(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 102959625 B
(45)授权公告日 2014.12.17

(21)申请号 201080030027.5
(22)申请日 2010.12.24
(85)PCT国际申请进入国家阶段日
2012.01.10
(86)PCT国际申请的申请数据
PCT/CN2010/080227 2010.12.24
(87)PCT国际申请的公布数据
WO2012/083555 EN 2012.06.28
(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
(72)发明人 王喆
(51)Int.Cl.
*G10L 25/78*(2013.01)
(56)对比文件
CN 101379548 A,2009.03.04,
CN 101320559 A,2008.12.10,
CN 101379548 A,2009.03.04,
CN 101320559 A,2008.12.10,
US 2004/0030544 A1,2004.02.12,
DE 10244699 A1,2004.04.01,
CN 101583996 A,2009.11.18,
EP 2113908 A1,2009.11.04,
审查员 赵云峰

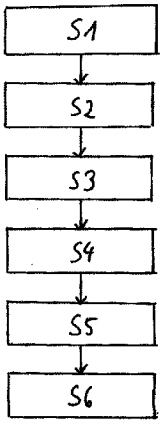
权利要求书4页 说明书10页 附图2页

(54)发明名称

自适应地检测输入音频信号中的话音活动的方法和设备

(57)摘要

本发明提供一种用于自适应地检测由帧组成的输入音频信号中的话音活动的方法和设备，包括以下步骤：至少基于所述所接收的输入音频信号的输入帧而确定所述输入信号的噪声特性(nc)；导出适于所述输入音频信号的所述噪声特性的或根据所述噪声特性而选择的VAD参数(vp)；以及将所述所导出的VAD参数与阈值进行比较，以提供话音活动检测决策。




CN 102959625 B

1. 一种用于自适应地检测由帧组成的输入音频信号中的话音活动的方法，其特征在于：所述方法包括以下步骤：

(a) 至少基于所述输入音频信号的输入帧而确定所述输入音频信号的噪声特性；

(b) 将所接收的所述音频信号的输入帧分成若干个子带；

(c) 获取所述输入帧的每一子带的信噪比；

(d) 基于每一子带的对应子带的信噪比使用自适应函数来计算子带特定参数，其中，根据所述输入音频信号的所述噪声特性来选择所述自适应函数的至少一个参数；以及

(e) 通过添加所述计算出的子带特定参数获取作为 VAD 参数的经修改的分段信噪比；

(f) 比较所获取的 VAD 参数及阈值，以提供话音活动检测决策。

2. 根据权利要求 1 所述的方法，其特征在于：

所述输入音频信号的所述噪声特性为长期信噪比和 / 或背景噪声变化。

3. 根据权利要求 1 所述的方法，其特征在于：

其中所述自适应函数为非线性函数。

4. 根据权利要求 1 或 3 所述的方法，其特征在于：

通过以下步骤来获取所述输入帧的每一子带的所述信噪比：

获取每一子带的信号能量，

估算每一子带的背景噪声能量，以及

根据相应子带的所述信号能量和所述背景噪声能量来计算每一子带的所述信噪比。

5. 根据权利要求 4 所述的方法，其特征在于：

其中所述输入帧的每一子带的所述信号能量为平滑信号能量，所述平滑信号能量形成所述输入帧与至少一个先前帧之间的加权平均值。

6. 根据权利要求 1 所述的方法，其特征在于：

通过添加子带特定参数来计算所述经修改的分段信噪比具体如下：

$$mssnr = \sum_{i=0}^{N} sbsp(i)$$

其中 N 为所述输入帧被分成的子频带的数目，

其中 sbsp(i) 为基于相应子带的所述信噪比而使用自适应函数计算出的子带特定参数。

7. 根据权利要求 6 所述的方法，其特征在于：

通过如下所示的方式来计算所述经修改的分段信噪比：

$$mssnr = \sum_{i=o}^{N} \left( f(snr(i)) + \alpha \right)^{\beta}$$

其中 snr(i) 为所述输入帧的第 i 个子带的信噪比，

N 为所述输入帧被分成的子频带的所述数目，

$(f(snr(i)+\alpha)^{\beta}$ 为用以计算所述子带特定参数的所述自适应函数，且

α、β 为所述自适应函数的两个可配置变量。

8. 根据权利要求 7 所述的方法，其特征在于：

其中所述自适应函数的第一变量 α 如下所示取决于所述输入音频信号的长期信噪

比：

$\alpha = g(i, lsnr)$

其中 g 为线性或非线性函数，lsnr 为所述输入音频信号的长期信噪比，且

其中所述自适应函数的第二变量 β 取决于所述长期信噪比和$\varphi$：

$$\beta = h(lsnr, \varphi)$$

其中 h 为非线性函数，且

$$\varphi = f(snr(i)) + \alpha。$$

9. 根据权利要求 8 所述的方法，其特征在于：

其中通过以下公式计算所述自适应函数的所述第一变量 α：

$\alpha = g(i, lnsr) = a(i) \cdot lsnr + b(i)$

其中 a(i)、b(i) 为取决于子带索引 i 的实数，且

通过以下公式计算所述自适应函数的所述第二变量 β：

$$\beta = h(lsnr, \varphi) = \begin{cases} \beta_1 & \varphi \geq d \text{且} lsnr > e_2 \\ \beta_2 & \varphi \geq d \text{且} e_1 < lsnr \leq e_2 \\ \beta_3 & \varphi \geq d \text{且} lsnr \leq e_1 \\ \beta_4 & \text{其它情况} \end{cases}$$

其中 $\beta_1$、$\beta_2$、$\beta_3$、$\beta_4$、d、$e_1$ 和 $e_2$ 为整数或浮点数，且 $\beta_1 < \beta_2 < \beta_3$，$e_1 < e_2$，所述 lsnr 为所述长期信噪比。

10. 根据权利要求 9 所述的方法，其特征在于：

其中将所述所获取的经修改的分段信噪比与阈值进行比较，所述阈值被设置为：

$$thr = \begin{cases} k_1 & lsnr > e_2 \\ k_2 & e_1 < lsnr \leq e_2 \\ k_3 & lsnr \leq e_1 \end{cases}$$

