

(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 6 月 25 日 (25.06.2020)



(10) 国际公布号
WO 2020/125376 A1

(51) 国际专利分类号:
G10L 21/0216 (2013.01)

(21) 国际申请号: PCT/CN2019/121953

(22) 国际申请日: 2019 年 11 月 29 日 (29.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811548802.0 2018 年 12 月 18 日 (18.12.2018) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(72) 发明人: 纪璇(JI, Xuan); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。于蒙(YU, Meng); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 北京德琦知识产权代理有限公司 (DEQI INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区知春路 1 号学院国际大厦 7 层, Beijing 100083 (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, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

(54) Title: VOICE DENOISING METHOD AND APPARATUS, COMPUTING DEVICE AND COMPUTER READABLE STORAGE MEDIUM

(54) 发明名称: 语音降噪的方法和装置、计算设备和计算机可读存储介质

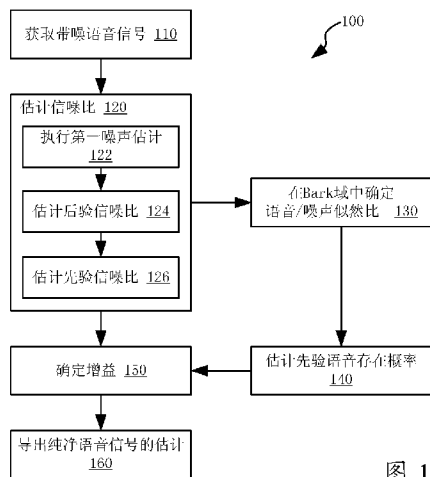


图 1B

110 Acquire a noise-bearing voice signal
120 Estimate a signal-to-noise ratio
122 Execute first noise estimation
124 Estimate a posterior signal-to-noise ratio
126 Estimate a priori signal-to-noise ratio
130 Determine a voice/noise likelihood ratio in a Bark domain
140 Estimate a priori voice presence probability
150 Determine a gain
160 Derive a pure voice signal estimate

(57) Abstract: A voice denoising method and apparatus. The method comprises: acquiring a noise-bearing voice signal, the noise-bearing voice signal comprising a pure voice signal and a noise signal (110); estimating a posterior signal-to-noise ratio and a priori signal-to-noise ratio of the noise-bearing voice signal (120); determining a voice/noise likelihood ratio in a Bark domain on the basis of the estimated posterior signal-to-noise ratio and the estimated prior signal-to-noise ratio (130); estimating a priori voice presence probability on the basis of the determined voice/noise likelihood ratio (140); determining a gain on the basis of the estimated posterior signal-to-noise ratio, the estimated prior signal-to-noise ratio and the estimated prior voice presence probability, the gain being an estimated frequency domain transfer function for transforming the noise-bearing voice signal into a pure voice signal (150); and deriving a pure voice signal estimate from the noise-bearing voice signal on the basis of the gain (160).

(57) 摘要: 一种语音降噪的方法和装置, 该方法包括: 获取带噪语音信号, 该带噪语音信号包括纯净语音信号和噪声信号 (110); 估计带噪语音信号的后验信噪比和先验信噪比 (120); 基于所估计的后验信噪比和所估计的先验信噪比在 Bark 域中确定语音/噪声似然比 (130); 基于所确定的语音/噪声似然比估计先验语音存在概率 (140); 基于所估计的后验信噪比、所估计的先验信噪比以及所估计的先验语音存在概率来确定增益, 该增益为用于将带噪语音信号变换成纯净语音信号的估计的频域传递函数 (150); 并且基于增益从带噪语音信号导出纯净语音信号估计 (160)。

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))。

语音降噪的方法和装置、计算设备和计算机可读存储介质

本申请要求于 2018 年 12 月 18 日提交中国专利局、申请号为 201811548802.0、发明名称为“语音降噪的方法和装置、计算设备和计算机可读存储介质”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
5 中。

技术领域

本申请涉及语音处理技术领域，具体来说涉及一种语音降噪方法、语音降噪装置、计算设备和计算机可读存储介质。

背景技术

10 在传统的语音降噪技术中通常存在两种处理方式。一种方式是在每个频点上都估计一个先验语音存在概率。在这种情况下，对于识别器而言在时间上和频率上的维纳增益波动越小，一般识别率越高；如果维纳增益波动比较大，反而会引入一些音乐噪声，可能导致识别率变差。另一种方式是使用全局的先验语音存在概率。这种方式比起前者而言在求
15 取维纳增益时更加鲁棒。然而，仅依赖全部频点上的先验信噪比来估计先验语音存在概率，可能不能很好地区分包含语音和噪声两者的帧和只含有噪声的帧。

发明内容

提供一种可以缓解、减轻或甚至消除上述问题中的一个或多个的机制将是有利的。
20

根据本申请的第一方面，提供了一种计算机实现的语音降噪方法，由计算设备执行，包括：获取带噪语音信号，所述带噪语音信号包括纯

净语音信号和噪声信号；估计所述带噪语音信号的后验信噪比和先验信噪比；基于所估计的后验信噪比和所估计的先验信噪比在 Bark 域中确定语音/噪声似然比；基于所确定的语音/噪声似然比估计先验语音存在概率；基于所估计的后验信噪比、所估计的先验信噪比以及所估计的先验语音存在概率来确定增益，所述增益为用于将所述带噪语音信号变换成所述纯净语音信号的估计的频域传递函数；并且基于所述增益从所述带噪语音信号导出所述纯净语音信号的所述估计。

根据本申请的另一方面，提供了一种语音降噪装置，包括：信号获取模块，被配置成获取带噪语音信号，所述带噪语音信号包括纯净语音信号和噪声信号；信噪比估计模块，被配置成估计所述带噪语音信号的先验信噪比和后验信噪比；似然比确定模块，被配置成基于所估计的先验信噪比和所估计的后验信噪比在 Bark 域中确定语音/噪声似然比；概率估计模块，被配置成基于所确定的语音/噪声似然比估计先验语音存在概率；增益确定模块，被配置成基于所估计的先验信噪比、所估计的后验信噪比以及所估计的先验语音存在概率来确定增益，所述增益为用于将所述带噪语音信号变换成所述纯净语音信号的估计的频域传递函数；以及语音信号导出模块，被配置成基于所述增益从所述带噪语音信号导出所述纯净语音信号的所述估计。

根据本申请的又另一方面，提供了一种计算设备，包括处理器和存储器，所述存储器被配置成存储计算机程序，所述计算机程序被配置成当在所述处理器上执行时使所述处理器执行如上所述的方法。

根据本申请的再另一方面，提供了一种计算机可读存储介质，被配置成存储计算机程序，所述计算机程序被配置成当在处理器上执行时使所述处理器执行如上所述的方法。

根据在下文中所描述的实施例，本申请的这些和其它方面将是清楚

明白的，并且将参考在下文中所描述的实施例而被阐明。

附图简要说明

在下面结合附图对于示例性实施例的描述中，本申请的更多细节、特征和优点被公开，在附图中：

5 图 1A 图示了根据本申请实施例的一种应用语音降噪方法的系统架构图。

图 1B 图示了根据本申请实施例的语音降噪方法的流程图；

图 2 更详细地图示了图 1B 的方法中执行第一噪声估计的步骤；

图 3 更详细地图示了图 1B 的方法中确定语音/噪声似然比的步骤；

10 图 4 更详细地图示了图 1B 的方法中估计先验语音存在概率的步骤；

图 5a、5b 和 5c 分别图示了一个示例的原始带噪语音信号、利用现有技术从该原始带噪语音信号导出的纯净语音信号的估计、以及利用图 1B 的方法从该原始带噪语音信号导出的纯净语音信号的估计的相应语谱图；

15 图 6 图示了根据本申请另一实施例的语音降噪方法的流程图；

图 7 图示了可以应用图 6 的方法的典型应用场景中的示例处理流程；

图 8 图示了根据本申请实施例的语音降噪装置的框图；并且

图 9 图示了根据本申请实施例的一个示例系统的结构图，该示例系统包括可以实现本文描述的各种技术的一个或多个系统和/或设备的示例计算设备。

20

实施方式

本申请的构思基于信号处理理论。设 $x(n)$ 和 $d(n)$ 分别表示纯净（即，无噪声）语音信号和不相关的加性噪声，则观察信号（下文中称为“带噪

语音信号”)可以表示为: $y(n)=x(n)+d(n)$ 。带噪语音信号 $y(n)$ 进行短时傅里叶变换得到频谱 $Y(k,l)$, 其中 k 表示频点, l 表示时间帧的序号。设 $X(k,l)$ 为纯净语音信号 $x(n)$ 的频谱, 那么通过估计增益 $G(k,l)$ 可以得到估计的纯净语音信号 $\hat{x}(n)$ 的频谱为 $\hat{X}(k,l)=G(k,l)*Y(k,l)$, 其中增益 $G(k,l)$ 为用于将带噪语音信号 $y(n)$ 变换成所述纯净语音信号 $x(n)$ 的估计的频域传递函数。然后, 通过逆短时傅里叶变换即可得到估计的纯净语音 $\hat{x}(n)$ 的时域信号。给出两个假设 $H_0(k,l)$ 和 $H_1(k,l)$, 分别表示语音不存在的事件

和语音存在的事件, 那么有如下表达式:
$$\begin{aligned} H_0(k,l): Y(k,l) &= D(k,l) \\ H_1(k,l): Y(k,l) &= X(k,l) + D(k,l) \end{aligned}$$

