(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102456351 A
(43)申请公布日 2012.05.16

(21)申请号 201010515293.9

(22)申请日 2010.10.14

(71)申请人 清华大学
地址 100084 北京市海淀区清华园一号

(72)发明人 梁维谦 胡奎 杨华中 蒋毅
陈卓

(74)专利代理机构 北京润泽恒知识产权代理有限公司 11319
代理人 苏培华

(51)Int.Cl.
G10L 21/02(2006.01)
G10L 19/00(2006.01)
H04R 3/00(2006.01)
H04R 25/00(2006.01)

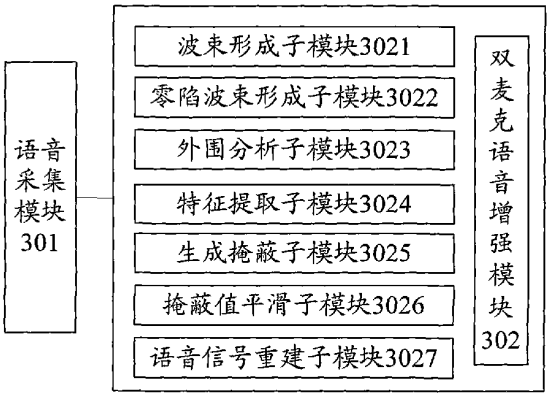
权利要求书 1 页 说明书 10 页 附图 5 页

(54)发明名称

一种语音增强的系统

(57)摘要

本发明提供了一种语音增强的系统，通过特定的双麦克结构，模拟人耳的听觉场景分析能力来实现语音增强，具有与噪声类型无关的特点，可广泛应用于各类噪声环境下的语音增强，利用计算听觉场景分析的原理，将目标语音从背景噪声中进行分离，从而实现去噪，从而实现了与噪声的具体类型、各种噪声源的个数、目标声源与噪声源的空间拓扑结构无关，在实现较好去噪效果的同时保证较高的语音质量。



