**); 中国北京市海淀区
学院路 40 号, Beijing 100191 (CN)。

(74) 代理人: 北京银龙知识产权代理有限公司 (**DRAGON INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区西直门北大街 32 号院
枫蓝国际中心 2 号楼 10 层, Beijing 100082 (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,

(54) **Title:** VOICE DEREVERBERATION METHOD AND DEVICE

(54) 发明名称: 语音去混响的方法及装置

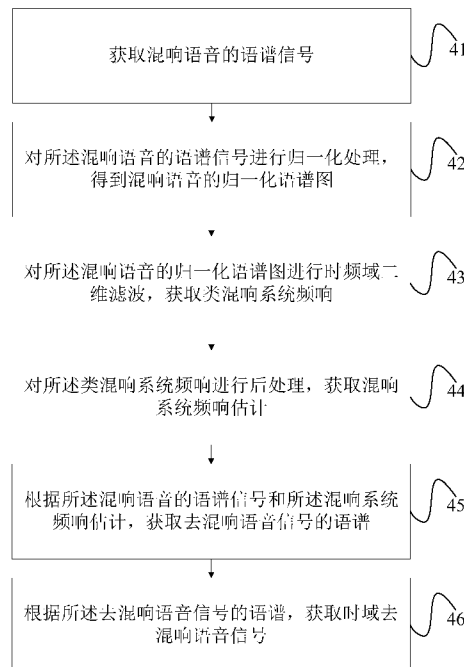


图 4

- 41 Acquiring the voice spectrum signal of the reverberant voice;
42 Normalizing the voice spectrum signal of the reverberant voice to obtain the normalized voice spectrum diagram of the reverberant voice;
43 Performing two-dimensional filtering in the time-frequency domain on the normalized voice spectrum diagram of the reverberant voice to acquire the frequency response of the reverberant-like system;
44 Post-processing the frequency response of the reverberant-like system to acquire the frequency response estimation of the reverberation system;
45 According to the voice spectrum signal of the reverberant voice and the frequency response estimation of the reverberation system, acquiring the voice spectrum of the dereverberation voice signal;
46 According to the voice spectrum of the dereverberation voice signal, acquiring the time domain dereverberation voice signal.

(57) **Abstract:** Provided is a voice dereverberation method, comprising: acquiring the voice spectrum signal of the reverberant voice (41); normalizing the voice spectrum signal of the reverberant voice to obtain the normalized voice spectrum diagram of the reverberant voice (42); performing two-dimensional filtering in the time-frequency domain on the normalized voice spectrum diagram of the reverberant voice to acquire the frequency response of the reverberant-like system (43); post-processing the frequency response of the reverberant-like system to acquire the frequency response estimation of the reverberation system (44); according to the voice spectrum

LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
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,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, 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))。

signal of the reverberant voice and the frequency response estimation of the reverberation system, acquiring the voice spectrum of the dereverberation voice signal (45); according to the voice spectrum of the dereverberation voice signal, acquiring the time domain dereverberation voice signal (46).

(57) 摘要: 一种语音去混响的方法, 包括: 获取混响语音的语谱信号(41); 对混响语音的语谱信号进行归一化处理, 得到混响语音的归一化语谱图(42); 对混响语音的归一化语谱图进行时频域二维滤波, 获取类混响系统频响(43); 对类混响系统频响进行后处理, 获取混响系统频响估计(44); 根据混响语音的语谱信号和混响系统频响估计, 获取去混响语音信号的语谱(45); 根据去混响语音信号的语谱, 获取时域去混响语音信号(46)。

语音去混响的方法及装置

相关申请的交叉引用

本申请主张在 2019 年 1 月 8 日在中国提交的中国专利申请号 No. 201910016620.7 的优先权，其全部内容通过引用包含于此。

技术领域

本公开涉及语音信号处理技术领域，特别涉及一种语音去混响的方法及装置。

背景技术

众所周知，语音去混响技术已广泛应用于免提电话、助听器、电话会议系统、高保真语音控制系统和自动语音识别系统。语音去混响方法又称混响消除法，通常大致分为有三大类，但是相关技术中的语音去混响方法又分别存在下列问题：

第一类基于麦克风阵列处理的去混响技术，其性能受限于阵列的麦克风数目，要获得令人满意的去混响结果，势必需要大量的麦克风，这便导致实际产品的成本提高和结构设计的困难增加。第二类在频域对后混响信号进行抑制处理的去混响技术需要首先估计出工作环境的混响时间参数 (RT60)，但由于目前尚缺乏关于工作环境中与频率关联的混响时间参数 (RT60) 的高精度实时估计算法，故该技术的去混响性能受限。第三类基于逆滤波思想的去混响技术中能实际应用的 WPE 方法涉及一个高阶观测数据相关矩阵的伪逆运算，因而在商用 DSP 上实现时通常耗费较多的计算资源。

发明内容

本公开一些实施例提供一种语音去混响的方法及装置，以解决相关技术中的语音去混响技术存在成本较高、结构设计较为复杂、去混响能力受限、实现耗费较多计算资源的问题。

为了解决上述技术问题，本公开一些实施例提供一种语音去混响的方法，

包括：

获取混响语音的语谱信号；

对所述混响语音的语谱信号进行归一化处理，得到混响语音的归一化语谱图；

对所述混响语音的归一化语谱图进行时频域二维滤波，获取类混响系统频响；

对所述类混响系统频响进行后处理，获取混响系统频响估计；

根据所述混响语音的语谱信号和所述混响系统频响估计，获取去混响语音信号的语谱；

根据所述去混响语音信号的语谱，获取时域去混响语音信号。

可选地，所述对所述混响语音的语谱信号进行归一化处理，得到混响语音归一化语谱图，包括：

根据公式： $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$ ，获取混响语音的归一化语谱图；

其中， $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $X(k, t)$ 为混响语音的语谱信号； X_{max} 为 $X(k, t)$ 的最大值； M 为预设的正整数，且 M 为灰度电平等级的个数； $\lfloor * \rfloor$ 为向下取整函数； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述对所述混响语音的归一化语谱图进行时频域二维滤波，获取类混响系统频响，包括：

根据公式： $H(k, t) = X_{norm}(k, t) * G(k, t)$ ，获取类混响系统频响；

其中， $H(k, t)$ 为类混响系统频响； $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $*$ 为线性卷积算子； $G(k, t)$ 为平滑核，且 $G(k, t) = \beta \cdot e^{-\frac{k^2+t^2}{2 \cdot \alpha^2}}$ ， $\iint G(k, t) dt dk = 1$ ， β 为归一化系数常数，且 β 使得 $e^{-\frac{k^2+t^2}{2 \cdot \alpha^2}}$ 的全积分等于1； α 为滤波半径，且 α 越大，图像越锐化； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述对所述类混响系统频响进行后处理，获取混响系统频响估计，包括：

根据公式：

$$\hat{H}(k, t) = \psi\{H(k, t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k, t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k, t) = H_{min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{min} < H(k, t) < H_{max} \end{cases}, \text{ 获取混}$$

响系统频响估计;

其中, $\hat{H}(k, t)$ 为混响系统频响估计; $H(k, t)$ 为类混响系统频响; $\hat{H}_{min}$ 为 $\hat{H}(k, t)$ 的最小值; $\hat{H}_{max}$ 为 $\hat{H}(k, t)$ 的最大值; H_{min} 为 $H(k, t)$ 的最小值; H_{max} 为 $H(k, t)$ 的最大值; $\Phi\{\cdot\}$ 为递减函数; k 为离散频率的索引; t 为时域信号帧索引。

可选地, 所述根据所述混响语音的语谱信号和所述混响系统频响估计, 获取去混响语音信号的语谱, 包括:

根据公式: $\hat{S}(k, t) = X(k, t)/\hat{H}(k, t)$, 获取去混响语音信号的语谱;

其中, $\hat{S}(k, t)$ 为去混响语音信号的语谱; $X(k, t)$ 为混响语音的语谱信号; $\hat{H}(k, t)$ 为混响系统频响估计; k 为离散频率的索引; t 为时域信号帧索引。

本公开一些实施例还提供一种语音去混响的装置, 包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序; 其中, 所述处理器执行所述计算机程序时实现以下步骤:

获取混响语音的语谱信号;

对所述混响语音的语谱信号进行归一化处理, 得到混响语音的归一化语谱图;

对所述混响语音的归一化语谱图进行时频域二维滤波, 获取类混响系统频响;

对所述类混响系统频响进行后处理, 获取混响系统频响估计;