其中 $D(k,l)$ 表示噪声信号的短时傅里叶频谱。假设在频域中带噪语音信

号服从高斯分布: $p(Y(k,l)|H_0(k,l)) = \frac{1}{\pi\lambda_d(k,l)} \exp\{-\frac{|Y(k,l)|^2}{\lambda_d(k,l)}\}$ 和

$p(Y(k,l)|H_1(k,l)) = \frac{1}{\pi(\lambda_x(k,l) + \lambda_d(k,l))} \exp\{-\frac{|Y(k,l)|^2}{\lambda_x(k,l) + \lambda_d(k,l)}\}$, 根据该条件概率

分布和贝叶斯假设, 可以得到语音存在概率为:

$p(k,l) = \{1 + \frac{q(k,l)}{1 - q(k,l)} (1 + \xi(k,l)) \exp(-\nu(k,l))\}^{-1}$, 其中 $\xi(k,l) = \frac{\lambda_x(k,l)}{\lambda_d(k,l)}$,

$\gamma(k,l) = \frac{|Y(k,l)|^2}{\lambda_d(k,l)}$, $\nu(k,l) = \frac{\gamma(k,l)\xi(k,l)}{1 + \xi(k,l)}$, $\lambda_x(k,l)$ 为带噪语音信号 $y(n)$ 的第 l 帧

在第 k 个频点上的语音方差, 并且 $\lambda_d(k,l)$ 为第 l 帧在第 k 个频点上的噪声方差。 $\xi(k,l)$ 和 $\gamma(k,l)$ 分别表示第 l 帧在第 k 个频点上的先验信噪比和后验信噪比, $q(k,l)$ 是先验语音不存在概率, 并且 $1 - q(k,l)$ 即先验语音存在概率。我们使用 $\log$ 频谱幅度估计对纯净语音信号 $x(n)$ 的频谱幅度进行估计: $\hat{A}(k,l) = \exp\{E[\log A(k,l)|Y(k,l)]\}$, 并且基于高斯模型假设可以得到增

益 $G(k,l) = \{G_{H_1}(k,l)\}^{p(k,l)} G_{\min}^{1-p(k,l)}$, 其中 $G_{H_1}(k,l) = \frac{\xi(k,l)}{1 + \xi(k,l)} \exp(\frac{1}{2} \int_{\nu(k,l)}^{\infty} \frac{e^{-t}}{t} dt)$, 并且

$G_{\min}$ 是经验值, 其用于当语音不存在的时候限制增益 $G(k, l)$ 不低于某个阈值。求解增益 $G(k, l)$ 涉及到对先验信噪比 $\xi(k, l)$ 、噪声方差 $\lambda_d(k, l)$ 和先验语音不存在概率 $q(k, l)$ 进行估计。

图 1A 图示了根据本申请实施例的一种应用语音降噪方法的系统架构图。如图 1A 所示, 该系统架构包括计算设备 910, 以及用户终端集群。所述用户终端集群可以包括多个具有语音采集功能的用户终端, 包括用户终端 100a, 用户终端 100b 以及用户终端 100c。

如图 1A 所示, 所述用户终端 100a, 用户终端 100b 以及用户终端 100c 可以分别与所述计算设备 910 进行网络连接, 并且通过所述网络连接分别与所述计算设备 910 进行数据交互。

以用户终端 100a 为例, 所述用户终端 100a 通过网络发送带噪语音信号至所述计算设备 910, 所述计算设备 910 采用图 1B 所述的语音降噪方法 100, 或者图 6 所示的语音降噪方法 600, 从所述带噪语音信号导出纯净语音信号, 以供后续设备 (图中未示出) 进行语音识别。

图 1B 图示了根据本申请实施例的语音降噪方法 100 的流程图, 该方法可由图 9 所示的计算设备 910 执行。

在步骤 110 处, 获取带噪语音信号 $y(n) = x(n) + d(n)$ 。取决于应用场景, 带噪语音信号 $y(n)$ 的获取可以通过各种不同的方式实现。在一些实施例中, 它可以通过 I/O 接口, 例如麦克风, 从说话人直接获取。在一些实施例中, 它可以经由有线或无线网络或者移动通信网络从远程设备接收。在一些实施例中, 它还可以从本地存储器中缓冲或存储的语音数据记录中检索得到。所获取的带噪语音信号 $y(n)$ 经短时傅里叶变换被转换成频谱 $Y(k, l)$ 以供处理。

在步骤 120 处, 估计所述带噪语音信号 $y(n)$ 的后验信噪比 $\gamma(k, l)$ 和先

验信噪比 $\xi(k,l)$ 。在该实施例中，这可以通过如下所述的步骤 122~126 实现。

在步骤 122 处，执行第一噪声估计，其中得到所述噪声信号的方差 $\lambda_d(k,l)$ 的第一估计。图 2 更详细地图示了如何执行第一噪声估计。

5 参考图 2，在步骤 122a 处，对所述带噪语音信号 $y(n)$ 的能量谱在频

域进行平滑： $S_f(k,l) = \sum_{i=-w}^w W(i) |Y(k-l,i)|^2$ ，其中 $W(i)$ 是长度为 $2*w+1$ 的窗。

然后， $S_f(k,l)$ 进行时域平滑，得到 $S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s) S_f(k,l)$ ，其中 α_s 是平滑因子。在步骤 122b 处，对经平滑的所述能量谱 $S(k,l)$ 执行最小跟

踪估计。具体地，进行如下最小跟踪估计： $S_{\min}(k,l) = \min\{S_{\min}(k,l-1), S(k,l)\}$ ，
 $S_{\text{tmp}}(k,l) = \min\{S_{\text{tmp}}(k,l-1), S(k,l)\}$

10 其中 $S_{\min}$ 和 S_{tmp} 的初始值取为 $S(k,0)$ 。经过 L 帧之后，最小跟踪估计的表

达式在第 L+1 帧被更新为 $S_{\min}(k,l) = \min\{S_{\text{tmp}}(k,l-1), S(k,l)\}$ ，
 $S_{\text{tmp}}(k,l) = S(k,l)$ 。然后，对于从

第 L+2 帧到第 2L+1 帧的 L 个帧，最小跟踪估计的表达式恢复为

$S_{\min}(k,l) = \min\{S_{\min}(k,l-1), S(k,l)\}$ 。在第 2(L+1) 帧，最小跟踪估计的表达式
 $S_{\text{tmp}}(k,l) = \min\{S_{\text{tmp}}(k,l-1), S(k,l)\}$

再次被更新为 $S_{\min}(k,l) = \min\{S_{\text{tmp}}(k,l-1), S(k,l)\}$ ，
 $S_{\text{tmp}}(k,l) = S(k,l)$ 。然后，对于随后的 L 个帧，

15 最小跟踪估计的表达式再次恢复为 $S_{\min}(k,l) = \min\{S_{\min}(k,l-1), S(k,l)\}$ ，并且
 $S_{\text{tmp}}(k,l) = \min\{S_{\text{tmp}}(k,l-1), S(k,l)\}$

以此类推。也即，最小跟踪估计的表达式以 L+1 帧为周期被周期性地更新。在步骤 122c 处，取决于经平滑的所述能量谱 $S(k,l)$ 与该经平滑的所

述能量谱的最小跟踪估计 $S_{\min}(k,l)$ 的比值，即 $S_r(k,l) = \frac{S(k,l)}{S_{\min}(k,l)}$ ，利用所述

带噪语音信号 $y(n)$ 的上一帧中的所述噪声信号的方差 $\lambda_d(k,l-1)$ 的所述第

一估计和所述带噪语音信号 $y(n)$ 的当前帧的所述能量谱 $|Y(k,l)|^2$ 来选择性地更新所述当前帧中的所述噪声信号的方差 $\lambda_d(k,l)$ 的所述第一估计。具体地, 如果比值 $\frac{|Y(k,l)|^2}{\lambda_d(k,l)}$ 大于或等于第一阈值就执行更新, 并且如果比值 $\frac{|Y(k,l)|^2}{\lambda_d(k,l)}$ 小于该第一阈值就不更新。噪声估计更新公式为:

- 5 $\hat{\lambda}_d(k,l) = \alpha_d \hat{\lambda}_d(k,l-1) + (1-\alpha_d) |Y(k,l)|^2$, 其中 α_d 是平滑因子。在工程实践中, 所获取的带噪语音信号 $y(n)$ 的起始的若干帧可以被估计为噪声信号的初始值。

返回参考图 1B, 在步骤 124 处, 利用所述噪声信号的方差 $\lambda_d(k,l)$ 的所述第一估计来估计所述后验信噪比 $\gamma(k,l)$ 。在步骤 122 中得到估计的噪声信号的方差 $\hat{\lambda}_d(k,l)$ 之后, 后验信噪比 $\gamma(k,l)$ 的估计可以计算为

$$\hat{\gamma}(k,l) = \frac{|Y(k,l)|^2}{\hat{\lambda}_d(k,l)}.$$

在步骤 126 处, 利用所估计的后验信噪比 $\hat{\gamma}(k,l)$ 来估计所述先验信噪比 $\xi(k,l)$ 。在该实施例中, 先验信噪比估计可以使用面向判决的 (decision-directed, DD) 估计:

- 15 $\hat{\xi}(k,l) = \alpha G_{H_1}^2(k,l-1) \hat{\gamma}(k,l-1) + (1-\alpha) \max\{\hat{\gamma}(k,l)-1, 0\}$ 。其中 $G_{H_1}^2(k,l-1) \hat{\gamma}(k,l-1)$ 表示上一帧的先验信噪比的估计, $\max\{\hat{\gamma}(k,l)-1, 0\}$ 是基于当前帧对先验信噪比的最大似然估计, 并且 α 是这两种估计的平滑因子。由此, 得到估计的先验信噪比 $\hat{\xi}(k,l)$ 。

在步骤 130 处, 基于所估计的后验信噪比 $\hat{\gamma}(k,l)$ 和所估计的先验信噪比 $\hat{\xi}(k,l)$ 在 Bark 域中确定语音/噪声似然比。似然比公式为

$$\Delta(k,l) = \frac{P(Y(k,l)|H_1(k,l))}{P(Y(k,l)|H_0(k,l))}, \text{ 其中 } Y(k,l) \text{ 为第 } l \text{ 帧在第 } k \text{ 个频点上的幅度谱,}$$

$H_1(k, l)$ 为第 l 帧在第 k 个频点假设是语音的状态, $H_0(k, l)$ 为第 l 帧在第 k 个频点假设是噪声的状态, $P(Y(k, l) | H_1(k, l))$ 为在语音存在的情况下的概率密度, 并且 $P(Y(k, l) | H_0(k, l))$ 为在噪声存在的情况下的概率密度。图 3 更详细地图示了如何确定语音/噪声似然比。

5 参考图 3, 在步骤 132 处, 对概率密度做高斯概率密度函数 (PDF)

假设, 似然比公式可变成: $\Delta(k, l) = \frac{P(Y(k, l) | H_1(k, l))}{P(Y(k, l) | H_0(k, l))} = \frac{\exp(\frac{\xi(k, l)\gamma(k, l)}{(1 + \xi(k, l))})}{(1 + \xi(k, l))}$ 。在

步骤 134 处, 将先验信噪比 $\xi(k, l)$ 和后验信噪比 $\gamma(k, l)$ 从线性频域转换到 Bark 域。Bark 域是使用听觉滤波器模拟出的听觉的 24 个临界频带, 并且因此具有 24 个频点。存在多种方式从线性频域转换到 Bark 域。在该

10 实施例中, 该转换可以基于以下等式:

$b = 13 * \arctan(0.76 * f_{kHz}) + 3.5 * \arctan(\frac{f_{kHz}}{7.5})^2$, 其中 f_{kHz} 为所述线性频域中的频

率, 并且 b 表示为 Bark 域中的 24 个频点。由此, 在 Bark 域上的似然比

公式可表达为 $\Delta(b, l) = \frac{\exp(\frac{\xi(b, l)\gamma(b, l)}{(1 + \xi(b, l))})}{(1 + \xi(b, l))}$ 。

返回参考图 1B, 在步骤 140 处, 基于所确定的语音/噪声似然比估计先验语音存在概率, 图 1B 所示的方法可以提升判断语音是否出现的准确率, 避免多次判断语音是否出现, 提高了资源利用率。图 4 更详细地图示了如何估计先验语音存在概率。

参考图 4, 在步骤 142 处, 在对数域中将 $\Delta(b, l)$ 平滑为 $\log(\Delta(b, l)) = \beta * \log(\Delta(b, l - 1)) + (1 - \beta) * \log(\Delta(b, l))$, 其中 β 为平滑因子。在步骤 144 处, 通过在 Bark 域的全带中映射 $\log(\Delta(b, l))$ 而得到所估计的先验语音存在概率 $P_{frame}(l)$ 。在该实施例中, 函数 $\tanh$ 可以

被用于所述映射, 得到 $P_{frame}(l) = \tanh(\frac{1}{24} \sum_{b=1}^{24} \log(\Delta(b, l)))$, 其中 $P_{frame}(l)$ 为所估计的先验语音存在概率, 也即具体实施方式的开头段落中提到的先验语音存在概率 $1-q(k, l)$ 的估计。在该实施例中函数 $\tanh$ 被使用是因为它能够将区间 $[0, +\infty)$ 映射为 0-1 的区间, 尽管其他实施例是可能的。

- 5 与现有技术的语音降噪方案相比, 方法 100 可以提升判断语音是否出现的准确率。这是因为 (1) 语音/噪声似然比能很好地区分有语音出现的状态和没有语音出现的状态, 并且 (2) Bark 域相比于线性频域更符合人耳的听觉掩蔽效应。Bark 域具有对低频的放大作用和对高频的压缩作用, 能更清晰地揭示哪些信号容易产生掩蔽和哪些噪声比较明显。
- 10 因此, 方法 100 可以提升判断语音是否出现的准确率, 从而得到更准确的先验语音存在概率。

- 返回参考图 1B, 在步骤 150 处, 基于在步骤 124 中得到的所估计的后验信噪比 $\hat{\gamma}(k, l)$ 、在步骤 126 中得到的所估计的先验信噪比 $\hat{\xi}(k, l)$ 以及在步骤 140 中得到的所估计的先验语音存在概率 $P_{frame}(l)$ 来确定增益
- 15 $G(k, l)$ 。这可以通过具体实施方式的开头段落中提到的以下等式来实现:

$$G(k, l) = \{G_{H_1}(k, l)\}^{p(k, l)} G_{\min}^{1-p(k, l)}, \text{ 其中 } G_{H_1}(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \exp\left(\frac{1}{2} \int_{\nu(k, l)}^{\infty} \frac{e^{-t}}{t} dt\right), \text{ 以}$$

及

$$p(k, l) = \left\{1 + \frac{q(k, l)}{1 - q(k, l)} (1 + \xi(k, l)) * \exp(-\nu(k, l))\right\}^{-1}, \text{ 其中 } \nu(k, l) = \frac{\gamma(k, l)\xi(k, l)}{1 + \xi(k, l)}.$$

- 在步骤 160 处, 基于增益 $G(k, l)$ 从所述带噪语音信号 $y(n)$ 导出所述纯净语音信号 $x(n)$ 的所述估计 $\hat{x}(n)$ 。具体地, 通过 $\hat{X}(k, l) = G(k, l) * Y(k, l)$ 可以得到估计的纯净语音信号 $\hat{x}(n)$ 的频谱, 并且然后通过逆短时傅里叶变换即可得到估计的纯净语音 $\hat{x}(n)$ 的时域信号。
- 20

图 5a、5b 和 5c 分别图示了一个示例的原始带噪语音信号、利用现有技术从该原始带噪语音信号导出的纯净语音信号的估计、以及利用方法 100 从该原始带噪语音信号导出的纯净语音信号的估计的相应语谱图。从这些图可以看出，在只有噪声存在的情况下，与在图 5b 中相比，噪声在图 5c 中被进一步抑制，而语音基本不变。这表明了方法 100 在估计语音是否存在方面的更好的表现以及在只有噪声的情况下对噪声的进一步抑制。这有利地增强了从带噪语音信号恢复出来的语音信号的质量。

图 6 图示了根据本申请另一实施例的语音降噪方法 600 的流程图，可以由图 9 所示的计算设备 910 执行。

参考图 6，与方法 100 类似，方法 600 也包括步骤 110~160，这些步骤的详情已经在上面关于图 1B-4 进行了描述并且因此在此被省略。方法 600 不同于方法 100 在于它还包括步骤 610 和 620，它们在下面被详细描述。

在步骤 610 处，执行第二噪声估计，其中得到所述噪声信号的方差 $\lambda_d(k, l)$ 的第二估计。第二噪声估计是独立于（并行于）第一噪声估计而被执行的，并且可以采用与步骤 122 中相同的噪声估计更新公式：

$\hat{\lambda}_d(k, l) = \alpha_d \hat{\lambda}_d(k, l-1) + (1 - \alpha_d) |Y(k, l)|^2$ 。然而，在第二噪声估计中采用不同于第一噪声估计的更新准则。具体地，在步骤 610 中，取决于步骤 140

中得到的所估计的先验语音存在概率 $P_{frame}(l)$ ，利用所述带噪语音信号 $y^{(n)}$ 的上一帧中的所述噪声信号的方差 $\lambda_d(k, l-1)$ 的所述第二估计和所述带噪语音信号 $y^{(n)}$ 的当前帧的能量谱 $|Y(k, l)|^2$ 来选择性地更新所述当前帧中的所述噪声信号的方差 $\lambda_d(k, l)$ 的所述第二估计。更具体地，如果所估计的先验语音存在概率 $P_{frame}(l)$ 大于或等于第二阈值 $spthr$ ，则执行所述更新，

并且如果所估计的先验语音存在概率 $P_{frame}(l)$ 小于所述第二阈值 $spthr$, 则不执行所述更新。

在步骤 620 处, 取决于所述噪声信号的方差 $\lambda_d(k,l)$ 的所述第一估计在预定频率范围内的量值之和, 利用所述噪声信号的方差 $\lambda_d(k,l)$ 的所述第二估计来选择性地重新估计所述后验信噪比 $\gamma(k,l)$ 和所述先验信噪比 $\xi(k,l)$ 。在一些实施例中所述预定频率范围可以例如为低频范围, 诸如 0 至 1 kHz, 尽管其他实施例是可能的。所述噪声信号的方差 $\lambda_d(k,l)$ 的所述第一估计在该预定频率范围内的量值之和可以指示噪声信号的预定频率分量的水平。在实施例中, 如果所述量值之和大于或等于第三阈值 $noithr$, 则执行所述重新估计, 并且如果所述量值之和小于所述第三阈值 $noithr$, 则不执行所述重新估计。后验信噪比 $\gamma(k,l)$ 和先验信噪比 $\xi(k,l)$ 的重新估计可以基于上面描述的步骤 124 和 126 中的操作, 只不过在步骤 610 的第二噪声估计中 (而不是在步骤 122 的第一噪声估计中) 得到的噪声方差的估计被使用。