CN 102456351 A

1. 一种语音增强的系统，其特征在于，所述系统包括：

语音采集模块，包括两路麦克，其中一路麦克置于目标声源的近端；另一路麦克置于目标声源的远端；所述采集模块用于采集两路语音信号；

双麦克语音增强模块，用于对采集的两路语音信号进行处理，以获取增强后的目标声源语音信号；所述双麦克语音增强模块包括以下子模块：

波束形成子模块，用于获得目标声源语音的参考信号；

零陷波束形成子模块，用于获得环境噪声的参考信号；

外围分析子模块，用于通过模拟声音进入人耳的过程，将两路语音进行分频及对语音信号进行变换；

特征提取子模块，用于提取分频后两路语音信号的延时差和能量差信息；

生成掩蔽子模块，根据提取的延时差和能量差信息获得不同时频区域对应的掩蔽值；

掩蔽值平滑子模块，用于对提取的掩蔽值进行平滑处理；

语音信号重建子模块，用于对由波束形成模块得到的参考信号进行掩蔽值处理，并合成增强后的语音信号作为输出。

2. 根据权利要求 1 所述的系统，其特征在于，所述语音采集模块还包括以下子模块：

滤波子模块，用于对两路麦克采集的语音信号进行带通滤波；

放大子模块，用于将滤波后的语音信号进行放大；

A/D 转换子模块，用于将放大后的语音信号转换为数字信号。

3. 根据权利要求 1 所述的系统，其特征在于，所述双麦克语音增强模块还包括以下子模块：

语音激活检测子模块，用于检测纯环境噪声语音段。

4. 根据权利要求 1 所述的系统，其特征在于，所述外围分析子模块包括以下单元：

内耳耳蜗模拟单元，用于将语音信号进行 gammatone 滤波分频，获取两路信号在 128 个不同子频带的语音信号；

内耳非线性神经传导模拟单元，用于将每个频率通道的子带信号进行非线性变换；

中耳模拟单元，用于通过对每个通道的 gammatone 滤波器的增益按照等响度曲线进行调整。

一种语音增强的系统

技术领域

[0001] 本发明涉及语音增强技术领域，特别是涉及一种语音增强的系统。

背景技术

[0002] 语音增强技术应用于噪声环境下的语音通信，可以提高通话质量；应用于人机对话，可以提高识别正确率。在人们的日常生活的各种噪声环境下，人们往往更希望获取经过降噪处理后的语音信息。语音增强的方法按通道个数可以分为单麦克语音增强与麦克风阵列增强技术。传统的单麦克语音增强技术如谱减法、维纳滤波等方法，都是先估计噪声幅值或能量，再将其从带噪语音中减去。对于平稳噪声如白噪声，可以达到一定的效果，然而对于诸如非目标人说话的噪声、音乐噪声等非平稳噪声则可能造成较严重的语音损伤。传统的多麦克语音增强技术是波速形成技术，包括延时相加、延时相减等固定波束，以及广义旁瓣消除等自适应波束。贝尔实验室研发了一种具有自适应特性的差分麦克风阵列(Gary W.Elko，Anh-Tho Nguyen Pong，A simple adaptive first-orderdifferential microphone，In：proc.1995Workshop on Applications of SignalProcessing to Audio and Acoustics，72-169)，此类方法通过自适应调整空间滤波的参数，对位于零陷方向的不同类型的噪声均有一定的去噪效果但仍存在例如声源定位、对于来自与主声源相近的角度方向的噪声抑制效果差等问题。此外还有多子带的处理技术，如华为公司申请的一个专利(200410034505.6)“一种语音增强方法”。该方法采用的是多子带处理技术，虽然也能取得一定的去噪效果，但由于其仍是基于对子带信噪比的估计，因而也无法广泛适用于各种噪声类型。

[0003] 因此，目前需要本领域技术人员迫切解决的一个技术问题就是：如何能够创新地提出一种语音增强的方法或者是系统，以满足各类噪声环境下的语音增强需求。

发明内容

[0004] 本发明所要解决的技术问题是提供一种语音增强的系统，用以满足各类噪声环境下的语音增强需求，在实现较好去噪效果的同时保证较高的语音质量。

[0005] 为了解决上述问题，本发明公开了一种语音增强的系统，所述系统包括：

[0006] 语音采集模块，包括两路麦克，其中一路麦克置于目标声源的近端；另一路麦克置于目标声源的远端；所述采集模块用于采集两路语音信号；

[0007] 双麦克语音增强模块，用于对采集的两路语音信号进行处理，以获取增强后的目标声源语音信号；所述双麦克语音增强模块包括以下子模块：

[0008] 波束形成子模块，用于获得目标声源语音的参考信号；

[0009] 零陷波束形成子模块，用于获得环境噪声的参考信号；

[0010] 外围分析子模块，用于通过模拟声音进入人耳的过程，将两路语音进行分频及对语音信号进行变换；

[0011] 特征提取子模块，用于提取分频后两路语音信号的延时差和能量差信息；

[0012] 生成掩蔽子模块，根据提取的延时差和能量差信息获得不同时频区域对应的掩蔽值；

[0013] 掩蔽值平滑子模块，用于对提取的掩蔽值进行平滑处理；

[0014] 语音信号重建子模块，用于对由波束形成模块得到的参考信号进行掩蔽值处理，并合成增强后的语音信号作为输出。

[0015] 优选的，所述语音采集模块还包括以下子模块：

[0016] 滤波子模块，用于对两路麦克采集的语音信号进行带通滤波；

[0017] 放大子模块，用于将滤波后的语音信号进行放大；

[0018] A/D 转换子模块，用于将放大后的语音信号转换为数字信号。

[0019] 优选的，所述双麦克语音增强模块还包括以下子模块：

[0020] 语音激活检测子模块，用于检测纯环境噪声语音段。

[0021] 优选的，所述外围分析子模块包括以下单元：

[0022] 内耳耳蜗模拟单元，用于将语音信号进行 gammatone 滤波分频，获取两路信号在 128 个不同子频带的语音信号；

[0023] 内耳非线性神经传导模拟单元，用于将每个频率通道的子带信号进行非线性变换；

[0024] 中耳模拟单元，用于通过对每个通道的 gammatone 滤波器的增益按照等响度曲线进行调整。

[0025] 与现有技术相比，本发明具有以下优点：

[0026] 本发明通过特定的双麦克结构，模拟人耳的听觉场景分析能力来实现语音增强，具有与噪声类型无关的特点，可广泛应用于各类噪声环境下的语音增强，利用计算听觉场景分析的原理，将目标语音从背景噪声中进行分离，从而实现去噪，从而实现了与噪声的具体类型、各种噪声源的个数、目标声源与噪声源的空间拓扑结构无关，在实现较好去噪效果的同时保证较高的语音质量。