根据所述混响语音的语谱信号和所述混响系统频响估计, 获取去混响语音信号的语谱;

根据所述去混响语音信号的语谱, 获取时域去混响语音信号。

可选地, 所述处理器执行所述计算机程序时实现以下步骤:

根据公式: $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$, 获取混响语音的归一化语谱图;

其中, $X_{norm}(k, t)$ 为混响语音的归一化语谱图; $X(k, t)$ 为混响语音的语谱信号; X_{max} 为 $X(k, t)$ 的最大值; M 为预设的正整数, 且 M 为灰度电平等级的

个数； $[*]$ 为向下取整函数； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述处理器执行所述计算机程序时实现以下步骤：

根据公式： $H(k, t) = X_{norm}(k, t) * G(k, t)$ ，获取类混响系统频响；

其中， $H(k, t)$ 为类混响系统频响； $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $*$ 为线性卷积算子； $G(k, t)$ 为平滑核，且 $G(k, t) = \beta \cdot e^{-\frac{k^2+t^2}{2\alpha^2}}$ ， $\iint G(k, t) dt dk = 1$ ， β 为归一化系数常数，且 β 使得 $e^{-\frac{k^2+t^2}{2\alpha^2}}$ 的全积分等于1； α 为滤波半径，且 α 越大，图像越锐化； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述处理器执行所述计算机程序时实现以下步骤：

根据公式：

$$\hat{H}(k, t) = \psi\{H(k, t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k, t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k, t) = H_{min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{min} < H(k, t) < H_{max} \end{cases}, \text{ 获取混}$$

响系统频响估计；

其中， $\hat{H}(k, t)$ 为混响系统频响估计； $H(k, t)$ 为类混响系统频响； $\hat{H}_{min}$ 为 $\hat{H}(k, t)$ 的最小值； $\hat{H}_{max}$ 为 $\hat{H}(k, t)$ 的最大值； H_{min} 为 $H(k, t)$ 的最小值； H_{max} 为 $H(k, t)$ 的最大值； $\Phi\{\cdot\}$ 为递减函数； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述处理器执行所述计算机程序时实现以下步骤：

根据公式： $\hat{S}(k, t) = X(k, t)/\hat{H}(k, t)$ ，获取去混响语音信号的语谱；

其中， $\hat{S}(k, t)$ 为去混响语音信号的语谱； $X(k, t)$ 为混响语音的语谱信号； $\hat{H}(k, t)$ 为混响系统频响估计； k 为离散频率的索引； t 为时域信号帧索引。

本公开一些实施例还提供一种计算机可读存储介质，其上存储有计算机程序，其中，所述计算机程序被处理器执行时实现上述的语音去混响的方法中的步骤。

本公开一些实施例还提供一种语音去混响的装置，包括：

第一获取模块，用于获取混响语音的语谱信号；

第二获取模块，用于对所述混响语音的语谱信号进行归一化处理，得到混响语音的归一化语谱图；

第三获取模块，用于对所述混响语音的归一化语谱图进行时频域二维滤

波，获取类混响系统频响；

第四获取模块，用于对所述类混响系统频响进行后处理，获取混响系统频响估计；

第五获取模块，用于根据所述混响语音的语谱信号和所述混响系统频响估计，获取去混响语音信号的语谱；

第六获取模块，用于根据所述去混响语音信号的语谱，获取时域去混响语音信号。

可选地，所述第二获取模块，用于：

根据公式： $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$ ，获取混响语音的归一化语谱图；

其中， $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $X(k, t)$ 为混响语音的语谱信号； X_{max} 为 $X(k, t)$ 的最大值； M 为预设的正整数，且 M 为灰度电平等级的个数； $\lfloor \cdot \rfloor$ 为向下取整函数； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述第三获取模块，用于：

根据公式： $H(k, t) = X_{norm}(k, t) * G(k, t)$ ，获取类混响系统频响；

其中， $H(k, t)$ 为类混响系统频响； $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $*$ 为线性卷积算子； $G(k, t)$ 为平滑核，且 $G(k, t) = \beta \cdot e^{-\frac{k^2+t^2}{2\cdot\alpha^2}}$ ， $\iint G(k, t) dt dk = 1$ ， β 为归一化系数常数，且 β 使得 $e^{-\frac{k^2+t^2}{2\cdot\alpha^2}}$ 的全积分等于1； α 为滤波半径，且 α 越大，图像越锐化； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述第四获取模块，用于：

根据公式： $\hat{H}(k, t) = \psi\{H(k, t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k, t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k, t) = H_{min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{min} < H(k, t) < H_{max} \end{cases}$ ，获取混响系统频响估计；

其中， $\hat{H}(k, t)$ 为混响系统频响估计； $H(k, t)$ 为类混响系统频响； $\hat{H}_{min}$ 为 $\hat{H}(k, t)$ 的最小值； $\hat{H}_{max}$ 为 $\hat{H}(k, t)$ 的最大值； H_{min} 为 $H(k, t)$ 的最小值； H_{max} 为 $H(k, t)$ 的最大值； $\Phi\{\cdot\}$ 为递减函数； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述第五获取模块，用于：

根据公式： $\hat{S}(k, t) = X(k, t) / \hat{H}(k, t)$ ，获取去混响语音信号的语谱；

其中， $\hat{S}(k, t)$ 为去混响语音信号的语谱； $X(k, t)$ 为混响语音的语谱信号； $\hat{H}(k, t)$ 为混响系统频响估计； k 为离散频率的索引； t 为时域信号帧索引。

本公开的有益效果是：

上述方案，通过对混响语音的语谱信号进行归一化处理，得到混响语音的归一化语谱图，在对混响语音的归一化语谱图进行时频域二维滤波，获取类混响系统频响，然后对类混响系统频响进行后处理，获取混响系统频响估计，最后得到时域去混响语音信号，此种方式，可以节省去混响的计算资源，降低了去混响成本，可以实现有效的去混响，即使在严重混响的环境中，自动语音识别的性能也有明显的提高。

附图说明

图 1 表示本公开一些实施例的语音去混响的方法的流程示意图；

图 2 表示应用 SR 模型来进行语音信号去混响处理的原理示意图；

图 3 表示变换函数 ψ 的第一种实现方式示意图；

图 4 表示变换函数 ψ 的第二种实现方式示意图；

图 5 表示本公开一些实施例的语音去混响的装置的模块示意图；

图 6 表示本公开一些实施例的语音去混响的装置的结构示意图。

具体实施方式

为使本公开的目的、技术方案和优点更加清楚，下面将结合附图及具体实施例对本公开进行详细描述。

在进行本公开一些实施例的说明时，首先对下面描述中所用到的一些概念进行解释说明。

语音去混响方法又称混响消除法，通常大致分为有三大类：

第一类是采用麦克风阵列处理技术，该技术首先估计声源相对麦克风阵列的方位 (Direction of Arrival, DOA)，通过控制麦克风阵列的方向性来增强来自声源方向的直达信号成分，并减小和消除来自其它方向的声源反射信号成分，从而达到去混响的目的，为了获得满意的去混响效果，该技术通

常需要大量数目的麦克风，以便阵列获得充分的方向性增益。

第二类去混响技术则是在频域对后混响信号进行抑制处理的方法，该方法首先估计出工作环境的混响时间参数 (RT60)，并据此估计出后混响信号的功率谱，然后应用噪声抑制中的谱减法对后混响信号进行抑制处理，尽管该技术不涉及信号的相位信息而使其处理性能具有较好的鲁棒性，但由于目前尚缺乏关于工作环境中与频率关联的混响时间参数 (RT60) 的高精度实时估计算法，故该技术的去混响性能受限。

第三类去混响技术则是基于逆滤波的思想，其目标是估计出引发混响的室内冲激响应 (Room Impulse Response, RIR) 的逆滤波器，用其对混响语音信号进行滤波处理以恢复源信号，在声源到麦克风的室内传递函数 (Room Transfer Function, RTF) 已知的情况下，用 RTF 的逆滤波器可以从观测的混响信号中精确地恢复出其源信号，业已证明：在麦克风数目大于已激活的声源数目、并且每个声源到每个麦克风的 RTF 不存在共同的零点的条件下，上述功能的逆滤波器解是存在的。然而在实际应用中，RTF (或其等效的逆滤波器) 是时变的、未知的，需要从已获的观测数据中估计出。为此，大量学者致力于该领域的探索和研究，提出了许多方法，最为引人注目的便是基于延时的线性预测 (Delayed Linear Prediction, DLP) 的后混响抑制技术，该技术能有效地抑制后混响成分而未明显地损伤语音的短时相关性，但它要求 DLP 的滤波器阶数很高 (滤波器通常有数千个系数)，因而需要很长的观测数据，由此导致该技术具有很高的计算负荷，难以在商用的数字信号处理器 (Digital Signal Processor, DSP) 芯片上实时实现。