在所述重新估计被执行的情况下, 在步骤 150 中基于所重新估计的后验信噪比 (而不是在步骤 124 中得到的后验信噪比)、所重新估计的先验信噪比 (而不是在步骤 126 中得到的先验信噪比) 以及在步骤 140 中得到的所估计的先验语音存在概率来确定增益 $G(k,l)$ 。在所述重新估计未被执行的情况下, 在步骤 150 中仍然基于在步骤 124 中得到的后验信噪比、在步骤 126 中得到的先验信噪比以及在步骤 140 中得到的所估计的先验语音存在概率来确定增益 $G(k,l)$ 。

方法 600 与直接地使用第二噪声估计来重新估计先验信噪比 $\xi(k,l)$ 和后验信噪比 $\gamma(k,l)$ (以及因此维纳增益 $G(k,l)$) 的方案相比能够导致在低信噪比情况下识别率的提升, 因为第二噪声估计可能导致噪声的过估

计，该过估计虽然在低信噪比情况下能进一步抑制噪声，但是在高信噪比的情况下可能损失语音信息。由于引入了噪声估计的判决，其中根据判决结果选择性地使用第一噪声估计或第二噪声估计来求维纳增益，方法 600 能确保在高低信噪比下都有比较好的性能表现。

5 图 7 图示了其中可以应用图 6 的方法 600 的典型应用场景中的示例处理流程 700。该典型应用场景例如为车载终端与用户之间的人机对话。在 710 处，对来自用户的语音输入进行回波抵消。语音输入可以是例如通过多个信号采集通道采集的带噪语音信号。回波抵消可以基于例如自动回波抵消（AEC）技术来实现。在 720 处，进行波束形成。通过对多
10 个信号采集通道采集的各路信号进行加权合成，形成所需的语音信号。在 730 处，对语音信号进行降噪。这可以通过图 6 的方法 600 来实现。在 740 处，基于经降噪的语音信号确定是否唤醒车载终端上安装的语音应用程序。例如，只有在经降噪的语音信号被识别为特定的语音口令（例如，“你好！XXX”）时，语音应用程序才被唤醒。语音口令的识别可以
15 通过车载终端上的本地语音识别软件来实现。如果语音应用程序未被唤醒，则继续接收和识别语音信号，直到所要求的语音口令被输入。如果语音应用程序被唤醒，则在 750 处触发云端语音识别功能，并且经降噪的语音信号被车载终端发送到云端进行识别。在识别来自车载终端的语音信号之后，云端可以将相应的语音应答内容回送给车载终端，从而实现
20 人机对话。在一种实施方式中，语音信号的识别和应答可以在车载终端本地执行。

图 8 图示了根据本申请实施例的语音降噪装置 800 的框图。参考图 8，语音降噪装置 800 包括信号获取模块 810、信噪比估计模块 820、似然比确定模块 830、概率估计模块 840、增益确定模块 850 以及语音信
25 号导出模块 860。

信号获取模块 810 被配置成获取带噪语音信号 $y(n)$ 。取决于应用场景,信号获取模块 810 可以通过各种不同的方式实现。在一些实施例中,它可以是诸如麦克风之类的语音拾取设备或者其他以硬件实现的接收机。在一些实施例中,它可以被实现为计算机指令,以例如从本地存储器中检索语音数据记录。在一些实施例中,它可以被实现为硬件和软件的组合。带噪语音信号 $y(n)$ 的获取涉及上面关于图 1B 描述的步骤 110 中的操作,并且在此不再赘述。

信噪比估计模块 820 被配置成估计所述带噪语音信号 $y(n)$ 的后验信噪比 $\gamma(k,l)$ 和先验信噪比 $\xi(k,l)$ 。这涉及在上面关于图 1B 和 2 描述的步骤 120 中的操作,并且在此不再赘述。在一些实施例中,信噪比估计模块 820 还可以被配置成执行上面关于图 6 描述的步骤 610 和 620 中的操作。具体地,信噪比估计模块 820 还可以被配置成 (1) 执行第二噪声估计,其中得到所述噪声信号的方差 $\lambda_d(k,l)$ 的第二估计,和 (2) 取决于所述噪声信号的方差 $\lambda_d(k,l)$ 的所述第一估计在预定频率范围内的量值之和,利用所述噪声信号的方差 $\lambda_d(k,l)$ 的所述第二估计来选择性地重新估计所述后验信噪比 $\gamma(k,l)$ 和所述先验信噪比 $\xi(k,l)$ 。

似然比确定模块 830 被配置成基于所估计的后验信噪比 $\hat{\gamma}(k,l)$ 和所估计的先验信噪比 $\hat{\xi}(k,l)$ 在 Bark 域中确定语音/噪声似然比。这涉及在上面关于图 1B 和 3 描述的步骤 130 中的操作,并且在此不再赘述。

概率估计模块 840 被配置成基于所确定的语音/噪声似然比估计先验语音存在概率。这涉及在上面关于图 1B 和 4 描述的步骤 140 中的操作,并且在此不再赘述。

增益确定模块 850 被配置成基于所估计的后验信噪比 $\hat{\gamma}(k,l)$ 、所估计

的先验信噪比 $\hat{\xi}(k, l)$ 以及所估计的先验语音存在概率 $P_{frame}(l)$ 来确定增益 $G(k, l)$ 。这涉及在上面关于图 1B 描述的步骤 150 中的操作，并且在此不再赘述。在后验信噪比和先验信噪比的重估已经由信噪比估计模块 820 执行的实施例中，增益确定模块 850 还被配置成基于所重新估计的后验信噪比、所重新估计的先验信噪比以及所估计的先验语音存在概率 $P_{frame}(l)$ 来确定增益 $G(k, l)$ 。

语音信号导出模块 860 被配置成基于增益 $G(k, l)$ 从所述带噪语音信号 $y(n)$ 导出所述纯净语音信号 $x(n)$ 的所述估计 $\hat{x}(n)$ 。这涉及在上面关于图 1B 描述的步骤 160 中的操作，并且在此不再赘述。

图 9 图示了根据本申请实施例的示例系统 900 的结构图，该系统 900 包括可以实现本文描述的各种技术的一个或多个系统和/或设备的示例计算设备 910。计算设备 910 可以是例如服务提供商的服务器设备、与客户端（例如，客户端设备）相关联的设备、片上系统、和/或任何其它合适的计算设备或计算系统。上面关于图 8 描述的语音降噪装置 800 可以采取计算设备 910 的形式。在一种实施方式中，语音降噪装置 800 可以以语音降噪应用 916 的形式被实现为计算机程序。

如图示的示例计算设备 910 包括彼此通信耦合的处理系统 911、一个或多个计算机可读介质 912 以及一个或多个 I/O 接口 913。尽管未示出，但是计算设备 910 还可以包括系统总线或其他数据和命令传送系统，其将各种组件彼此耦合。系统总线可以包括不同总线结构的任何一个或组合，所述总线结构诸如存储器总线或存储器控制器、外围总线、通用串行总线、和/或利用各种总线架构中的任何一种的处理器或局部总线。还构思了各种其他示例，诸如控制和数据线。

处理系统 911 代表使用硬件执行一个或多个操作的功能。因此，处

理系统 911 被图示为包括可被配置为处理器、功能块等的硬件元件 914。这可以包括在硬件中实现作为专用集成电路或使用一个或多个半导体形成的其它逻辑器件。硬件元件 914 不受其形成的材料或其中采用的处理机构的限制。例如，处理器可以由（多个）半导体和/或晶体管（例如，
5 电子集成电路（IC））组成。在这样的上下文中，处理器可执行指令可以是电子可执行指令。

计算机可读介质 912 被图示为包括存储器/存储装置 915。存储器/存储装置 915 表示与一个或多个计算机可读介质相关联的存储器/存储容量。存储器/存储装置 915 可以包括易失性介质（诸如随机存取存储器
10 （RAM））和/或非易失性介质（诸如只读存储器（ROM）、闪存、光盘、磁盘等）。存储器/存储装置 915 可以包括固定介质（例如，RAM、ROM、固定硬盘驱动器等）以及可移动介质（例如，闪存、可移动硬盘驱动器、光盘等）。计算机可读介质 912 可以以下面进一步描述的各种其他方式进行配置。

15 一个或多个 I/O 接口 913 代表允许用户向计算设备 910 输入命令和信息并且还允许使用各种输入/输出设备将信息呈现给用户和/或其他组件或设备的功能。输入设备的示例包括键盘、光标控制设备（例如，鼠标）、麦克风（例如，用于语音输入）、扫描仪、触摸功能（例如，被配置为检测物理触摸的容性或其他传感器）、相机（例如，可以采用可见
20 或不可见的波长（诸如红外频率）将不涉及触摸的运动检测为手势）等等。输出设备的示例包括显示设备（例如，监视器或投影仪）、扬声器、打印机、网卡、触觉响应设备等。因此，计算设备 910 可以以下面进一步描述的各种方式进行配置以支持用户交互。

计算设备 910 还包括语音降噪应用 916。语音降噪应用 916 可以例如
25 如是图 8 的语音降噪装置 800 的软件实例，并且与计算设备 910 中的其