附图说明

[0027] 图 1 是本发明具体实施方式中所述的人耳基本结构的示意图；

[0028] 图 2 是本发明具体实施方式中所述的外围分析的基本原理与人耳工作机理的对照示意图；

[0029] 图 3 是本发明实施例所述的一种语音增强的系统结构图；

[0030] 图 4 是本发明实施例所述的语音采集中麦克风阵列分布的结构示意图；

[0031] 图 5 是本发明实施例所述的语音信号采集部分结构框图；

[0032] 图 6 是本发明实施例所述的双麦克语音增强算法部分总体示意图；

[0033] 图 7 是本发明实施例所述的双麦克 VAD 模块原理示意图图；

[0034] 图 8 是本发明实施例所述的 Meddis 模型原理示意图；

[0035] 图 9 是本发明实施例所述的信号重建原理示意图。

具体实施方式

[0036] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实

施方式对本发明作进一步详细的说明。

[0037] 近年来，基于听觉场景分析的语音处理技术的研究日趋活跃。听觉场景分析(Auditory Scene Analysis)是指人类的听觉系统能够从复杂的混合声音中选择并跟踪某一说话人的声音，这一现象首先由 Cherry 发现，并称之为“鸡尾酒会效应”。听觉场景分析的概念首先是由著名心理听觉学家 Albert Bregman 在其专著《计算场景分析》中提出。听觉系统利用声音的各种特性（时域、频域、空间位置等），通过自下而上（分解）和自上而下（学习）的双向信息交流，对现实世界的混合声音进行分解，使各成分归属于各自的物理声源。

[0038] 此后，人们尝试用计算机模拟人的这种听觉特性，产生了计算听觉场景分析(Computational Auditory Scene Analysis,CASA)方法。既然是模拟人的一种生理机能，因此，这里对人耳的生理结构及声音进入人耳、引起神经冲动并由听神经传导、人脑的处理机能做一些介绍并用计算机算法的形式模拟实现。

[0039] 人耳的基本结构的示意图如图 1 所示，主要包括外耳、中耳、内耳。其中，外耳包括外耳道和鼓膜，鼓膜是中耳的门户。声音经鼓膜传到中耳，中耳主要由锥骨、镫骨、钻骨三块听小骨组成，其对声音的传播起到一个类似于杠杆的放大作用。内耳里最重要的器官是耳蜗，当声音引起内耳的卵圆窗振动后，这种振动通过耳蜗内的淋巴液的流动传递。而耳蜗内有细小的毛细胞把淋巴液流动转化为生物电信号产生神经冲动，最后由神经把信息送往大脑进一步处理。

[0040] 模拟声音由空间路径进行传播以及人的外耳部分对声音的影响我们用的是 HRTF(Head Related Transfer Function)，这个头相关传输函数是通过一个人头模型采集声音信号，再计算出来的一个传输函数。

[0041] 由于中耳的模拟是与内耳的工作相关的，我们先介绍内耳。内耳的第一个过程是进行耳蜗滤波，将声音分解到不同的频带上。例如，可以采用 128 个滤波器组成的非均匀的 gammatone 滤波器组，由于各个频带是依据人耳的听觉特性进行划分的，gammatone 滤波器组体现了人耳的听觉特性信息。