此外，人们还提出将时变语音信号源模型与多声道线性预测相结合来进行去混响的方法，该方法可以基于较短的观测数据有效地抑制后混响，而且对前混响也有抑制的效果，但它固有的计算复杂度致使其无法在实际中应用。最近，人们将基于 DLP 的去混响技术拓展到处理时变语音信号的场景，提出了一种称之为方差归一化延时的线性预测 (NDLP) 去混响技术，NDLP 的频域实现即为著名的加权预测误差 (Weighted Prediction Error, WPE) 去混响算法，尽管 WPE 性能具有较好的鲁棒性，但它涉及一个高阶观测数据相关矩阵的伪逆运算，因而在商用 DSP 上实现时通常耗费较多的计算资源。

正如上面所述，本公开针对相关技术中的语音去混响技术存在成本较高、结构设计较为复杂、去混响能力受限、实现耗费较多计算资源的问题，提供一种语音去混响的方法及装置。

具体地，本公开基于环绕视网膜皮层（Surround Retinex, SR）模型，提出了一种新颖实用的去混响方法。该方法与上述传统的去混响技术相比，具有更合理的计算复杂度，可以实现有效的去混响，即使在严重混响的环境中，自动语音识别（Automatic Speech Recognition, ASR）的性能也有所提高。

下面对本公开一些实施例的实现原理进行说明如下。

本公开基于环绕视网膜皮层（Surround Retinex, SR）模型，提出了一种新颖实用的去混响技术。与上述传统的去混响技术相比，具有更合理的计算复杂度，可以实现有效的去混响，即使在严重混响的环境中，自动语音识别（Automatic Speech Recognition, ASR）的性能也有所提高。

本公开的主要思想是：既然环绕视网膜皮层（SR）模型已被证明是一种有效的图像增强工具，它能够从退化图像中估计出图像照明源，那么鉴于作为混响语音信号时-频域有效表征的“语谱图”（Spectrogram）可以看成类似于具有被污染的场景照明图像，一种自然的想法便是应用 SR 模型从混响语音信号的“语谱图”中估计出去混响的语音信号“语谱图”，进而获得时域去混响语音信号。

一、SR 模型简介：

“视网膜皮层”模型是著名学者 E. H. Land 和 J. J. McCann 于 1971 年提出的一种基于人视觉系统的图像增强理论，这一理论指出：尽管到达眼睛的视觉光量取决于反射率和光照，但是一个自然场景中感知的图像与反射率有很强的相关性。换言之，即使在困难的光照情况下，人类视觉系统也能通过依靠场景的反射率和忽视场景照明的方式能感知颜色。这个理论是基于反射率图像模型，在数学上可以表述为：

$$\text{公式一、} F(x,y) = R(x,y) \cdot I(x,y)$$

其中， $F(x,y)$ 表示一个自然场景的感知图像； $R(x,y)$ 表示一个反射图像，它仅取决于场景表面的反射率，对应于高频的反射亮度； $I(x,y)$ 表示一个光照图像，它由照明光源决定并与照明量有关，对应于低频的亮度。

SR 模型的关键技术是基于感知图像 $F(x, y)$ 来估计光照图像 $I(x, y)$ 。D. J. Jobson 等人建议：光照图像 $I(x, y)$ 可以估计为感知图像 $F(x, y)$ 的一种模糊方案，即利用公式二表示为：

$$\text{公式二: } \hat{I}(x, y) = F(x, y) * G(x, y)$$

其中， $*$ 为线性卷积算子， $G(x, y)$ 为平滑核， $G(x, y)$ 通常取为下述公式三的 Gauss 核形式：

$$\text{公式三: } G(x, y) = \beta \cdot e^{-\frac{x^2+y^2}{2\alpha^2}}, \quad \iint G(x, y) dx dy = 1$$

其中， α 是滤波半径， α 越大，图像越锐化； β 是为归一化系数常数，使得公式后半部分的全积分等于 1。因此，反射图像 $R(x, y)$ 的估计可用公式四表达为：

$$\text{公式四: } \hat{R}(x, y) = F(x, y) / \hat{I}(x, y)$$

其中， $\hat{R}(x, y)$ 为反射图像 $R(x, y)$ 的估计，这里 $\hat{R}(x, y)$ 恢复的是独立于光照的成分。

二、基于 SR 模型的语音信号去混响技术

通常情况下，一个混响时域数字语音信号 $x(n)$ 数学上可表征为：

$$\text{公式五: } x(n) = s(n) * h(n)$$

其中， $*$ 为线性卷积算子， $s(n)$ 为源语音数字信号， $h(n)$ 为信号源与麦克风间的信道冲激响应。可见，混响语音信号 $x(n)$ 是“干净”语音信号 $s(n)$ 和冲击响应 $h(n)$ 的线性卷积。

对公式五两边进行短时离散傅里叶变化（STDFT）便得：

$$\text{公式六: } X(k, t) = S(k, t) \cdot H(k, t)$$

其中 $X(k, t)$ 、 $S(k, t)$ 和 $H(k, t)$ 分别为信号 $x(n)$ 、 $s(n)$ 和 $h(n)$ 的 STDFT， k 为离散频率的索引， t 为时域信号帧索引。

比较公式一和公式六，我们可以看出 SR 模型和混响语音模型之间存在表 1 所示的对应关系。

数学模型	环绕视网膜皮层模型	混响语音模型
信号类别	图像	语音
模型	$F(x, y) = R(x, y) \cdot I(x, y)$	$X(k, t) = S(k, t) \cdot H(k, t)$

获取的信号	退化图像: $F(x, y)$	混响语音语谱: $X(k, t)$
源信号	“干净”图像: $R(x, y)$	“干净”语音语谱: $S(k, t)$
退化源	光照图像: $I(x, y)$	混响系统频响: $H(k, t)$

表 1 SR 模型和混响语音模型之间的对应关系

由此我们提出应用 SR 模型来进行语音信号去混响处理的算法, 其原理框图如图 1 所示, 其中“STDFT”模块将时域混响语音 $x(n)$ 转化为语谱信号 $X(k, t)$; “语谱归一化处理”模块将语谱信号 $X(k, t)$ 归一化为具有 M 个灰度电平等级的信号 $X_{\text{norm}}(k, t)$, 然后用公式三推理得到的 Gauss 平滑核 $G(k, t)$ (注: 这里取 $x=k$, $y=t$) 对 $X_{\text{norm}}(k, t)$ 进行时-频域二维滤波获得类似于 SR 模型中“光照图像”的类混响系统频响 $H(k, t)$ 。

作为一个实施例, 我们给出混响语音语谱图的一种归一化方法如下:

对混响语音信号语谱图 $X(k, t)$ 而言, 记 $\max\{|X(k, t)|\}$ 为 X_{max} , 那么 $X(k, t)$ 对应的具有 M 个灰度电平的归一化语谱图 $X_{\text{norm}}(k, t)$ 定义为:

$$\text{公式七: } X_{\text{norm}}(k, t) \triangleq \left\lfloor \left[\frac{|X(k, t)|}{X_{\text{max}}} \right] (M - 1) + 0.5 \right\rfloor$$

其中 $\lfloor x \rfloor$ 表示不大于 x 的最大整数。

考虑到光照图像中的重要信息 (例如人脸图像中眼和嘴) 通常由低灰度值像素来表征, 而语音信号语谱图中的高灰度值像素则代表了重要的语音信息, 那么我们需要对估计的“光照图像” $H(k, t)$ 进行后处理以便完成这种对应关系的转换, “后处理变换器 ψ ” 模块便是为实现这种转化而设计的, 这里将转换函数 ψ 定义为:

$$\text{公式八: } \hat{H}(k, t) \triangleq \psi\{H(k, t)\} = \begin{cases} \hat{H}_{\min}, & \text{若 } H(k, t) = H_{\max} \\ \hat{H}_{\max}, & \text{若 } H(k, t) = H_{\min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{\min} < H(k, t) < H_{\max} \end{cases}$$