他元件相组合地实现本文描述的技术。

本文可以在软件硬件元件或程序模块的一般上下文中描述各种技术。一般地，这些模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、元素、组件、数据结构等。本文所使用的术语“模块”，
5 “功能”和“组件”一般表示软件、固件、硬件或其组合。本文描述的技术的特征是与平台无关的，意味着这些技术可以在具有各种处理器的各种计算平台上实现。

所描述的模块和技术的实现可以存储在某种形式的计算机可读介质上或者跨某种形式的计算机可读介质传输。计算机可读介质可以包括
10 可由计算设备 910 访问的各种介质。作为示例而非限制，计算机可读介质可以包括“计算机可读存储介质”和“计算机可读信号介质”。

与单纯的信号传输、载波或信号本身相反，“计算机可读存储介质”是指能够持久存储信息的介质和/或设备，和/或有形的存储装置。因此，计算机可读存储介质是指非信号承载介质。计算机可读存储介质包括诸如
15 如易失性和非易失性、可移动和不可移动介质和/或以适用于存储信息（诸如计算机可读指令、数据结构、程序模块、逻辑元件/电路或其他数据）的方法或技术实现的存储设备之类的硬件。计算机可读存储介质的示例可以包括但不限于 RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字通用盘（DVD）或其他光学存储装置、硬盘、盒式
20 磁带、磁带，磁盘存储装置或其他磁存储设备，或其他存储设备、有形介质或适于存储期望信息并可以由计算机访问的制品。

“计算机可读信号介质”是指被配置为诸如经由网络将指令发送到计算设备 910 的硬件的信号承载介质。信号介质典型地可以将计算机可读指令、数据结构、程序模块或其他数据体现在诸如载波、数据信号或其它
25 它传输机制的调制数据信号中。信号介质还包括任何信息传递介质。术

语“调制数据信号”是指以这样的方式对信号中的信息进行编码来设置或改变其特征中的一个或多个的信号。作为示例而非限制，通信介质包括诸如有线网络或直接连线的有线介质以及诸如声、RF、红外和其它无线介质的无线介质。

- 5 如前所述，硬件元件 914 和计算机可读介质 912 代表以硬件形式实现的指令、模块、可编程器件逻辑和/或固定器件逻辑，其在一些实施例中可以用于实现本文描述的技术的至少一些方面。硬件元件可以包括集成电路或片上系统、专用集成电路(ASIC)、现场可编程门阵列(FPGA)、复杂可编程逻辑器件(CPLD)以及硅中的其它实现或其他硬件设备的
- 10 组件。在这种上下文中，硬件元件可以作为执行由硬件元件所体现的指令、模块和/或逻辑所定义的程序任务的处理设备，以及用于存储用于执行的指令的硬件设备，例如，先前描述的计算机可读存储介质。

 前述的组合也可以用于实现本文所述的各种技术和模块。因此，可以将软件、硬件或程序模块和其它程序模块实现为在某种形式的计算机

15 可读存储介质上和/或由一个或多个硬件元件 914 体现的一个或多个指令和/或逻辑。计算设备 910 可以被配置为实现与软件和/或硬件模块相对应的特定指令和/或功能。因此，例如通过使用处理系统的计算机可读存储介质和/或硬件元件 914，可以至少部分地以硬件来实现将模块实现为可由计算设备 910 作为软件执行的模块。指令和/或功能可以由一个或

20 多个制品(例如，一个或多个计算设备 910 和/或处理系统 911)可执行/可操作以实现本文所述的技术、模块和示例。

 在各种实施方式中，计算设备 910 可以采用各种不同的配置。例如，计算设备 910 可以被实现为包括个人计算机、台式计算机、多屏幕计算机、膝上型计算机、上网本等的计算机类设备。计算设备 910 还可以被

25 实现为包括诸如移动电话、便携式音乐播放器、便携式游戏设备、平板

计算机、多屏幕计算机等移动设备的移动装置类设备。计算设备 910 还可以实现为电视类设备，其包括具有或连接到休闲观看环境中的一般地较大屏幕的设备。这些设备包括电视、机顶盒、游戏机等。

本文描述的技术可以由计算设备 910 的这些各种配置来支持，并且
5 不限于本文所描述的技术的具体示例。功能还可以通过使用分布式系统、诸如通过如下所述的平台 922 而在“云”920 上全部或部分地实现。

云 920 包括和/或代表用于资源 924 的平台 922。平台 922 抽象云 920 的硬件（例如，服务器设备）和软件资源的底层功能。资源 924 可以包括在远离计算设备 910 的服务器设备上执行计算机处理时可以使用的应
10 用和/或数据。资源 924 还可以包括通过因特网和/或通过诸如蜂窝或 Wi-Fi 网络的订户网络提供的服务。

平台 922 可以抽象资源和功能以将计算设备 910 与其他计算设备连接。平台 922 还可以用于抽象资源的分级以提供遇到的对于经由平台 922 实现的资源 924 的需求的相应水平的分级。因此，在互连设备实施例
15 中，本文描述的功能的实现可以分布在整个系统 900 内。例如，功能可以部分地在计算设备 910 上以及通过抽象云 920 的功能的平台 922 来实现。在一些实施例中，计算设备 910 可以将导出的纯净语音信号发送到驻留在云 920 上的语音识别应用（未示出）以供识别。在一种实施方式中，计算设备 910 也可以包括本地的语音识别应用（未示出）。

20 在本文的讨论中，描述了各种不同的实施例。应当领会和理解，本文描述的每个实施例可以单独使用或与本文所述的一个或多个其他实施例相关联地使用。

尽管已经以结构特征和/或方法动作特定的语言描述了主题，但是应当理解，所附权利要求中限定的主题不一定限于上述具体特征或动作。
25 相反，上述具体特征和动作被公开为实现权利要求的示例形式。虽然各

个操作在附图中被描绘为按照特定的顺序执行，但是这不应理解为要求这些操作必须以所示的特定顺序或者按顺行次序执行，也不应理解为要求必须执行所有示出的操作以获得期望的结果。

- 5 通过研究附图、公开内容和所附的权利要求书，本领域技术人员在实践所要求保护的主题时，能够理解和实现对于所公开的实施例的变型。在权利要求书中，词语“包括”不排除其他元件或步骤，并且不定冠词“一”或“一个”不排除多个。在相互不同的从属权利要求中记载了某些措施的仅有事实并不表明这些措施的组合不能用来获利。

权利要求书

1、一种语音降噪方法，由计算设备执行，包括：

获取带噪语音信号，所述带噪语音信号包括纯净语音信号和噪声信号；

5 估计所述带噪语音信号的后验信噪比和先验信噪比；

基于所估计的后验信噪比和所估计的先验信噪比在 Bark 域中确定语音/噪声似然比；

基于所确定的语音/噪声似然比估计先验语音存在概率；

10 基于所估计的后验信噪比、所估计的先验信噪比以及所估计的先验语音存在概率来确定增益，所述增益为用于将所述带噪语音信号变换成所述纯净语音信号的估计的频域传递函数；并且

基于所述增益从所述带噪语音信号导出所述纯净语音信号的所述估计。

2、如权利要求 1 所述的方法，其中所述估计所述带噪语音信号的后验信噪比和先验信噪比包括：

执行第一噪声估计，其中得到所述噪声信号的方差的第一估计；

利用所述噪声信号的方差的所述第一估计来估计所述后验信噪比；并且

利用所估计的后验信噪比来估计所述先验信噪比。

20 3、如权利要求 2 所述的方法，其中所述执行第一噪声估计包括：

对所述带噪语音信号的能量谱在频域和时域上进行平滑；

对经平滑的所述能量谱执行最小跟踪估计；并且

取决于经平滑的所述能量谱与该经平滑的所述能量谱的最小跟踪估计的比值，利用所述带噪语音信号的上一帧中的所述噪声信号的方差

的所述第一估计和所述带噪语音信号的当前帧的所述能量谱来选择性地更新所述带噪语音信号的当前帧中的所述噪声信号的方差的所述第一估计。

4、如权利要求 3 所述的方法，其中所述选择性地更新包括：

5 响应于所述比值大于或等于第一阈值而执行所述更新。

5、如权利要求 3 所述的方法，其中所述选择性地更新包括：

响应于所述比值小于所述第一阈值而不执行所述更新。

6、如权利要求 2 所述的方法，其中所述在 Bark 域中确定语音/噪声似然比包括：

10 基于高斯概率密度假设，将所述语音/噪声似然比计算为

$$\Delta(k, l) = \frac{\exp(\frac{\hat{\xi}(k, l)\hat{\gamma}(k, l)}{(1 + \hat{\xi}(k, l))})}{(1 + \hat{\xi}(k, l))}, \text{ 其中 } \Delta(k, l) \text{ 为所述带噪语音信号的第 } l \text{ 帧在第 } k$$

个频点上的所述语音/噪声似然比， $\hat{\xi}(k, l)$ 为所述第 l 帧在所述第 k 个频点上的所估计的先验信噪比，并且 $\hat{\gamma}(k, l)$ 为所述第 l 帧在所述第 k 个频点上的所估计的后验信噪比；并且

15 通过将 $\hat{\xi}(k, l)$ 和 $\hat{\gamma}(k, l)$ 从线性频域转换到 Bark 域而将 $\Delta(k, l)$ 变换为

$$\Delta(b, l) = \frac{\exp(\frac{\hat{\xi}(b, l)\hat{\gamma}(b, l)}{(1 + \hat{\xi}(b, l))})}{(1 + \hat{\xi}(b, l))}, \text{ 其中 } b \text{ 为 Bark 域中的频点。}$$