其中 $k_1$、$k_2$、$k_3$、$e_1$ 和 $e_2$ 为整数或浮点数，且 $k_1 > k_2 > k_3$，$e_1 < e_2$，其中如下所示方式生成所述话音活动检测决策 VADD：

$$VADD = \begin{cases} 1 & mssnr > thr \\ 0 & mssnr \leq thr \end{cases}$$

其中 VADD = 1 表示存在话音活动的主动帧，

且 VADD = 0 表示不存在话音活动的被动帧。

11. 根据权利要求 8 所述的方法，其特征在于：

其中通过以下公式计算所述自适应函数的所述第一变量 α

$\alpha = g(i, lsnr, \varepsilon) = a(i) \cdot lsnr + b(i) + c(\varepsilon)$

其中 a(i)、b(i) 为取决于子带索引 i 的实数，且

c(ε) 为取决于所述输入音频信号的背景噪声的估算波动的实数，且

其中通过以下公式计算所述自适应函数的所述第二变量 β：

$$\beta = h(lsnr, \varphi, \varepsilon) = \begin{cases} \beta_1 & \varphi \geq d \text{且} lsnr > e_2 \text{且} \varepsilon \leq p \\ \beta_2 & \varphi \geq d \text{且} lsnr > e_2 \text{且} \varepsilon > p \\ \beta_3 & \varphi \geq d \text{且} e_1 < lsnr < e_2 \text{且} \varepsilon \leq p \\ \beta_4 & \varphi \geq d \text{且} e_1 < lsnr < e_2 \text{且} \varepsilon > p \\ \beta_5 & \varphi \geq d \text{且} lsnr \leq e_1 \text{且} \varepsilon \leq p \\ \beta_6 & \varphi \geq d \text{且} lsnr \leq e_1 \text{且} \varepsilon > p \\ \beta_7 & \varphi < d \end{cases}$$

其中 $\varphi = f(snr(i)) + \alpha$，且 ε 为所述背景噪声的所述估算波动，且

d、$e_1$、$e_2$ 和 p 为整数或浮点数，且 $e_1 < e_2$，$\beta_1 = 3$、$\beta_2 = 4$、$\beta_3 = 7$、$\beta_4 = 10$、$\beta_5 = 8$、$\beta_6 = 15$、$\beta_7 = 15$。

12. 根据权利要求 11 所述的方法，其特征在于：

其中将所述所获取的经修改的分段信噪比与阈值进行比较，所述阈值被设置为：

$$thr = \begin{cases} q_1 + r_1 \cdot Min\left[\frac{lsnr - v_1}{W_1}, 1\right] & lsnr > e_2 \\ q_2 + r_2 Min\left[\frac{lsnr - v_2}{W_2}, 1\right] & e_1 < lsnr \leq e_2 \\ q_3 + r_3 \cdot Max\left[Min\frac{lsnr - v_3}{W_3}, 1\right] & lsnr \leq e_1 \end{cases}$$

其中 $q_1$、$q_2$、$q_3$、$r_1$、$r_2$、$r_3$、$e_1$、$e_2$、

$v_1$、$v_2$、$v_3$、$W_1$、$W_2$ 和 $W_3$ 为整数或浮点数，且 $e_1 < e_2$；

其中如下所示生成所述话音活动检测决策 VADD：

$$VADD = \begin{cases} 1 & mssnr > thr \\ 0 & mssnr \leq thr \end{cases}$$

其中 VADD = 1 表示存在话音活动的主动帧，

且 VADD = 0 表示不存在话音活动的被动帧。

13. 一种用于检测由帧组成的输入音频信号中的话音活动的话音活动检测设备，其特征在于：

所述话音活动检测设备包括：

(a) 基于信噪比的 VAD 参数计算单元，其基于每一子带的相应信噪比而使用自适应函数来计算所应用的输入帧的每一子带的所述信噪比和子带特定参数，并通过添加所述子带特定参数而导出经修改的分段信噪比；所述设备包括噪声特性确定单元，所述噪声特性确定单元至少基于所述输入音频信号的输入帧而确定所述输入音频信号的噪声特性，所述自适应函数是根据由所述噪声特性确定单元确定的至少一个噪声特性而选择的；以及 (b) 话音活动检测决策生成单元，其通过将所述经修改的分段信噪比与阈值进行比较而生成话音活动检测决策。

14. 根据权利要求 13 所述的话音活动检测设备，其特征在于：

所述噪声特性确定单元包括长期信噪比估算单元，所述长期信噪比估算单元计算所述

输入音频信号的长期信噪比。

15. 根据权利要求 13 所述的话音活动检测设备，其特征在于：

所述噪声特性确定单元包括背景噪声变化估算单元，所述背景噪声变化估算单元计算所述输入音频信号的所述背景噪声的稳定性或波动。

16. 一种音频信号处理装置，其特征在于，所述音频信号处理装置包括音频信号处理单元，所述音频信号处理单元包括根据前述权利要求 13 到 15 中任一权利要求所述的话音活动检测设备，并根据根据所述的话音活动检测设备所提供的话音活动检测决策来处理音频输入信号。

# 自适应地检测输入音频信号中的话音活动的方法和设备

## 技术领域

[0001] 本发明涉及一种用于自适应地检测由帧组成的输入音频信号中的话音活动的方法和设备，尤其涉及一种使用经非线性处理的子带分段信噪比参数的话音活动检测方法和设备。

## 背景技术

[0002] 话音活动检测 (VAD) 一般来说是一种供检测信号中的话音活动的技术。话音活动检测器广泛用于电信行业中。话音活动检测器的功能是在通信信道中检测例如语音或音乐等有源信号的有无。话音活动检测器可应用于通信网络内，以使所述网络可在不存在有源信号的周期中压缩传输带宽，或者根据指示是否存在有源信号的话音活动检测决策执行其它处理。话音活动检测器可将从输入信号中提取的特征参数或特征参数集与对应的阈值进行比较，并基于比较结果来确定所述输入是否包括有源信号。话音活动检测器的性能在很大程度上取决于所使用的特征参数的选择。已有许多特征参数被提出应用于话音活动检测，例如基于能量的参数、基于谱包络的参数、基于熵的参数或基于较高阶统计的参数。一般来说，基于能量的参数提供良好的话音活动检测性能。近年来，作为一种基于能量的参数的基于子带 SNR 的参数已广泛用于电信行业中。在基于子带 SNR 的话音活动检测器中，检测用于输入帧的每一子频带的 SNR，并添加所有子带的 SNR 以提供分段 SNR。此分段 SNR 参数 SSNR 可与阈值进行比较，以作出话音活动检测决策 VADD。所使用的阈值通常为变量，其根据输入信号的长期 SNR 或背景噪声的电平而自适应。