其中 $H_{\max} \triangleq \max\{H(k, t)\}$ 为 $H(k, t)$ 的最大值, $H_{\min} \triangleq \min\{H(k, t)\}$ 为 $H(k, t)$ 的最小值, 函数 $\Phi\{\cdot\}$ 为递减函数, $\hat{H}_{\min}$ 和 $\hat{H}_{\max}$ 分别为预设的 $\hat{H}(k, t)$ 最小值和最大值参数。

如图 2 所示, 为公式八的一种可选地实现形式, 采用图 2 中的抛物线曲线的左半部实现部作为变换函数 ψ , 其中, 图 2 中的抛物线曲线的具体定义函数为: $\hat{H}(k, t) = A1 \cdot [H(k, t) - H_0]^2 + B1$, $A1$ 为预设常数; $B1$ 为预设常数,

H_0 为抛物线曲线的最小值。

如图 3 所示, 为公式八的另一种可选地实现形式, 采用图 3 中的一段直线作为变换函数 ψ , 其中, 图 3 中的直线的具体定义函数为: $\hat{H}(k, t) = A2 \cdot H(k, t) + B2$, $A2$ 为预设常数; $B2$ 为预设常数。

根据已求得的 $X(k, t)$ 与 $\hat{H}(k, t)$, “去混响语音语谱估计器”模块按公式九计算去混响语音信号的语谱 $\hat{S}(k, t)$:

$$\text{公式九: } \hat{S}(k, t) = X(k, t) / \hat{H}(k, t)$$

最后应用“逆短时离散傅里叶变换 (ISTDFT)”模块将去混响语音信号的语谱 $\hat{S}(k, t)$ 变换回时域即得时域去混语音信号 $\hat{S}(n)$ 。

下面对本公开一些实施例的具体实现过程说明如下。

如图 4 所示, 本公开一些实施例提供一种语音去混响的方法, 包括:

步骤 41, 获取混响语音的语谱信号;

步骤 42, 对所述混响语音的语谱信号进行归一化处理, 得到混响语音的归一化语谱图;

步骤 43, 对所述混响语音的归一化语谱图进行时频域二维滤波, 获取类混响系统频响;

步骤 44, 对所述类混响系统频响进行后处理, 获取混响系统频响估计;

步骤 45, 根据所述混响语音的语谱信号和所述混响系统频响估计, 获取去混响语音信号的语谱;

步骤 46, 根据所述去混响语音信号的语谱, 获取时域去混响语音信号。

具体地, 所述步骤 42 的实现方式为:

根据上述公式七: $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$, 获取混响语音的归一化语谱图;

其中, $X_{norm}(k, t)$ 为混响语音的归一化语谱图; $X(k, t)$ 为混响语音的语谱信号; X_{max} 为 $X(k, t)$ 的最大值; M 为预设的正整数, 且 M 为灰度电平等级的个数; $\lfloor * \rfloor$ 为向下取整函数; k 为离散频率的索引; t 为时域信号帧索引。

具体地, 所述步骤 43 的实现方式为:

根据公式十: $H(k, t) = X_{norm}(k, t) * G(k, t)$, 获取类混响系统频响;

其中, $H(k, t)$ 为类混响系统频响; $X_{norm}(k, t)$ 为混响语音的归一化语谱

图；*为线性卷积算子； $G(k,t)$ 为平滑核，且 $G(k,t) = \beta \cdot e^{-\frac{k^2+t^2}{2\alpha^2}}$ ， $\iint G(k,t) dt dk = 1$ ， β 为归一化系数常数，且 β 使得 $e^{-\frac{k^2+t^2}{2\alpha^2}}$ 的全积分等于1； α 为滤波半径，且 α 越大，图像越锐化； k 为离散频率的索引； t 为时域信号帧索引。

具体地，所述步骤44的实现方式为：

根据上述公式八：

$$\hat{H}(k,t) = \psi\{H(k,t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k,t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k,t) = H_{min} \\ \Phi\{H(k,t)\}, & \text{若 } H_{min} < H(k,t) < H_{max} \end{cases}, \text{ 获取混}$$

响系统频响估计；

其中， $\hat{H}(k,t)$ 为混响系统频响估计； $H(k,t)$ 为类混响系统频响； $\hat{H}_{min}$ 为 $\hat{H}(k,t)$ 的最小值； $\hat{H}_{max}$ 为 $\hat{H}(k,t)$ 的最大值； H_{min} 为 $H(k,t)$ 的最小值； H_{max} 为 $H(k,t)$ 的最大值； $\Phi\{\cdot\}$ 为递减函数； k 为离散频率的索引； t 为时域信号帧索引。

具体地，所述步骤45的实现方式为：

根据公式十一： $\hat{S}(k,t) = X(k,t)/\hat{H}(k,t)$ ，获取去混响语音信号的语谱；

其中， $\hat{S}(k,t)$ 为去混响语音信号的语谱； $X(k,t)$ 为混响语音的语谱信号； $\hat{H}(k,t)$ 为混响系统频响估计； k 为离散频率的索引； t 为时域信号帧索引。

需要说明的是，步骤46的具体实现方式为：应用逆短时离散傅里叶变换（ISTDFT）将去混响语音信号的语谱 $\hat{S}(k,t)$ 变换回时域，便可得到时域去混响语音信号。

本公开一些实施例通过对比图像增强技术中的SR模型与混响语音信号的语谱图模型，给出了两者之间的对应关系，据此将SR模型的相关算法应用于混响语音语谱图以便完成混响语音信号的去混响任务；为便于应用SR模型，本公开一些实施例首先将混响语音信号的语谱图 $X(k,t)$ 转化为一个M级灰度的语谱图 $X_{norm}(k,t)$ ，然后用Gauss平滑核函数对之进行时-频域二维滤波处理而获得“光照图像” $H(k,t)$ ；考虑到图像中重要信息表征与语音语谱图中重要信息表征的逆对应关系（即：图像中重要信息由低灰度级的像素来表征，而语音语谱图中的重要信息由高灰度级的像素来表征），采用后处理变换的方式将 $H(k,t)$ 转化为所需的混响系统频响 $\hat{H}(k,t)$ ，并根据混响系统频响

$\hat{H}(k, t)$ 和混响语音的语谱信号来计算去混响语音信号的语谱,从而获得去混响的时域去混响语音信号。

本公开一些实施例可以节省去混响的计算资源,降低去混响成本,进而实现有效的去混响,即使在严重混响的环境中,自动语音识别的性能也有明显的提高。

如图5所示,本公开一些实施例还提供一种语音去混响的装置,包括:

第一获取模块51,用于获取混响语音的语谱信号;

第二获取模块52,用于对所述混响语音的语谱信号进行归一化处理,得到混响语音的归一化语谱图;

第三获取模块53,用于对所述混响语音的归一化语谱图进行时频域二维滤波,获取类混响系统频响;

第四获取模块54,用于对所述类混响系统频响进行后处理,获取混响系统频响估计;

第五获取模块55,用于根据所述混响语音的语谱信号和所述混响系统频响估计,获取去混响语音信号的语谱;

第六获取模块56,用于根据所述去混响语音信号的语谱,获取时域去混响语音信号。

进一步地,所述第二获取模块52,用于:

根据公式: $X_{norm}(k, t) = \left\lfloor \frac{|X(k, t)|}{X_{max}} \right\rfloor (M - 1) + 0.5$, 获取混响语音的归一化语谱图;

其中, $X_{norm}(k, t)$ 为混响语音的归一化语谱图; $X(k, t)$ 为混响语音的语谱信号; X_{max} 为 $X(k, t)$ 的最大值; M 为预设的正整数,且 M 为灰度电平等级的个数; $\lfloor * \rfloor$ 为向下取整函数; k 为离散频率的索引; t 为时域信号帧索引。

进一步地,所述第三获取模块53,用于:

根据公式: $H(k, t) = X_{norm}(k, t) * G(k, t)$, 获取类混响系统频响;

其中, $H(k, t)$ 为类混响系统频响; $X_{norm}(k, t)$ 为混响语音的归一化语谱图; $*$ 为线性卷积算子; $G(k, t)$ 为平滑核,且 $G(k, t) = \beta \cdot e^{-\frac{k^2+t^2}{2\cdot\alpha^2}}$, $\int \int G(k, t) dt dk = 1$, β 为归一化系数常数,且 β 使得 $e^{-\frac{k^2+t^2}{2\cdot\alpha^2}}$ 的全积分等于1; α 为滤波半径,且 α 越大,图像越锐化; k 为离散频率的索引; t 为时域信