7、如权利要求 6 所述的方法，其中从线性频域到 Bark 域的所述转换基于以下等式：

$$b = 13 * \arctan(0.76 * f_{kHz}) + 3.5 * \arctan(\frac{f_{kHz}}{7.5})^2,$$

20 其中 f_{kHz} 为所述线性频域中的频率。

8、如权利要求 6 所述的方法，其中所述估计先验语音存在概率包括：

在对数域中将 $\Delta(b, l)$ 平滑为 $\log(\Delta(b, l)) = \beta * \log(\Delta(b, l-1)) + (1 - \beta) * \log(\Delta(b, l))$, 其中 β 为平滑因子; 并且

通过在 Bark 域的全带中映射 $\log(\Delta(b, l))$ 而得到所估计的所述先验语音存在概率。

9、如权利要求 8 所述的方法, 其中所述映射为 $P_{frame}(l) = \tanh(\frac{1}{24} \sum_{b=1}^{24} \log(\Delta(b, l)))$, 其中 $P_{frame}(l)$ 为所估计的先验语音存在概率。

10、如权利要求 2 所述的方法, 还包括:

独立于所述第一噪声估计执行第二噪声估计, 其中得到所述噪声信号号的方差的第二估计; 并且

取决于所述噪声信号的方差的所述第一估计在预定频率范围内的量值之和, 利用所述噪声信号的方差的所述第二估计来选择性地重新估计所述后验信噪比和所述先验信噪比,

其中所述确定增益包括: 响应于所述重新估计被执行而基于所重新估计的后验信噪比、所重新估计的先验信噪比以及所估计的先验语音存在概率来确定所述增益。

11、如权利要求 10 所述的方法, 其中所述执行第二噪声估计包括:

取决于所估计的先验语音存在概率, 利用所述带噪语音信号的上一帧中的所述噪声信号的方差的所述第二估计和所述带噪语音信号的当前帧的能量谱来选择性地更新所述带噪语音信号的当前帧中的所述噪声信号的方差的所述第二估计。

12、如权利要求 11 所述的方法, 其中所述选择性地更新包括:

响应于所估计的先验语音存在概率大于或等于第二阈值而执行所述更新。

13、如权利要求 11 所述的方法，其中所述选择性地更新包括：

响应于所估计的先验语音存在概率小于所述第二阈值而不执行所述更新。

14、如权利要求 10 所述的方法，其中所述选择性地重新估计所述
5 先验信噪比和所述后验信噪比包括：

响应于所述噪声信号的方差的所述第一估计在所述预定频率范围内的所述量值之和大于或等于第三阈值而执行所述重新估计。

15、如权利要求 10 所述的方法，其中所述选择性地重新估计所述先验信噪比和所述后验信噪比包括：

10 响应于所述噪声信号的方差的所述第一估计在所述预定频率范围内的所述量值之和小于所述第三阈值而不执行所述重新估计。

16、一种语音降噪装置，包括：

信号获取模块，被配置成获取带噪语音信号，所述带噪语音信号包括纯净语音信号和噪声信号；

15 信噪比估计模块，被配置成估计所述带噪语音信号的先验信噪比和后验信噪比；

似然比确定模块，被配置成基于所估计的先验信噪比和所估计的后验信噪比在 Bark 域中确定语音/噪声似然比；

20 概率估计模块，被配置成基于所确定的语音/噪声似然比估计先验语音存在概率；

增益确定模块，被配置成基于所估计的先验信噪比、所估计的后验信噪比以及所估计的先验语音存在概率来确定增益，所述增益为用于将所述带噪语音信号变换成所述纯净语音信号的估计的频域传递函数；以及

25 语音信号导出模块，被配置成基于所述增益从所述带噪语音信号导

出所述纯净语音信号的所述估计。

17、一种计算设备，包括处理器和存储器，所述存储器被配置成存储计算机程序，所述计算机程序被配置成当在所述处理器上执行时使所述处理器执行权利要求 1-15 中任一项所述的方法。