[0003] 在最近完成的 ITU-T 规范 G.720.1 中，已通过应用非线性处理而改进常规 SSNR 参数，从而获得经修改的 SSNR。还将计算出的经修改的分段 SNR 与阈值进行比较，所述阈值是从根据输入信号的长期 SNR、背景噪声变化以及话音活动检测操作点的阈值表而确定的，其中 VAD 操作点定义有源检测与无源检测之间的 VAD 决策的权衡，举例来说，质量优先的操作点将使 VAD 支持有源信号的决策，且反之亦然。

[0004] 尽管由 G.720.1 所使用的经修改的分段 SNR 参数改进了话音活动检测的性能，但不稳定和低 SNR 背景环境中的 VAD 性能仍需要改进。常规话音活动检测器经设计以平衡其在各种背景噪声条件下的性能。因此，常规话音活动检测器在特定条件下尤其是在不稳定和低 SNR 背景环境中的性能不够理想。

[0005] 因此，本发明的目的是提供一种具有高 VAD 性能的用于检测输入音频信号中的话音活动的方法和设备。

## 发明内容

[0006] 根据第一，本发明提供一种用于自适应地检测由帧组成的输入音频信号中的话音活动的方法，所述方法包括以下步骤：

[0007] (a) 至少基于所接收的所接收的输入音频信号的输入帧确定所述输入信号的噪声特性，

[0008] (e) 确定适于输入音频信号的所述噪声特性的或根据所述噪声特性而选择的 VAD 参数 (vp) ;以及

[0009] (f) 比较所获取的 VAD 参数及阈值进行,以提供话音活动检测决策。

[0010] 第一实施方案形式可将基于能量的参数、基于谱包络的参数、基于熵的参数或基于较高阶统计的参数用作 VAD 参数。

[0011] 在本发明的第一可实施方案中,本发明提供一种用于自适应性地检测由帧组成的输入音频信号中的话音活动的方法,所述方法包括以下步骤 :

[0012] (a) 所接收的输入音频信号的输入帧至少基于所接收的输入音频信号的输入帧而确定所述输入信号的噪声特性,

[0013] (b) 将所接收的所述音频信号的输入帧分成若干个子带,

[0014] (c) 获取所述输入帧的每一子带的 SNR,

[0015] (d) 基于每一子带的所述相应子带的 SNR 而使用自适应函数来计算子带特定参数,其中,所述自适应函数中的至少一个参数是根据所述噪音特性所选取的,

[0016] (e) 通过添加子带特定参数而获取作为所述 VAD 参数的经修改的分段 SNR ;以及

[0017] (f) 将所获取的经修改的分段 SNR 与阈值进行比较,以提供 VAD 决策。

[0018] 根据本发明的第一,本发明提供效率更高且质量更好的 VAD。VAD 的效率是检测噪声特性 ( 例如,背景噪声 ) 的能力,而 VAD 的质量与检测有源信号 ( 例如,输入音频信号中的语音或音乐 ) 的能力有关。

[0019] 在本发明的第一可实施方案中,所述所确定输入音频信号的噪声特性由所述输入音频信号的长期 SNR 形成。

[0020] 在本发明的第一另一可实施方案中,所述所确定输入音频信号的噪声特性由所述输入音频信号的背景噪声变化形成。

[0021] 在本发明的第一又一可实施方案中,所述所确定输入音频信号的噪声特性由所述输入音频信号的长期 SNR 和背景噪声变化的组合形成。

[0022] 在本发明的第一实施方案中,用于计算子带特定参数的自适应函数由非线性函数形成。

[0023] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中,通过获取每一子带的信号能量 ( 例如,输入帧的每一子带的信号能量 ) 来获取输入帧的每一子带的 SNR。

[0024] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的另一可实施方案中,通过估算每一子带的背景噪声能量来获取所述输入帧的每一子带的 SNR。

[0025] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的另一可实施方案中,通过根据相应子带的信号能量和背景噪声能量来计算每一子带的 SNR 来获取所述输入帧的每一子带的 SNR。

[0026] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的另一可实施方案中,所述输入帧的每一子带的信号能量为平滑信号能量,所述平滑信号能量形成所述输入帧与至少一个先前帧之间的加权平均值。

[0027] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的另一可实施方案中,通过如下所示添加子带特定参数来计算所述经修改的 SSNR :

[0028] $$mssnr = \sum_{i=0}^{N} sbsp(i)$$

[0029] 其中 N 为所述输入帧被分成的子频带的数目，

[0030] 其中 sbsp(i) 为子带特定参数，子带特定参数是使用自适应函数基于每一子带的子带 SNR 计算出的。

[0031] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中，所述修改的分段 SNR 的计算如下所示：

[0032] $$mssnr = \sum_{i=o}^{N} \left(f\left(snr(i)\right)+\alpha\right)^{\beta}$$

[0033] 其中 snr(i) 为输入帧的第 i 个子带的 SNR，

[0034] N 为所述输入帧被分成的子频带的数目，

[0035] $(f(snr(i)+\alpha)^{\beta}$ 为用以计算子带特定参数 sbsp(i) 的自适应函数 (AF)，且

[0036] α、β 为所述自适应函数 (AF) 的两个可配置变量。

[0037] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中，自适应函数 (AF) 的第一变量 α 如下所示取决于输入音频信号的长期 SNR(lsnr)：

[0038] α = g(i, lsnr)

[0039] 其中 g 为线性或非线性函数，且

[0040] 其中所述自适应函数 (AF) 的第二变量 β 取决于长期 SNR(lsnr) 和 $\varphi$：

[0041] $$\beta = h\,(lsnr,\ \varphi)$$

[0042] 其中 h 为非线性函数，且

[0043] $$\varphi = f(snr\,(i)) + \alpha。$$

[0044] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的另一实施方案中，通过以下公式计算自适应函数 ((AF)) 的第一变量 α：