号帧索引。

进一步地, 所述第四获取模块 54, 用于:

$$\begin{aligned} & \text{根据公式: } \hat{H}(k, t) = \psi\{H(k, t)\} = \\ & \begin{cases} \hat{H}_{\min}, & \text{若 } H(k, t) = H_{\max} \\ \hat{H}_{\max}, & \text{若 } H(k, t) = H_{\min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{\min} < H(k, t) < H_{\max} \end{cases}, \text{ 获取混响系统频响估计;} \end{aligned}$$

其中, $\hat{H}(k, t)$ 为混响系统频响估计; $H(k, t)$ 为类混响系统频响; $\hat{H}_{\min}$ 为 $\hat{H}(k, t)$ 的最小值; $\hat{H}_{\max}$ 为 $\hat{H}(k, t)$ 的最大值; $H_{\min}$ 为 $H(k, t)$ 的最小值; $H_{\max}$ 为 $H(k, t)$ 的最大值; $\Phi\{\cdot\}$ 为递减函数; k 为离散频率的索引; t 为时域信号帧索引。

进一步地, 所述第五获取模块 55, 用于:

根据公式: $\hat{S}(k, t) = X(k, t)/\hat{H}(k, t)$, 获取去混响语音信号的语谱;

其中, $\hat{S}(k, t)$ 为去混响语音信号的语谱; $X(k, t)$ 为混响语音的语谱信号; $\hat{H}(k, t)$ 为混响系统频响估计; k 为离散频率的索引; t 为时域信号帧索引。

需要说明的是, 该装置的实施例是与上述方法实施例一一对应的装置, 上述方法实施例中所有实现方式均适用于该装置的实施例中, 也能达到相同的技术效果。

如图 6 所示, 本公开一些实施例还提供一种语音去混响的装置, 包括处理器 61、存储器 62 及存储在所述存储器 62 上并可在所述处理器 61 上运行的计算机程序; 其中, 所述处理器 61 用于读取存储器中的程序, 执行下列过程:

获取混响语音的语谱信号;

对所述混响语音的语谱信号进行归一化处理, 得到混响语音的归一化语谱图;

对所述混响语音的归一化语谱图进行时频域二维滤波, 获取类混响系统频响;

对所述类混响系统频响进行后处理, 获取混响系统频响估计;

根据所述混响语音的语谱信号和所述混响系统频响估计, 获取去混响语音信号的语谱;

根据所述去混响语音信号的语谱, 获取时域去混响语音信号。

需要说明的是，在图 6 中，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 61 代表的一个或多个处理器和存储器 62 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。针对不同的装置，处理器 61 负责管理总线架构和通常的处理，存储器 62 可以存储处理器 61 在执行操作时所使用的数据。

可选地，所述处理器 61 执行所述计算机程序时实现以下步骤：

根据公式： $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$ ，获取混响语音的归一化语谱图；

其中， $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $X(k, t)$ 为混响语音的语谱信号； X_{max} 为 $X(k, t)$ 的最大值； M 为预设的正整数，且 M 为灰度电平等级的个数； $\lfloor \cdot \rfloor$ 为向下取整函数； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述处理器 61 执行所述计算机程序时实现以下步骤：

根据公式： $H(k, t) = X_{norm}(k, t) * G(k, t)$ ，获取类混响系统频响；

其中， $H(k, t)$ 为类混响系统频响； $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $*$ 为线性卷积算子； $G(k, t)$ 为平滑核，且 $G(k, t) = \beta \cdot e^{-\frac{k^2+t^2}{2\alpha^2}}$ ， $\iint G(k, t) dt dk = 1$ ， β 为归一化系数常数，且 β 使得 $e^{-\frac{k^2+t^2}{2\alpha^2}}$ 的全积分等于 1； α 为滤波半径，且 α 越大，图像越锐化； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述处理器 61 执行所述计算机程序时实现以下步骤：

根据公式： $\hat{H}(k, t) = \psi\{H(k, t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k, t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k, t) = H_{min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{min} < H(k, t) < H_{max} \end{cases}$ ，获取混响系统频响估计；

其中， $\hat{H}(k, t)$ 为混响系统频响估计； $H(k, t)$ 为类混响系统频响； $\hat{H}_{min}$ 为 $\hat{H}(k, t)$ 的最小值； $\hat{H}_{max}$ 为 $\hat{H}(k, t)$ 的最大值； H_{min} 为 $H(k, t)$ 的最小值； H_{max} 为 $H(k, t)$ 的最大值； $\Phi\{\cdot\}$ 为递减函数； k 为离散频率的索引； t 为时域信号帧索引。

可选地，所述处理器 61 执行所述计算机程序时实现以下步骤：

根据公式： $\hat{S}(k, t) = X(k, t) / \hat{H}(k, t)$ ，获取去混响语音信号的语谱；
其中， $\hat{S}(k, t)$ 为去混响语音信号的语谱； $X(k, t)$ 为混响语音的语谱信号；
 $\hat{H}(k, t)$ 为混响系统频响估计； k 为离散频率的索引； t 为时域信号帧索引。

本公开一些实施例还提供一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述的语音去混响的方法。

以上所述的是本公开的一些实施方式，应当指出对于本技术领域的普通人员来说，在不脱离本公开所述的原理前提下还可以作出若干改进和润饰，这些改进和润饰也在本公开的保护范围内。

权利要求书

1. 一种语音去混响的方法，包括：

获取混响语音的语谱信号；

对所述混响语音的语谱信号进行归一化处理，得到混响语音的归一化语谱图；

对所述混响语音的归一化语谱图进行时频域二维滤波，获取类混响系统频响；

对所述类混响系统频响进行后处理，获取混响系统频响估计；

根据所述混响语音的语谱信号和所述混响系统频响估计，获取去混响语音信号的语谱；

根据所述去混响语音信号的语谱，获取时域去混响语音信号。

2. 根据权利要求1所述的语音去混响的方法，其中，所述对所述混响语音的语谱信号进行归一化处理，得到混响语音归一化语谱图，包括：

根据公式： $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$ ，获取混响语音的归一化语谱图；

其中， $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $X(k, t)$ 为混响语音的语谱信号； X_{max} 为 $X(k, t)$ 的最大值； M 为预设的正整数，且 M 为灰度电平等级的个数； $\lfloor * \rfloor$ 为向下取整函数； k 为离散频率的索引； t 为时域信号帧索引。

3. 根据权利要求1所述的语音去混响的方法，其中，所述对所述混响语音的归一化语谱图进行时频域二维滤波，获取类混响系统频响，包括：

根据公式： $H(k, t) = X_{norm}(k, t) * G(k, t)$ ，获取类混响系统频响；

其中， $H(k, t)$ 为类混响系统频响； $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $*$ 为线性卷积算子； $G(k, t)$ 为平滑核，且 $G(k, t) = \beta \cdot e^{-\frac{k^2 + t^2}{2 \cdot \alpha^2}}$ ， $\iint G(k, t) dt dk = 1$ ， β 为归一化系数常数，且 β 使得 $e^{-\frac{k^2 + t^2}{2 \cdot \alpha^2}}$ 的全积分等于1； α 为滤波半径，且 α 越大，图像越锐化； k 为离散频率的索引； t 为时域信号帧索引。

4. 根据权利要求1所述的语音去混响的方法，其中，所述对所述类混响系统频响进行后处理，获取混响系统频响估计，包括：

根据公式：

$$\hat{H}(k, t) = \psi\{H(k, t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k, t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k, t) = H_{min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{min} < H(k, t) < H_{max} \end{cases}, \text{ 获取混}$$

响系统频响估计;

其中, $\hat{H}(k, t)$ 为混响系统频响估计; $H(k, t)$ 为类混响系统频响; $\hat{H}_{min}$ 为 $\hat{H}(k, t)$ 的最小值; $\hat{H}_{max}$ 为 $\hat{H}(k, t)$ 的最大值; H_{min} 为 $H(k, t)$ 的最小值; H_{max} 为 $H(k, t)$ 的最大值; $\Phi\{\cdot\}$ 为递减函数; k 为离散频率的索引; t 为时域信号帧索引。

5. 根据权利要求 1 所述的语音去混响的方法, 其中, 所述根据所述混响语音的语谱信号和所述混响系统频响估计, 获取去混响语音信号的语谱, 包括:

根据公式: $\hat{S}(k, t) = X(k, t)/\hat{H}(k, t)$, 获取去混响语音信号的语谱;

其中, $\hat{S}(k, t)$ 为去混响语音信号的语谱; $X(k, t)$ 为混响语音的语谱信号; $\hat{H}(k, t)$ 为混响系统频响估计; k 为离散频率的索引; t 为时域信号帧索引。