- 5 18、一种计算机可读存储介质，被配置成存储计算机程序，所述计算机程序被配置成当在处理器上执行时使所述处理器执行权利要求 1-15 中任一项所述的方法。

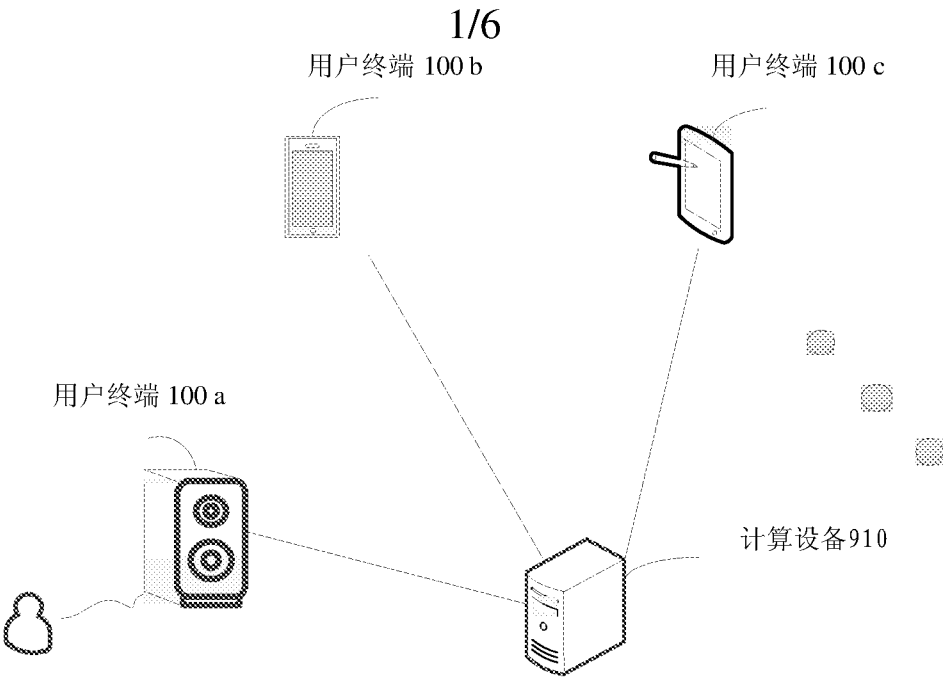


图 1A

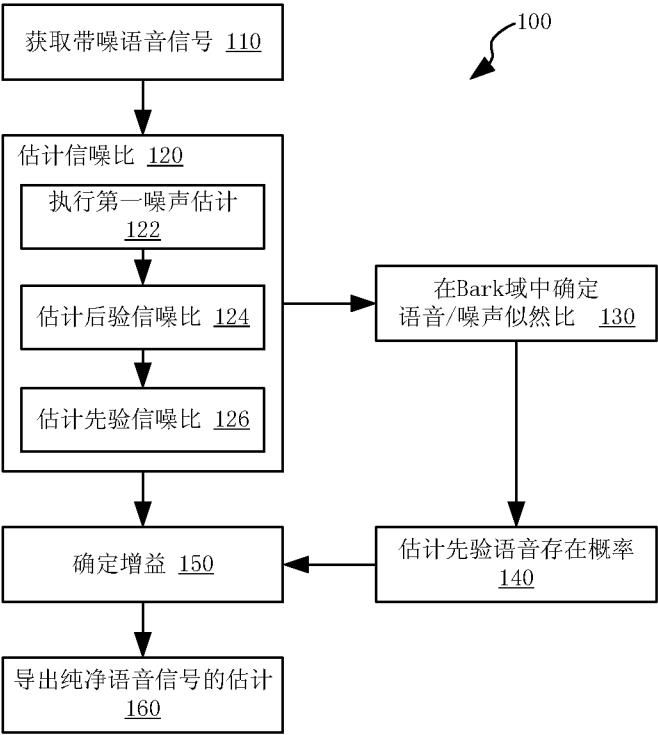


图 1B

2/6

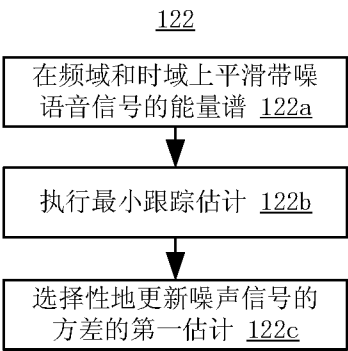


图 2

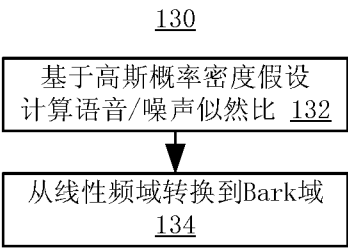


图 3

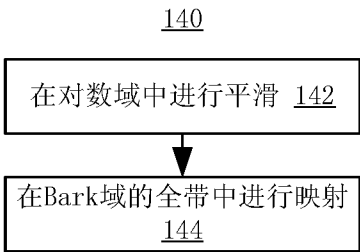


图 4

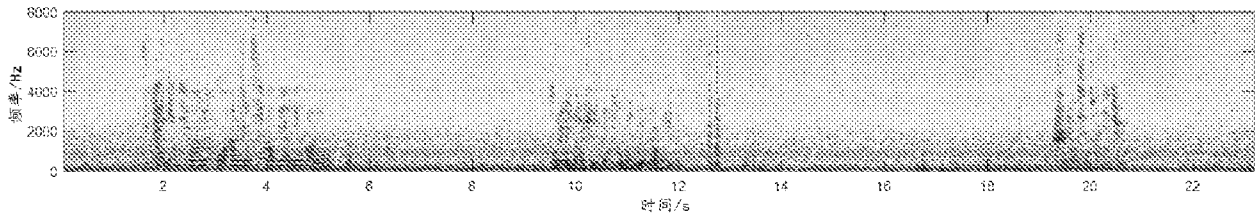


图 5a

3/6

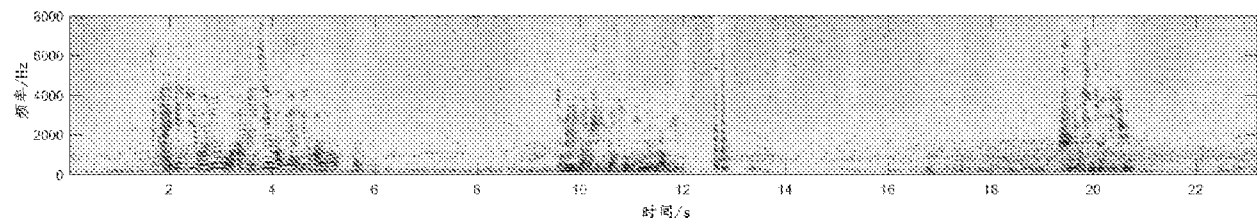


图 5b

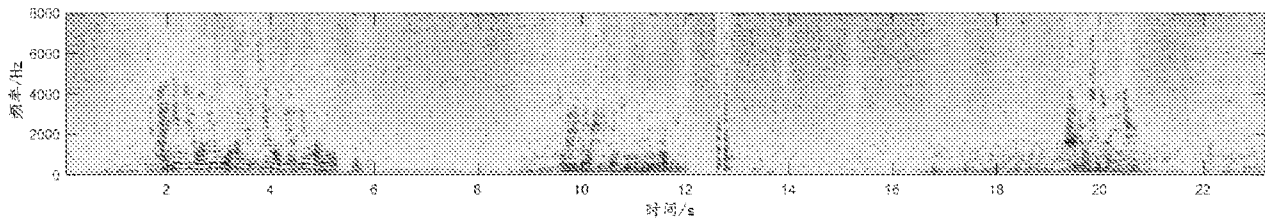


图 5c

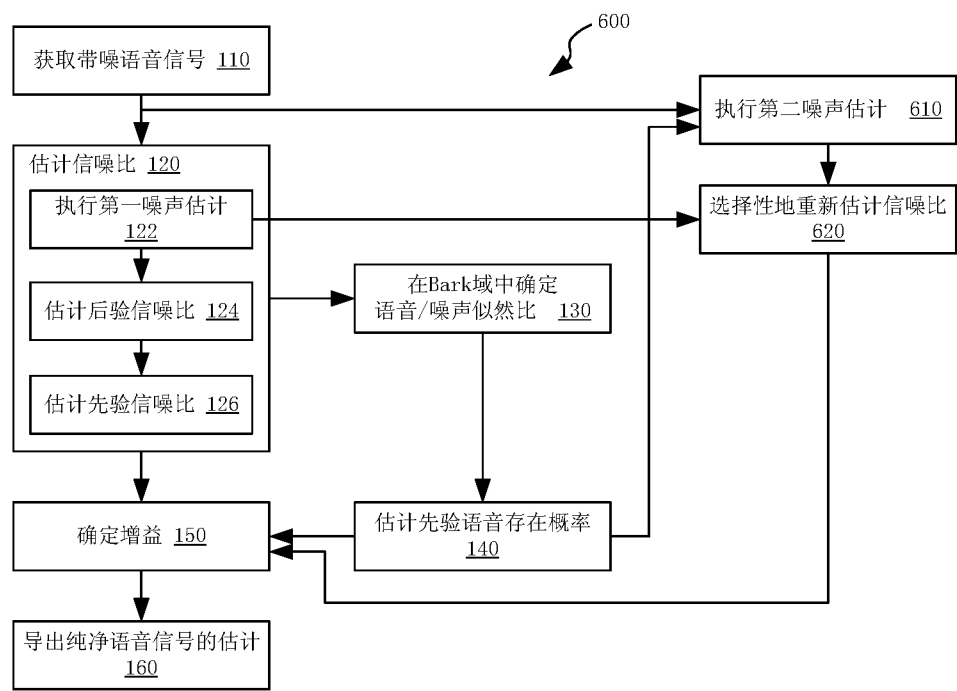


图 6

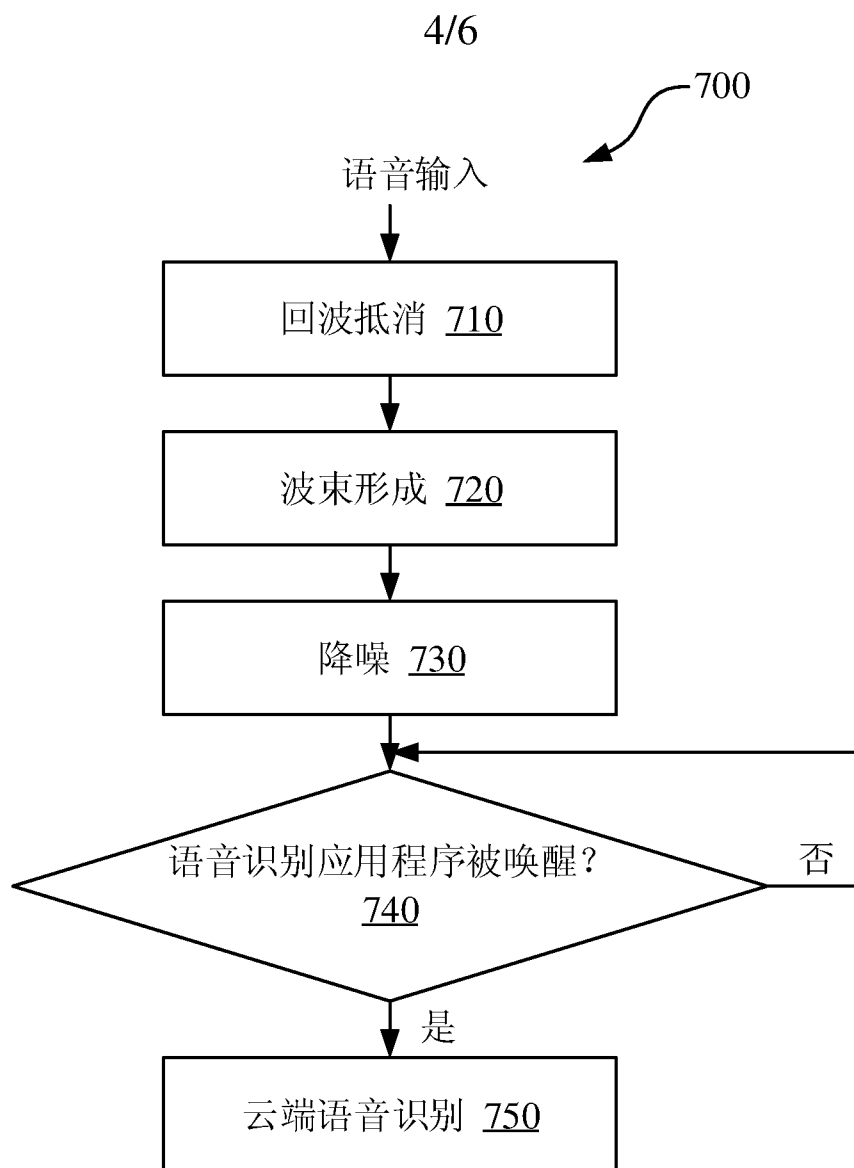


图 7

5/6

800

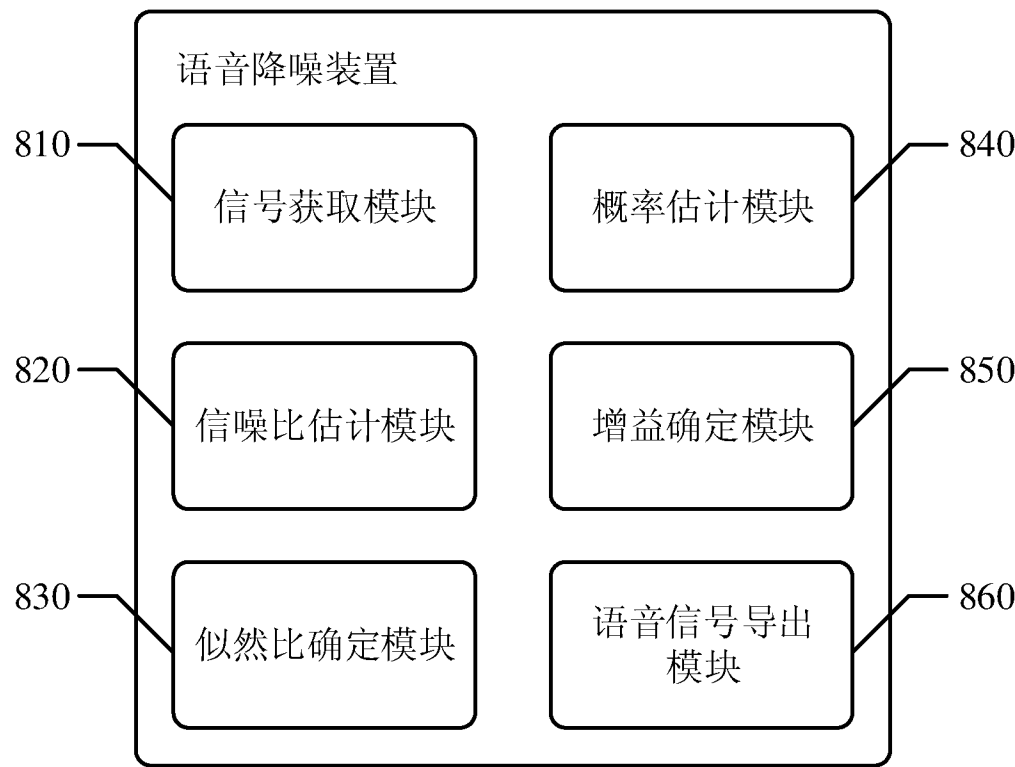


图 8

6/6

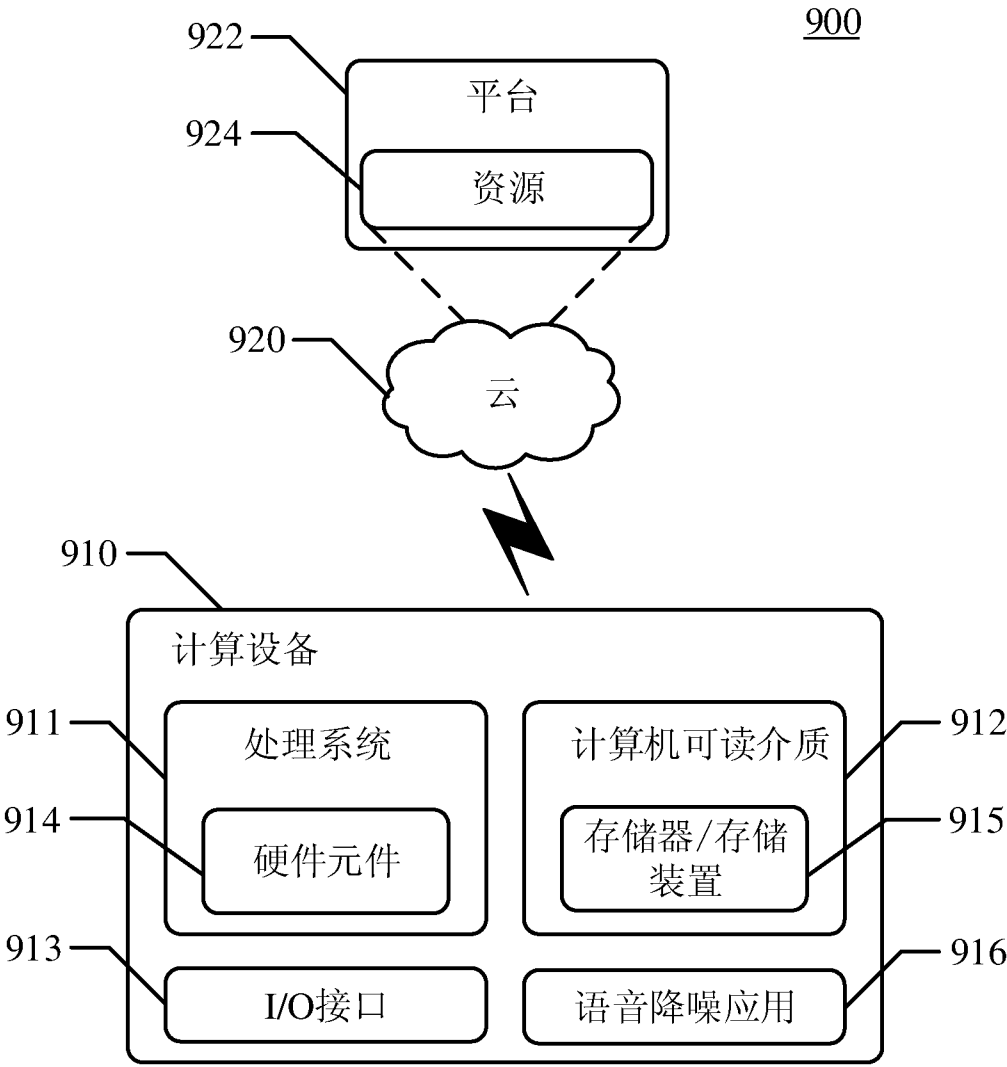


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/121953

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/0216(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; CNTXT; DWPI; CNKI; IEEE; 百度学术; 噪声, 噪音, 语音, 似然, 去噪, 除噪, 先验, 后验, 信噪比, 存在概率, 频域, 增益, noise, cancel, speech, audio, voice, likelihood, priori, posterior, SNR, existence probability, speech probability, gain transfer function

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110164467 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 23 August 2019 (2019-08-23) claims 1-15	1-18
A	CN 103650040 A (GOOGLE INC.) 19 March 2014 (2014-03-19) description, paragraphs [0046]-[0105], [0141]-[0151], figures 1-2	1-18
A	CN 108831499 A (THE 10TH RESEARCH INSTITUTE OF CHINA ELECTRONICS TECHNOLOGY GROUP CORPORATION) 16 November 2018 (2018-11-16) entire document	1-18
A	CN 103730124 A (SHANGHAI JIAO TONG UNIVERSITY WUXI RESEARCH INSTITUTE) 16 April 2014 (2014-04-16) entire document	1-18
A	US 2016042746 A1 (OKI ELECTRIC IND .CO., LTD.) 11 February 2016 (2016-02-11) entire document	1-18



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

26 February 2020

Date of mailing of the international search report

12 March 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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/121953

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110164467	A	23 August 2019	None			
CN	103650040	A	19 March 2014	CN	103650040	B	25 August 2017
				WO	2012158156	A1	22 November 2012
CN	108831499	A	16 November 2018	None			
CN	103730124	A	16 April 2014	None			
US	2016042746	A1	11 February 2016	US	9418677	B2	16 August 2016
				JP	2016038551	A	22 March 2016
				JP	6379839	B2	29 August 2018

国际检索报告

国际申请号

PCT/CN2019/121953

A. 主题的分类 G10L 21/0216(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;CNTXT;DWPI;CNKI;IEEE;百度学术:噪声, 噪音, 语音, 似然, 去噪, 除噪, 先验, 后验, 信噪比, 存在概率, 频域, 增益, noise, cancel, speech, audio, voice, likelihood, priori, posterior, SNR, existence probability, speech probability, gain transfer function																				