[0045] α = g(i, lnsr) = a(i) • lsnr+b(i)

[0046] 其中 a(i)、b(i) 为取决于子带索引 i 的实数，且

[0047] 通过以下公式计算自适应函数 ((AF)) 的第二变量 β：

[0048] $$\beta = h(lsnr,\varphi) = \begin{cases} \beta_1 & \varphi \geq d \text{ 且 } lsnr > e_2 \\ \beta_2 & \varphi \geq d \text{ 且 } e_1 < lsnr \leq e_2 \\ \beta_3 & \varphi \geq d \text{ 且 } lsnr \leq e_1 \\ \beta_4 & \text{其它情况} \end{cases}$$

[0049] 其中 $\beta_1 < \beta_2 < \beta_3$ 以及 $\beta_4$ 和 d 以及 $e_1 < e_2$ 为整数或浮点数，且其中 lsnr 为输入音频信号的长期 SNR。

[0050] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中，将所获取的经修改的分段 SNR(mssnr) 与阈值 (thr) 进行比较，所述阈值 (thr) 被设置为：

[0051] $$thr=\begin{cases}k_1 & lsnr>e_2\\ k_2 & e_1<lsnr\le e_2\\ k_3 & lsnr\le e_1\end{cases}$$

[0052] 其中 $k_1>k_2>k_3$ 以及 $e_1<e_2$ 为整数或浮点数，其中生成话音活动检测决策(VADD)通过下述方式生成：

[0053] $$VADD=\begin{cases}1 & mssnr>thr\\ 0 & mssnr\le thr\end{cases}$$

[0054] 其中 VADD = 1 表示存在话音活动的主动帧，且

[0055] VADD = 0 表示不存在话音活动的被动帧。

[0056] 在根据本发明的第一用于自适应地检测话音活动输入音频信号的方法的一可实施方案中，通过以下公式计算自适应函数((AF))的第一变量 α：

[0057] $\alpha = g(i,lsnr,\varepsilon) = a(i)\cdot lsnr+b(i)+c(\varepsilon)$

[0058] 其中 a(i)、b(i) 为取决于子带索引 i 的实数，且

[0059] $c(\varepsilon)$ 为取决于估算处得所述输入音频信号的所述背景噪声的波动的实数，且

[0060] 其中通过以下公式计算所述自适应函数((AF))的第二变量 β：

[0061] $$\beta=h(lsnr,\varphi,\varepsilon)=\begin{cases}\beta_1 & \varphi\ge d\text{且}lsnr>e_2\text{且}\varepsilon\le p\\ \beta_2 & \varphi\ge d\text{且}lsnr>e_2\text{且}\varepsilon>p\\ \beta_3 & \varphi\ge d\text{且}e_1<lsnr<e_2\text{且}\varepsilon\le p\\ \beta_4 & \varphi\ge d\text{且}e_1<lsnr<e_2\text{且}\varepsilon>p\\ \beta_5 & \varphi\ge d\text{且}lsnr\le e_1\text{且}\varepsilon\le p\\ \beta_6 & \varphi\ge d\text{且}lsnr\le e_1\text{且}\varepsilon>p\\ \beta_7 & \varphi<d\end{cases}$$

[0062] 其中 $\varphi = f(snr(i))+\alpha$，且 ε 为所述估算出的背景噪声的波动，且

[0063] d 和 $e_1<e_2$ 以及 p 为整数或浮点数。

[0064] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中，将所获取的经修改的分段 SNR(mssnr) 与阈值 (thr) 进行比较，所述阈值 (thr) 被设置为：

[0065] $$thr=\begin{cases}q_1+r_1\cdot Min\left[\dfrac{lsnr-v_1}{W_1},1\right] & lsnr>e_2\\ q_2+r_2 Min\left[\dfrac{lsnr-v_2}{W_2}\right] & e_1<lsnr\le e_2\\ q_3+r_3\cdot Max\left[Min\dfrac{lsnr-v_3}{W_3},1\right] & lsnr\le e_1\end{cases}$$

[0066] 其中 $q_1$、$q_2$、$q_3$ 以及 $r_1$、$r_2$、$r_3$ 以及 $e_1<e_2$ 以及

[0067] $v_1$、$v_2$、$v_3$ 以及 $w_1$、$w_2$、$w_3$ 为整数或浮点数，

[0068] 其中如下所示生成所述话音活动检测决策 (VADD)：

[0069] $$VADD = \begin{cases} 1 & mssnr > thr \\ 0 & mssnr \le thr \end{cases}$$

[0070] 其中 VADD = 1 表示存在话音活动的主动帧，且

[0071] VADD = 0 表示不存在话音活动的被动帧。

[0072] 根据第二，本发明进一步提供一种用于检测由帧组成的输入音频信号中的话音活动的 VAD 设备，

[0073] 其中所述 VAD 设备包括：

[0074] 基于 SNR 的 VAD 参数计算单元，其基于每一子带的所述相应子带 SNR(snr) 而使用自适应函数 (AF) 来计算所应用的输入帧的每一子带的 SNR(snr) 和子带特定参数 (sbsp)，并通过添加子带的特定参数而获取经修改的分段 SNR(mssnr)；以及

[0075] VAD 决策生成单元，其通过将所述经修改的分段 SNR(mssnr) 与阈值进行比较而生成 VAD 决策 (VADD)。

[0076] 在根据本发明的第二的 VAD 设备的一可实施方案中，所述设备包括噪声特性确定单元，其所接收的输入音频信号的输入帧至少基于所接收的输入音频信号的输入帧确定输入信号的噪声特性 (nc)。

[0077] 在根据本发明的第二的 VAD 设备的一可实施方案中，噪声特性确定单元包括长期 SNR 估算单元，所述长期 SNR 估算单元计算所述输入音频信号的长期 SNR。

[0078] 在根据本发明的第二的 VAD 设备的另一可实施方案中，噪声特性确定单元包括背景噪声变化估算单元，所述背景噪声变化估算单元计算所述输入音频信号的背景噪声的稳定性或波动。