6. 一种语音去混响的装置, 包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序; 其中, 所述处理器执行所述计算机程序时实现以下步骤:

获取混响语音的语谱信号;

对所述混响语音的语谱信号进行归一化处理, 得到混响语音的归一化语谱图;

对所述混响语音的归一化语谱图进行时频域二维滤波, 获取类混响系统频响;

对所述类混响系统频响进行后处理, 获取混响系统频响估计;

根据所述混响语音的语谱信号和所述混响系统频响估计, 获取去混响语音信号的语谱;

根据所述去混响语音信号的语谱, 获取时域去混响语音信号。

7. 根据权利要求 6 所述的语音去混响的装置, 其中, 所述处理器执行所述计算机程序时实现以下步骤:

根据公式: $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$, 获取混响语音的归一化语谱图;

其中, $X_{norm}(k, t)$ 为混响语音的归一化语谱图; $X(k, t)$ 为混响语音的语谱信号; X_{max} 为 $X(k, t)$ 的最大值; M 为预设的正整数, 且 M 为灰度电平等级的个数; $[*]$ 为向下取整函数; k 为离散频率的索引; t 为时域信号帧索引。

8. 根据权利要求 6 所述的语音去混响的装置, 其中, 所述处理器执行所述计算机程序时实现以下步骤:

根据公式: $H(k, t) = X_{norm}(k, t) * G(k, t)$, 获取类混响系统频响;

其中, $H(k, t)$ 为类混响系统频响; $X_{norm}(k, t)$ 为混响语音的归一化语谱图; $*$ 为线性卷积算子; $G(k, t)$ 为平滑核, 且 $G(k, t) = \beta \cdot e^{-\frac{k^2+t^2}{2\alpha^2}}$, $\iint G(k, t) dt dk = 1$, β 为归一化系数常数, 且 β 使得 $e^{-\frac{k^2+t^2}{2\alpha^2}}$ 的全积分等于 1; α 为滤波半径, 且 α 越大, 图像越锐化; k 为离散频率的索引; t 为时域信号帧索引。

9. 根据权利要求 6 所述的语音去混响的装置, 其中, 所述处理器执行所述计算机程序时实现以下步骤:

根据公式:

$$\hat{H}(k, t) = \psi\{H(k, t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k, t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k, t) = H_{min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{min} < H(k, t) < H_{max} \end{cases}, \text{ 获取混}$$

响系统频响估计;

其中, $\hat{H}(k, t)$ 为混响系统频响估计; $H(k, t)$ 为类混响系统频响; $\hat{H}_{min}$ 为 $\hat{H}(k, t)$ 的最小值; $\hat{H}_{max}$ 为 $\hat{H}(k, t)$ 的最大值; H_{min} 为 $H(k, t)$ 的最小值; H_{max} 为 $H(k, t)$ 的最大值; $\Phi\{\cdot\}$ 为递减函数; k 为离散频率的索引; t 为时域信号帧索引。

10. 根据权利要求 6 所述的语音去混响的装置, 其中, 所述处理器执行所述计算机程序时实现以下步骤:

根据公式: $\hat{S}(k, t) = X(k, t) / \hat{H}(k, t)$, 获取去混响语音信号的语谱;

其中, $\hat{S}(k, t)$ 为去混响语音信号的语谱; $X(k, t)$ 为混响语音的语谱信号; $\hat{H}(k, t)$ 为混响系统频响估计; k 为离散频率的索引; t 为时域信号帧索引。

11. 一种计算机可读存储介质, 其上存储有计算机程序, 其中, 所述计算机程序被处理器执行时实现如权利要求 1 至 5 任一项所述的语音去混响的方法中的步骤。

12. 一种语音去混响的装置，包括：

第一获取模块，用于获取混响语音的语谱信号；

第二获取模块，用于对所述混响语音的语谱信号进行归一化处理，得到混响语音的归一化语谱图；

第三获取模块，用于对所述混响语音的归一化语谱图进行时频域二维滤波，获取类混响系统频响；

第四获取模块，用于对所述类混响系统频响进行后处理，获取混响系统频响估计；

第五获取模块，用于根据所述混响语音的语谱信号和所述混响系统频响估计，获取去混响语音信号的语谱；

第六获取模块，用于根据所述去混响语音信号的语谱，获取时域去混响语音信号。

13. 根据权利要求 12 所述的语音去混响的装置，其中，所述第二获取模块，用于：

根据公式： $X_{norm}(k, t) = \left\lfloor \left[\frac{|X(k, t)|}{X_{max}} \right] (M - 1) + 0.5 \right\rfloor$ ，获取混响语音的归一化语谱图；

其中， $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $X(k, t)$ 为混响语音的语谱信号； X_{max} 为 $X(k, t)$ 的最大值； M 为预设的正整数，且 M 为灰度电平等级的个数； $\lfloor * \rfloor$ 为向下取整函数； k 为离散频率的索引； t 为时域信号帧索引。

14. 根据权利要求 12 所述的语音去混响的装置，其中，所述第三获取模块，用于：

根据公式： $H(k, t) = X_{norm}(k, t) * G(k, t)$ ，获取类混响系统频响；

其中， $H(k, t)$ 为类混响系统频响； $X_{norm}(k, t)$ 为混响语音的归一化语谱图； $*$ 为线性卷积算子； $G(k, t)$ 为平滑核，且 $G(k, t) = \beta \cdot e^{-\frac{k^2+t^2}{2 \cdot \alpha^2}}$ ， $\iint G(k, t) dt dk = 1$ ， β 为归一化系数常数，且 β 使得 $e^{-\frac{k^2+t^2}{2 \cdot \alpha^2}}$ 的全积分等于 1； α 为滤波半径，且 α 越大，图像越锐化； k 为离散频率的索引； t 为时域信号帧索引。

15. 根据权利要求 12 所述的语音去混响的装置，其中，所述第四获取模块，用于：

根据公式：

$$\hat{H}(k, t) = \psi\{H(k, t)\} = \begin{cases} \hat{H}_{min}, & \text{若 } H(k, t) = H_{max} \\ \hat{H}_{max}, & \text{若 } H(k, t) = H_{min} \\ \Phi\{H(k, t)\}, & \text{若 } H_{min} < H(k, t) < H_{max} \end{cases}, \text{ 获取混}$$

响系统频响估计;

其中, $\hat{H}(k, t)$ 为混响系统频响估计; $H(k, t)$ 为类混响系统频响; $\hat{H}_{min}$ 为 $\hat{H}(k, t)$ 的最小值; $\hat{H}_{max}$ 为 $\hat{H}(k, t)$ 的最大值; H_{min} 为 $H(k, t)$ 的最小值; H_{max} 为 $H(k, t)$ 的最大值; $\Phi\{\cdot\}$ 为递减函数; k 为离散频率的索引; t 为时域信号帧索引。

16. 根据权利要求 12 所述的语音去混响的装置, 其中, 所述第五获取模块, 用于:

根据公式: $\hat{S}(k, t) = X(k, t) / \hat{H}(k, t)$, 获取去混响语音信号的语谱;

其中, $\hat{S}(k, t)$ 为去混响语音信号的语谱; $X(k, t)$ 为混响语音的语谱信号; $\hat{H}(k, t)$ 为混响系统频响估计; k 为离散频率的索引; t 为时域信号帧索引。