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>PX</td> <td>CN 110164467 A (腾讯科技深圳有限公司) 2019年 8月 23日 (2019 - 08 - 23) 权利要求1-15</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 103650040 A (谷歌公司) 2014年 3月 19日 (2014 - 03 - 19) 说明书第[0046]-[0105], [0141]-[0151]段, 附图1-2</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 108831499 A (西南电子技术研究所中国电子科技集团公司第十研究所) 2018年 11月 16日 (2018 - 11 - 16) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 103730124 A (上海交通大学无锡研究院) 2014年 4月 16日 (2014 - 04 - 16) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>US 2016042746 A1 (OKI ELECTRIC IND .CO., LTD.) 2016年 2月 11日 (2016 - 02 - 11) 全文</td> <td>1-18</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	PX	CN 110164467 A (腾讯科技深圳有限公司) 2019年 8月 23日 (2019 - 08 - 23) 权利要求1-15	1-18	A	CN 103650040 A (谷歌公司) 2014年 3月 19日 (2014 - 03 - 19) 说明书第[0046]-[0105], [0141]-[0151]段, 附图1-2	1-18	A	CN 108831499 A (西南电子技术研究所中国电子科技集团公司第十研究所) 2018年 11月 16日 (2018 - 11 - 16) 全文	1-18	A	CN 103730124 A (上海交通大学无锡研究院) 2014年 4月 16日 (2014 - 04 - 16) 全文	1-18	A	US 2016042746 A1 (OKI ELECTRIC IND .CO., LTD.) 2016年 2月 11日 (2016 - 02 - 11) 全文	1-18
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 110164467 A (腾讯科技深圳有限公司) 2019年 8月 23日 (2019 - 08 - 23) 权利要求1-15	1-18																		
A	CN 103650040 A (谷歌公司) 2014年 3月 19日 (2014 - 03 - 19) 说明书第[0046]-[0105], [0141]-[0151]段, 附图1-2	1-18																		
A	CN 108831499 A (西南电子技术研究所中国电子科技集团公司第十研究所) 2018年 11月 16日 (2018 - 11 - 16) 全文	1-18																		
A	CN 103730124 A (上海交通大学无锡研究院) 2014年 4月 16日 (2014 - 04 - 16) 全文	1-18																		
A	US 2016042746 A1 (OKI ELECTRIC IND .CO., LTD.) 2016年 2月 11日 (2016 - 02 - 11) 全文	1-18																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td> * 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 </td> <td> “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 </td> </tr> </table>			* 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																
* 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 26日		国际检索报告邮寄日期 2020年 3月 12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/CN2019/121953

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	110164467	A	2019年 8月 23日	无	
CN	103650040	A	2014年 3月 19日	CN	103650040 B 2017年 8月 25日
				WO	2012158156 A1 2012年 11月 22日
CN	108831499	A	2018年 11月 16日	无	
CN	103730124	A	2014年 4月 16日	无	
US	2016042746	A1	2016年 2月 11日	US	9418677 B2 2016年 8月 16日
				JP	2016038551 A 2016年 3月 22日
				JP	6379839 B2 2018年 8月 29日