[0079] 在根据本发明的第二的 VAD 设备的另一可实施方案中，噪声特性确定单元包括长期 SNR 估算单元和背景噪声变化估算单元，所述长期 SNR 估算单元计算所述输入音频信号的长期 SNR，所述背景噪声变化估算单元计算所述输入音频信号的背景噪声的稳定性或波动。

[0080] 在根据本发明的第二的 VAD 设备的另一可实施方案中，根据由所述噪声特性确定单元确定的至少一个噪声特性 (nc) 来选择自适应函数 ((AF))。

[0081] 根据本发明的第三，本发明进一步提供一种音频信号处理装置，其中所述音频信号处理装置包括音频信号处理单元，所述音频信号处理单元用于根据由本发明的第二的 VAD 设备提供的 VAD 决策 (VADD) 来处理音频输入信号。

## 附图说明

[0082] 下文参看附图较详细地描述了本发明的不同方面的可实施方案。

[0083] 图 1 展示用于说明根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的可实施方案的流程图；

[0084] 图 2 展示根据本发明的第二的用于检测输入音频信号中的话音活动的 VAD 设备的框图；

[0085] 图 3 展示根据本发明的第三音频信号处理装置的框图。

## 具体实施方式

[0086] 图 1 展示根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的可实施方案的流程图。在本发明的第一示范性实施方案的第一步骤 S1 中，所接收的输入音频信号的输入帧至少基于所接收的输入音频信号的输入帧确定输入音频信号的噪声特性 nc。所述输入音频信号包括信号帧。在一可实施方案中，输入信号被分段成具有预定长度（例如 20ms）的帧，且被逐帧输入。在其它实施方案中，输入帧的长度可变化。步骤 S1 中所确定的输入音频信号的噪声特性 nc 可为由长期 SNR 估算单元计算出的长期 SNR lsnr。在另一可实施方案中，在步骤 S1 中所确定的噪声特性 nc 由背景噪声变化估算单元计算出的背景噪声变化形成，所述背景噪声变化估算单元计算输入音频信号的背景噪声 bn 的稳定性或波动 ε 。在步骤 S1 中所确定的噪声特性 nc 也可能即包含长期 SNR lsnr 也包括背景噪声变化。

[0087] 在另一步骤 S2 中，所接收的输入音频信号的输入帧被分成若干个子频带。

[0088] 在另一步骤 S3 中，基于每一子带的子带 SNR 而使用自适应函数 AF 来计算子带特定参数 sbsp。在一可实施方案中，通过快速傅里叶变换 (FFT) 为每一输入帧获取功率谱，且所获取的功率谱被分成具有非线性宽度的预定数目的子带。计算每一子带的能量，其中在一可实施方案中，输入帧的每一子带的能量可由平滑能量形成，所述平滑能量是由输入帧与至少一个先前帧之间的同一子带的能量的加权平均值形成的。在本发明的第一可实施方案中，可将子带 SNR(snr) 作为子频带的经修改的对数 SNR 而进行计算 ：

[0089] $$snr(i) = \log_{10}\left(\frac{E(i)}{E_n(i)}\right)$$

[0090] 其中 E(i) 为输入帧的第 i 个子带的能量，且 $E_n(i)$ 为背景噪声估算值 (background noise estimate) 的第 i 个子带的能量。可由背景噪声估算单元计算出背景噪声估算值，其中通过对所检测的背景噪声帧中每一子带的能量求移动平均值以计算背景噪声估算值的每一子带的能量。这可表达为 ：

[0091] $E_n(i) = \lambda \cdot E_n(i) + (1-\lambda) \cdot E(i)$

[0092] 其中 E(i) 为经检测后做为背景噪声的帧的第 i 个子带的能量，λ 为通常处于 0.9 到 0.99 范围内的“遗忘因子”。

[0093] 在步骤 S3 中已获取所述输入帧的每一子带的 SNR(snr) 之后，在步骤 S4 中基于相应子带的相应的 SNR(snr) 而使用自适应函数 (AF) 来计算子带特定参数 (sbsp)。在用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中，根据所确定输入音频信号的噪声特性而选择自适应函数 (AF) 的至少一个参数。在步骤 S1 中所确定的噪声特性 nc 可包括输入音频信号的长期 SNR 和 / 或背景噪声变化。自适应函数 AF 为非线性函数。

[0094] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中，在步骤 S5 中，通过如下所示的添加子带的特定参数 (sbsp) 而获取经修改的分段 SNR(mssnr) ：

[0095] $$mssnr = \sum_{i=0}^{N} sbsp(i)$$

[0096] 其中 N 为由所述输入帧分成的子频带的数目，且

[0097] 其中 sbsp(i) 为基于每一子带的子带 SNR 而使用自适应函数 (AF) 计算出的子带特定参数。在本发明的第一可实施方案中，所述经修改的分段 SNR(mssnr) 的计算如下 ：

[0098] $$mssnr = \sum_{i=o}^{N} \left( f\left( snr(i) \right) + \alpha \right)^{\beta}$$

[0099] 其中 snr(i) 为输入帧的第 i 个子带的 SNR，

[0100] N 为所述输入帧被分成的子频带的数目，且：

[0101] AF = $(f(snr(i)+\alpha)^{\beta}$ 为用以计算子带特定参数 sbsp(i) 的自适应函数，

[0102] 其中 α、β 为自适应函数 (AF) 的两个可配置变量。

[0103] 在本发明的第一可实施方案中，自适应函数 (AF) 的第一变量 α 如下所示取决于输入音频信号的长期 SNR(lsnr)：

[0104] α = g(i,lsnr)

[0105] 其中 g 为线性或非线性函数，且

[0106] 其中自适应函数 ((AF)) 的第二变量 β 取决于长期 SNR(lsnr) 和值$\varphi$：

[0107] $$\beta = h\,(\mathrm{lsnr},\ \varphi)$$

[0108] 其中 h 为非线性函数，且

[0109] $$\varphi = f(\mathrm{snr}\,(i)) + \alpha\,。$$

[0110] 在根据本发明的第一可实施方案中，通过以下公式计算自适应函数 (AF) 的第一变量 α：

[0111] α = g(i,lnsr) = a(i)·lsnr+b(i)