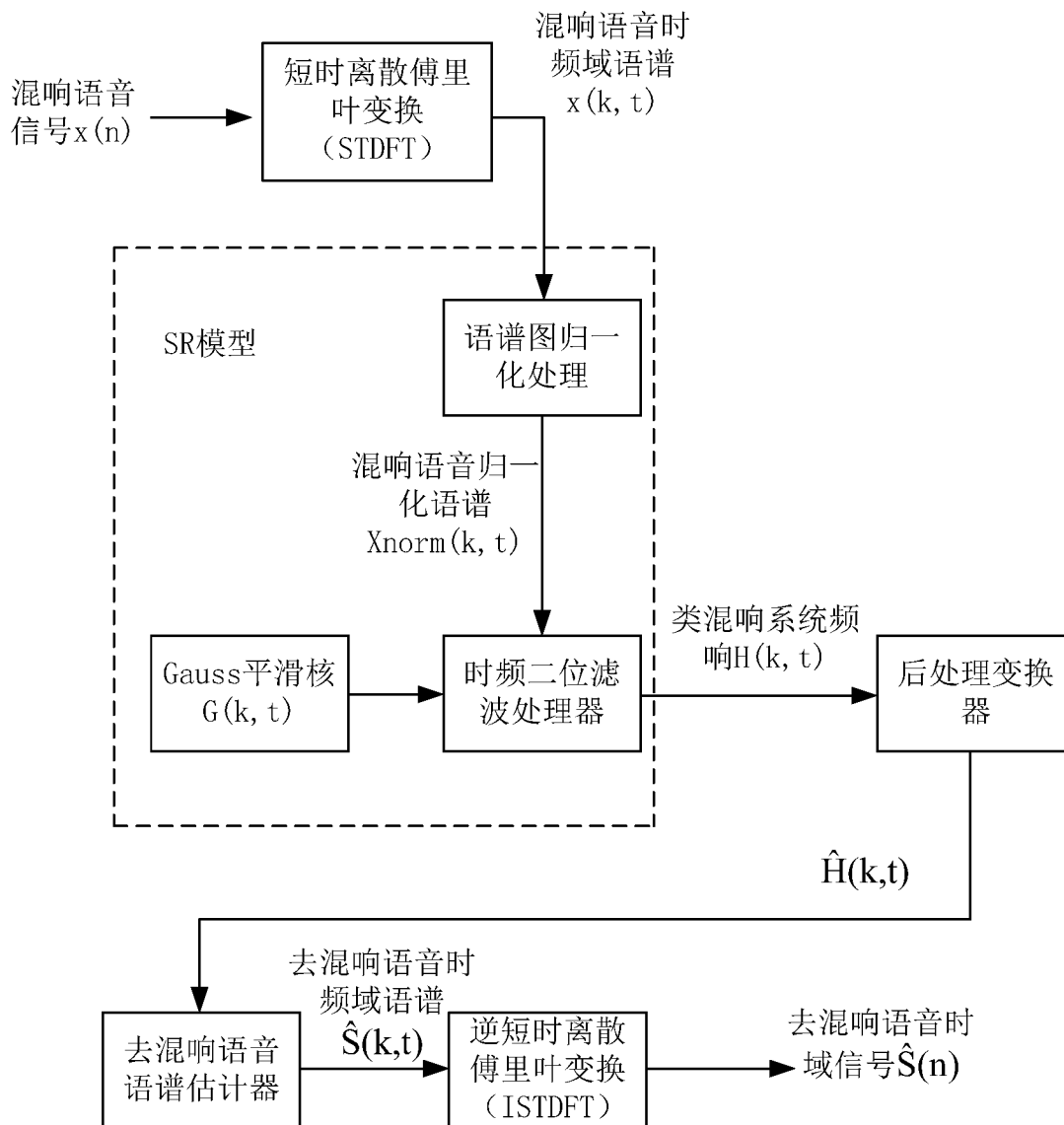


图 1

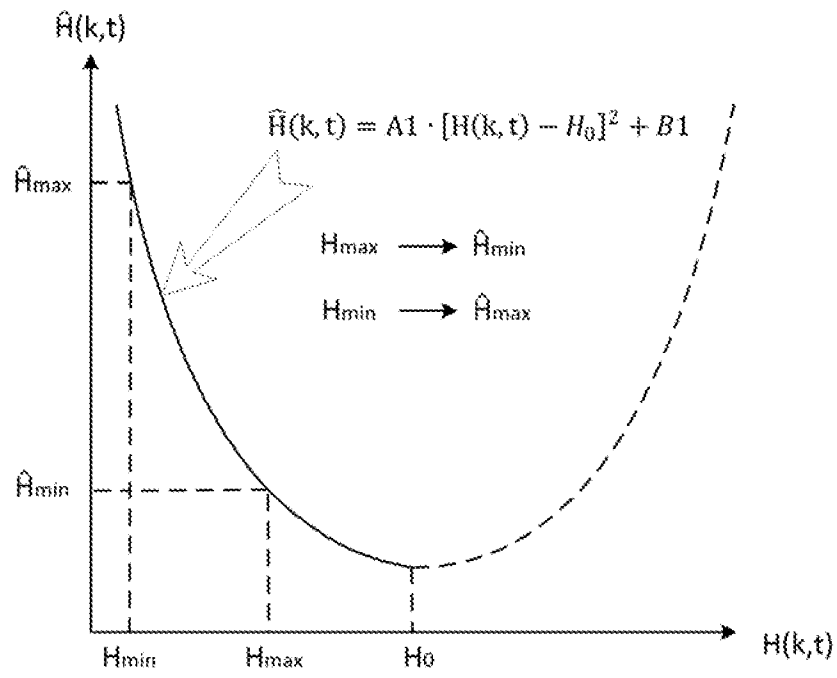


图 2

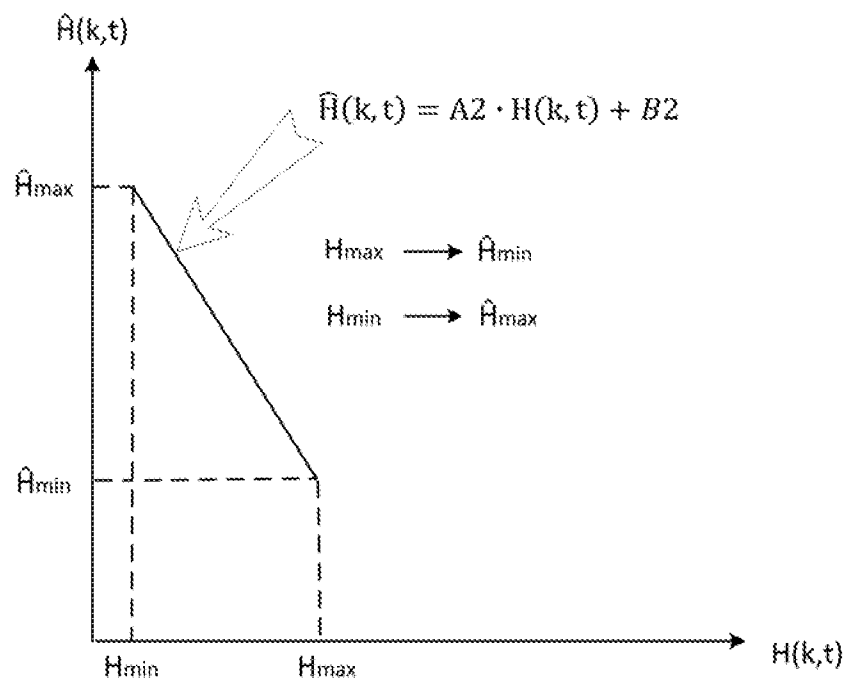


图 3

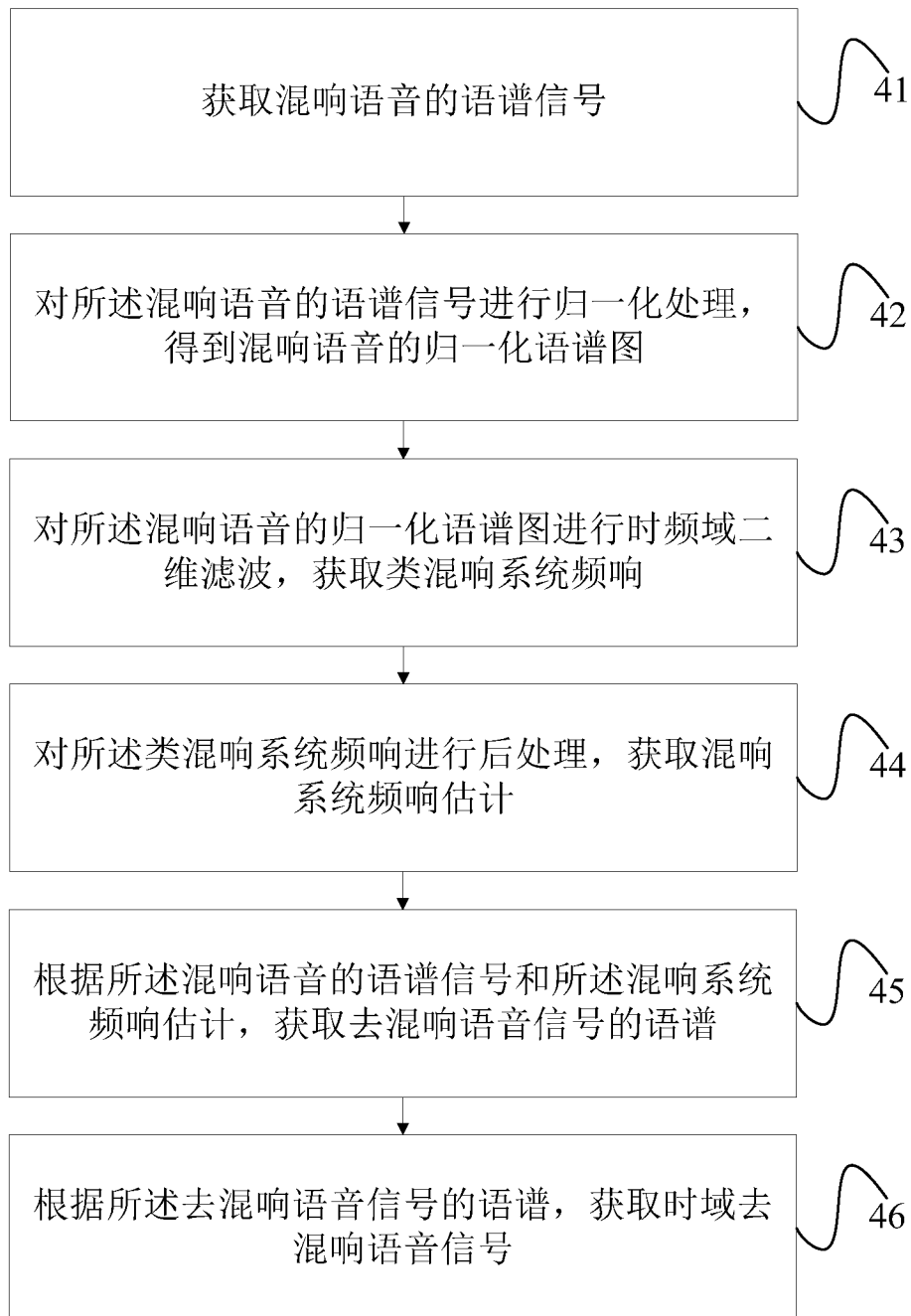


图 4

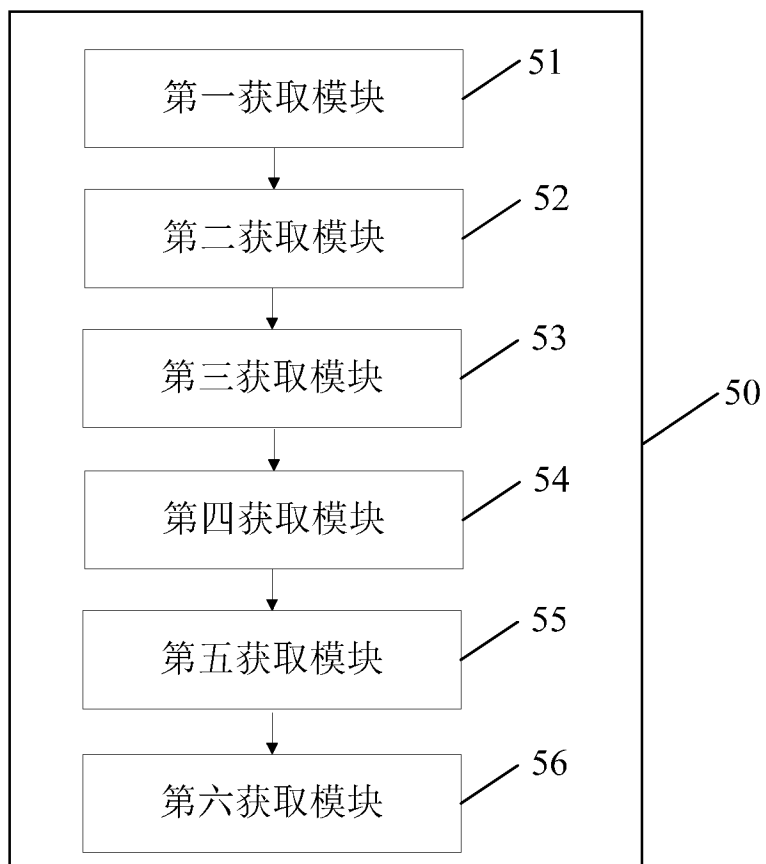


图 5

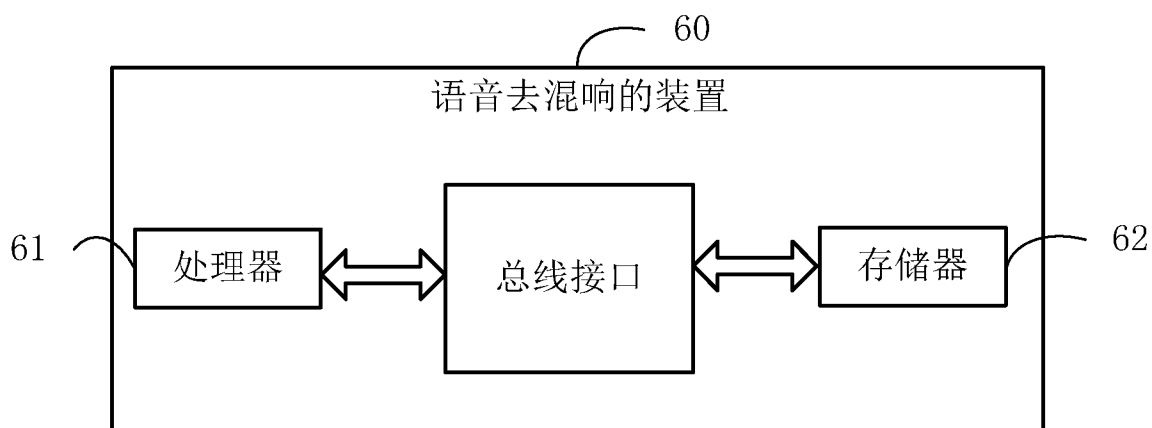


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070922

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/0208(2013.01)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)

G10L 21

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; DWPI; CNKI; 百度学术; BAIDU: 去混响, 消除, 混响, 语谱图, 二维滤波, dereverberation, cancel, reverb, spectrogram, speech, audio, enhancement, surround retinex

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109637553 A (TELECOMMUNICATIONS SCIENCE AND TECHNOLOGY RESEARCH INSTITUTE CO., LTD.) 16 April 2019 (2019-04-16) claims 1-16	1-16
X	ZHANG, Mingming et al. "Frequency-domain Dereverberation on Speech Signal using Surround Retinex" <i>2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference</i> , 01 November 2013 (2013-11-01), sections 2 and 3, and figure 5	1-16
A	肖纯智 等 (XIAO, Chunzhi et al.). "一种基于语谱图分析的语音增强算法 (A Speech Enhancement Algorithm Based on Speech Spectrogram)" <i>电声技术 (Audio Engineering)</i> , Vol. 36, No. 9, 30 September 2012 (2012-09-30), pp. 44-48	1-16
A	CN 106485674 A (TIANJIN UNIVERSITY) 08 March 2017 (2017-03-08) entire document	1-16

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search