[0112] 其中 a(i)、b(i) 为取决于子带索引 i 的实数，且

[0113] 通过以下公式计算自适应函数 ((AF)) 的第二变量 β：

[0114] $$\beta = h(lsnr,\varphi) = \begin{cases} \beta_1 & \varphi \geq d \text{且} lsnr > e_2 \\ \beta_2 & \varphi \geq d \text{且} e_1 < lsnr \leq e_2 \\ \beta_3 & \varphi \geq d \text{且} lsnr \leq e_1 \\ \beta_4 & \text{其它情况} \end{cases}$$

[0115] 其中 $\beta_1 < \beta_2 < \beta_3$ 以及 $\beta_4$ 和 d 以及 $e_1 < e_2$ 为整数或浮点数，且其中 lsnr 为输入音频信号的长期 SNR。

[0116] 在一具体可实施方案中，$\beta_1 = 4$、$\beta_2 = 10$、$\beta_3 = 15$ 且 $\beta_4 = 9$。在此具体实施方案中，将 d 设置为 1，且 $e_1 = 8$ 且 $e_2 = 18$。

[0117] 在步骤 S5 中，通过添加子带的特定参数 (sbsp) 而获取经修改的分段 SNR(msnr)。在用于自适应地检测如图 1 中所示的输入音频信号中的话音活动的方法的实施方案的另一步骤 S6 中，将所获取的经修改的分段 SNR(mssnr) 与阈值 thr 进行比较，以提供 VAD 决策 (VADD)。

[0118] 在一可实施方案中，将所获取的经修改的分段 SNR(mssnr) 与阈值 thr 进行比较，所述阈值 thr 被设置为：

[0119] $$thr = \begin{cases} k_1 & lsnr > e_2 \\ k_2 & e_1 < lsnr \leq e_2 \\ k_3 & lsnr \leq e_1 \end{cases}$$

[0120] 其中 $k_1 > k_2 > k_3$ 以及 $e_1 < e_2$ 为整数或浮点数，且其中如下所示生成 VAD 决策 (VADD)：

[0121] $$VADD = \begin{cases} 1 & mssnr > thr \\ 0 & mssnr \le thr \end{cases}$$

[0122] 其中 VADD = 1 表示存在话音活动的主动帧，

[0123] 且 VADD = 0 表示不存在话音活动的被动帧。

[0124] 在一可能的具体实施方案中，$k_1 = 135$、$k_2 = 35$、$k_3 = 10$ 且 $e_1$ 被设置为 8 而 $e_2$ 被设置为 18。

[0125] 在用于自适应地检测输入音频信号中的话音活动的方法的另一可实施方案中，通过以下公式计算自适应函数 (AF) 的第一变量 α：

[0126] $\alpha = g(i, lsnr, \varepsilon) = a(i) \cdot lsnr + b(i) + c(\varepsilon)$

[0127] 其中 a(i)、b(i) 为取决于子带索引 i 的实数，且

[0128] c(ε) 为取决于输入音频信号的背景噪声 bn 的估算波动的实数，且

[0129] 其中通过以下公式计算自适应函数 (AF) 的第二变量 β：

[0130] $$\beta = h(lsnr, \varphi, \varepsilon) = \begin{cases} \beta_1 & \varphi \ge d \text{且} lsnr > e_2 \text{且} \varepsilon \le p \\ \beta_2 & \varphi \ge d \text{且} lsnr > e_2 \text{且} \varepsilon > p \\ \beta_3 & \varphi \ge d \text{且} e_1 < lsnr < e_2 \text{且} \varepsilon \le p \\ \beta_4 & \varphi \ge d \text{且} e_1 < lsnr < e_2 \text{且} \varepsilon > p \\ \beta_5 & \varphi \ge d \text{且} lsnr \le e_1 \text{且} \varepsilon \le p \\ \beta_6 & \varphi \ge d \text{且} lsnr \le e_1 \text{且} \varepsilon > p \\ \beta_7 & \varphi < d \end{cases}$$

[0131] 其中 $\varphi = f(snr(i)) + \alpha$，和 ε 估算出的背景噪声 bn 的波动，且

[0132] d 和 $e_1 < e_2$ 以及 p 为整数或浮点数。

[0133] 在特定实施方案中，如下所示设置参数：

[0134] $\beta_1 = 3$、$\beta_2 = 4$、$\beta_3 = 7$、$\beta_4 = 10$、$\beta_5 = 8$、$\beta_6 = 15$、$\beta_7 = 15$ 且

[0135] $d = 1$ 且 $e_1 = 8$ 且 $e_2 = 18$ 且 $p = 40$。

[0136] 在根据本发明的第一自适应地检测输入音频信号中的话音活动的方法的一实施方案中，将所获取的经修改的分段 SNR(mssnr) 与阈值 thr 进行比较，所述阈值被设置为：

[0137] $$thr = \begin{cases} q_1 + r_1 \cdot Min\left[\dfrac{lsnr - v_1}{W_1}, 1\right] & lsnr > e_2 \\ q_2 + r_2 Min\left[\dfrac{lsnr - v_2}{W_2}, 1\right] & e_1 < lsnr \le e_2 \\ q_3 + r_3 \cdot Max\left[Min\dfrac{lsnr - v_3}{W_3}, 1\right] & lsnr \le e_1 \end{cases}$$

[0138] 其中 $q_1$、$q_2$、$q_3$ 以及 $r_1$、$r_2$、$r_3$ 以及 $e_1 < e_2$ 以及

[0139] $v_1$、$v_2$、$v_3$ 以及 $w_1$、$w_2$、$w_3$ 为整数或浮点数。

[0140] 在本发明的第一具体实施方案中，$q_1 = 20$、$q_2 = 30$、$q_3 = 9$ 且 $r_1 = 30$、$r_2 = 10$ 且

$r_3 = 2$。另外，$v_1 = 18$、$v_2 = 8$ 且 $v_3 = 5$ 且 $w_1 = 8$、$w_2 = 10$ 且 $w_3 = 3$。另外，参数 $e_1$、$e_2$ 经设置为 $e_1 = 8$ 且 $e_2 = 18$。

[0141] 因此，在一可能的实施例中，不仅执行了背景噪声估算和长期 SNR 估算，而且还另外执行了背景噪声变化估算，以确定输入音频信号中背景噪声的背景噪声波动 ε。

[0142] 自适应函数 (AF) 的两个因子 α、β 调整经修改的分段 SNR 参数的辨别能力的权衡。不同的权衡表示所述检测更有利于对所接收的帧的主动检测或非主动检测。一般来说，输入音频信号的长期 SNR(lsnr) 越高，借助于调整自适应函数 (AF) 的对应的系数 α、β 而针对主动检测来调整经修改的分段 SNR(mssnr) 就越有利。

[0143] 在步骤 S6 中执行的 VAD 决策可进一步经历硬释放延迟 (hard hang-over) 程序。硬释放延迟程序迫使针对若干个帧的 VAD 决策在步骤 S6 中所获取的 VAD 决策从主动变为非主动之后立刻变为主动。

[0144] 在根据本发明的第一用于自适应地检测输入音频信号中的话音活动的方法的一可实施方案中，分析输入音频信号的背景噪声，并生成表示背景噪声的稳定性或波动（由 ε 表示）的程度的数字。可通过（例如）以下来计算背景噪声 bn 的此波动 ε：

[0145] $\varepsilon = \omega \cdot \varepsilon + (1-\omega) \cdot ssnr_n$

[0146] 其中 ω 为通常介于 0.9-0.99 之间的遗忘因子，且 $ssnr_n$ 为在被检测为背景帧的帧的所有子带上的 snr(i) 的总和乘以（例如）10 的因子。

[0147] 图 2 展示根据本发明的第二的 VAD 设备 1 的框图。所述 VAD 设备 1 包括基于 SNR 的 VAD 参数计算单元 2，所述基于 SNR 的 VAD 参数计算单元 2 接收施加到 VAD 设备 1 的入口 3 的输入音频信号。基于 SNR 的 VAD 参数计算单元 2 基于每一子带的所述相应子带 SNR(snr) 而使用自适应函数 (AF) 来计算输入音频信号的输入帧的每一子带的 SNR 以及子带的特定参数 (sbsp)，并通过添加子带的特定参数 (sbsp) 获取经修改的分段 SNR(mssnr)。基于 SNR 的 VAD 参数计算单元 2 将所获取的经修改的分段 SNR(mssnr) 提供给 VAD 设备 1 的 VAD 决策生成单元 4。所述 VAD 决策生成单元 4 通过将经修改的分段 SNR(mssnr) 与阈值 (thr) 进行比较而生成 VAD 决策 (VADD)。VAD 设备 1 在出口 5 处输出所生成的 VAD 决策 (VADD)。

[0148] 在根据本发明的第二的 VAD 设备 1 的一可实施方案中，VAD 检测设备 1 进一步包括如图 2 中所示的噪声特性确定单元 6。所述噪声特性确定单元 6 至少基于提供至到 VAD 设备 1 的入口 3 的所接收的输入音频信号的输入帧而确定输入信号的噪声特性 (nc)。在一替代实施方案中，将噪声特性 (nc) 从外部噪声特性确定实体提供给基于 SNR 的 VAD 参数计算单元 2。在根据本发明的第二的 VAD 设备 1 的一可实施方案中，如图 2 中所示的噪声特性确定单元 6 可包括长期 SNR 估算单元，所述长期 SNR 估算单元计算输入音频信号的长期 SNR(lsnr)。在另一可实施方案中，噪声特性确定单元 6 还可包括背景噪声变化估算单元，所述背景噪声变化估算单元计算输入音频信号的背景噪声 bn 的稳定性或波动 ε。因此，由噪声特性确定单元 6 提供的噪声特性 (nc) 可包括输入音频信号的长期 SNR(lsnr) 和 / 或输入音频信号的背景噪声的稳定性或波动（ε）。在一可实施方案中，根据由所述噪声特性确定单元 6 确定的至少一个噪声特性 nc 来选择由基于 SNR 的 VAD 参数计算单元 2 所使用的自适应函数 (AF)。

[0149] 图 3 展示根据本发明的第三音频信号处理装置 7 的框图，其包括 VAD 设备 1，所述 VAD 设备 1 为音频信号处理装置 7 内的音频信号处理单元 8 提供 VAD 决策 (VADD)。音频信

号处理单元 8 根据所接收的由本发明的第一 VAD 设备 1 生成的 VAD 决策 (VADD) 来执行对输入音频信号的音频信号处理。音频信号处理单元 8 可基于所述 VAD 决策 (VADD) 而执行 ( 例如 ) 对输入音频信号的编码。音频信号处理装置 7 可形成例如移动电话等语音通信装置的一部分。另外，音频信号处理装置 7 可提供于语音通信系统内，例如，音频会议系统、回声信号消除系统、语音降噪系统、语音识别系统或语音编码系统。在一可实施方案中，由 VAD 设备 1 生成的 VAD 决策 (VADD) 可控制实体 ( 例如，蜂窝式无线电系统 ( 例如，GSM 或 LTE 或 CDMA 系统 ) 中的实体 ) 的不连续传输 DTX 模式。VAD 设备 1 可通过减少共信道干扰来增强例如蜂窝式无线电系统等系统的系统容量。此外，可显著减少蜂窝式无线电系统内的便携式数字装置的功耗。