24 February 2020

Date of mailing of the international search report

28 February 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/070922

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109637553	A	16 April 2019	None	
CN	106485674	A	08 March 2017	None	

国际检索报告

国际申请号

PCT/CN2020/070922

A. 主题的分类

G10L 21/0208(2013.01)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)

G10L 21

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;DWPI;CNKI;百度学术: 去混响, 消除, 混响, 语谱图, 二维滤波, dereverberation, cancel, reverb, spectrogram, speech, audio, enhancement, surround retinex

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
PX	CN 109637553 A (电信科学技术研究院有限公司) 2019年 4月 16日 (2019 - 04 - 16) 权利要求1-16	1-16
X	Mingming Zhang 等. "Frequency-domain Dereverberation on Speech Signal using Surround Retinex" 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2013年 11月 1日 (2013 - 11 - 01), 第2-3部分, 图5	1-16
A	肖纯智 等. "一种基于语谱图分析的语音增强算法" 电声技术, 第36卷, 第9期, 2012年 9月 30日 (2012 - 09 - 30), 第44-48页	1-16
A	CN 106485674 A (天津大学) 2017年 3月 8日 (2017 - 03 - 08) 全文	1-16

☐ 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

2020年 2月 24日

国际检索报告邮寄日期

2020年 2月 28日

ISA/CN的名称和邮寄地址

中国国家知识产权局(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

游晓梅

电话号码 (86-10)62089539

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/070922

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	109637553	A	2019年 4月 16日	无	
CN	106485674	A	2017年 3月 8日	无	