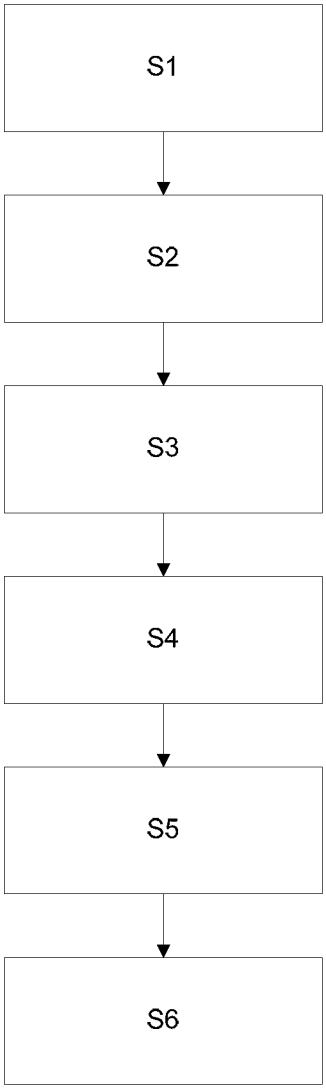

S1
S2
S3
S4
S5
S6

图 1

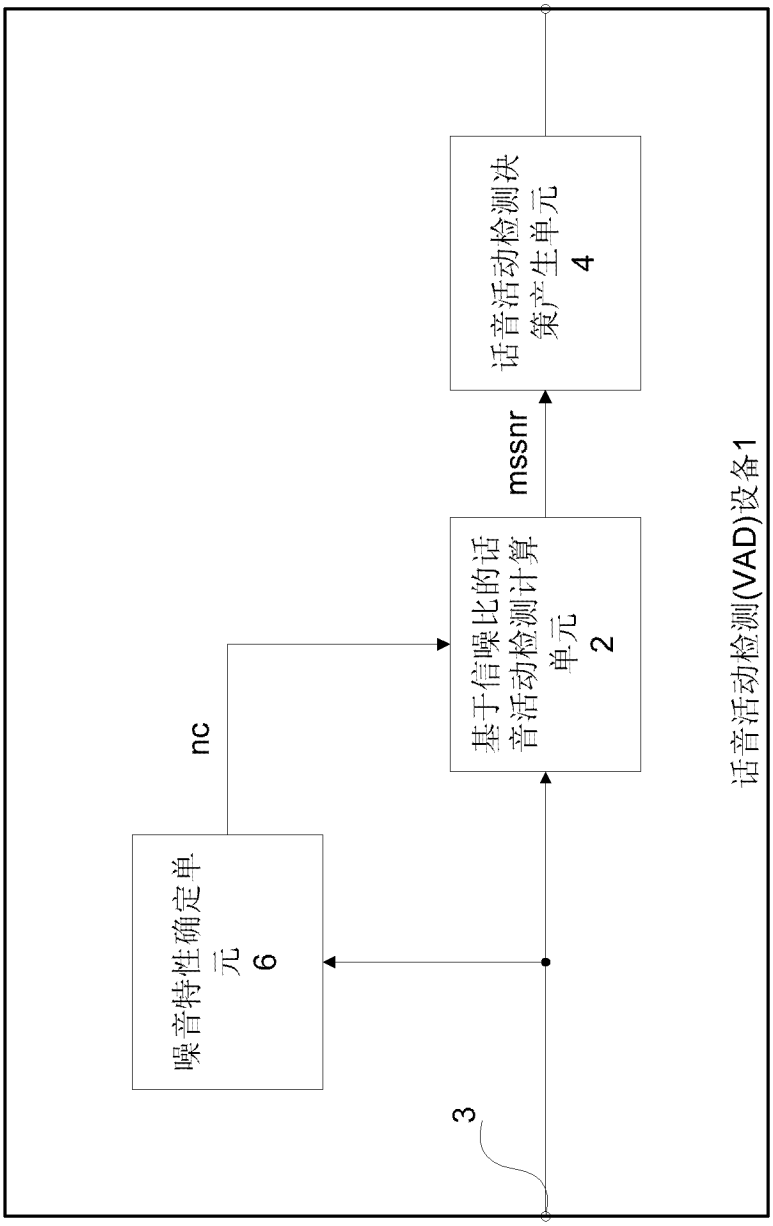

话音活动检测(VAD)设备1
话音活动检测决策产生单元
4
mssnr
基于信噪比的话音活动检测计算单元
2
nc
噪音特性确定单元
6
3

图 2

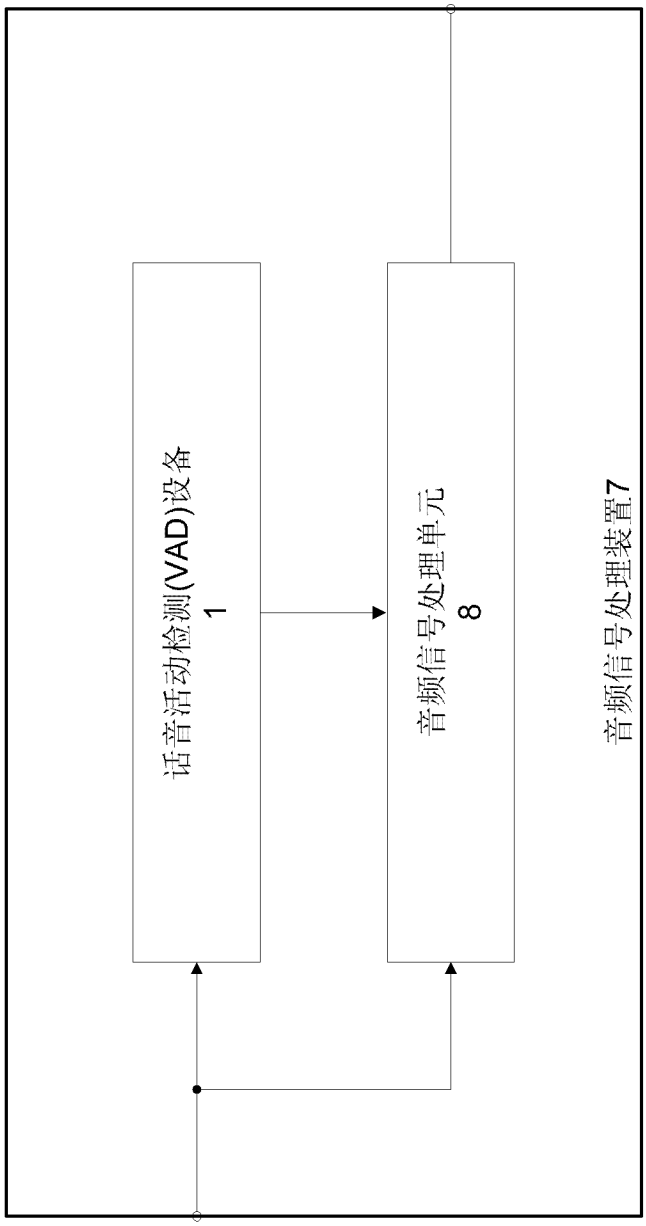



图 3