[0042] 而中耳的工作可以对于各个频率通道按照等响度曲线对 gammatone 滤波器的增益进行调整来进行模拟。

[0043] 内耳的另一个过程就是产生神经冲动的过程，这是一个非线性变换的过程。可以采用 Meddis 模型进行模拟。

[0044] 以上模拟人耳的三个过程我们称为外围分析，外围分析的基本原理与人耳工作机理的示意图如图 2 所示。

[0045] 实施例：

[0046] 参照图 3，示出了本发明的一种语音增强的系统结构图，所述系统具体包括：

[0047] 语音采集模块 301，包括两路麦克，其中一路麦克置于目标声源的近端；另一路麦克置于目标声源的远端；所述采集模块用于采集两路语音信号；

[0048] 本发明提出的语音增强技术采用两路麦克风，因而属于麦克风阵列语音增强技术的一种。

[0049] 语音采集部分的麦克风阵列分布结构如图 4 所示。其中一路麦克置于目标声源的近端，另一路麦克置于目标说话人的远端。

[0050] 优选的，所述语音采集模块 301 还包括以下子模块：

[0051] 滤波子模块 3011，用于对两路麦克采集的语音信号进行带通滤波；

[0052] 放大子模块 3012，用于将滤波后的语音信号进行放大；

[0053] A/D 转换子模块 3013，用于将放大后的语音信号转换为数字信号。

[0054] 两个麦克采集两路语音信号，所采集的两路语音信号首先经过滤波和放大处理，再通过 A/D 变换得到语音数字信号，以备进一步处理。语音信号采集部分的结构框图如图 5 所示。近端麦克风主要是采集的目标声源语音信号但混杂有环境噪声。为使最后的处理效果更好，先是使用直接采集到的两路语音信号进行波束形成计算，形成主瓣方向对准目标声源的一个波束，以抑制掉一部分环境噪声。

[0055] 远端麦克风主要是采集的环境噪声的参考信号，但混杂有目标声源语音信号。为使最后的处理效果更好，使用直接采集到的两路语音信号进行零陷波束形成计算，形成零瓣方向对准目标声源的一个波束，以抑制掉一部分目标声源信号。

[0056] 双麦克语音增强模块 302，用于对采集的两路语音信号进行处理，以获取增强后的目标声源语音信号；

[0057] 所述双麦克语音增强模块 302 包括以下子模块：

[0058] 波束形成子模块 3021，用于获得目标声源语音的参考信号；

[0059] 零陷波束形成子模块 3022，用于获得环境噪声的参考信号；

[0060] 波束形成子模块 3021 与零陷波束形成子模块 3022 的原理相似，大致如下：

[0061] 对于位置 c 处的声源发出的语音信号由采集电路采集到的两路信号 $x_1(n)$ 与 $x_2(n)$ 的频域表达为 $X_i(k)$ $(i = 1,2)$ 如式 (13) 所示：

[0062] $$X_i(k) = D_i(k,c)A_i(k)U_i(k,c)S(k) \quad i = 1,2 \tag{13}$$

[0063] 其中，$c = \{x, y, z\}$ 是直角系中声源的坐标，$p_i = \{x_i, y_i, z_i\}$ 是第 i 个麦克在直角坐标系中的坐标，$S(k)$ 是声源信号，$D_i(k,c)$ 是表示声音在空间中传播时幅度与相位的变化，其表达式如式 (14) 所示：

[0064] $$D_i(k,c) = \frac{e^{-j2\pi f_k v\|c-p_i\|}}{\|c-p_i\|} \qquad i = 1,2 \tag{14}$$

[0065] 上式中 f_k 表示对应第 k 个频点的时间频率值，v 是声音传播速度值的倒数。$A_i(k)$ 是表示图 5 采集电路中前置放大与 A/D 转换的影响，$U_i(k, c)$ 表示第 i 个麦克风自身固有的方向性。

[0066] 对于每个麦克，对应有一个滤波器，其参数矢量用 $W_i(k)$ 表示，则经麦克阵列波束形成处理后的结果 $Y(k)$ 如式 (15) 所示：

[0067] $$Y(k) = \sum_{i=1}^{2} X_i(k)W_i(k) \tag{15}$$

[0068] 归一化后得到对于声源位于位置 c 时的波束形状表达式如式 (16) 所示：

[0069] $$B(k,c) = \sum_{i=1}^{2} W_i(k)D_i(k,c)U_i(k,c) \tag{16}$$

[0070] 对于假定的噪声模型，采用最小噪声能量、以所需的波束形状（波束主瓣对准目标声源或是零瓣对准目标声源）为约束条件，由上式，可以得到波束形成或是零陷波束形成滤波处理时，两个麦克对应的滤波器参数 $W_i(k)$。

[0071] 外围分析子模块 3023，用于通过模拟声音进入人耳的过程，将两路语音进行分频及对语音信号进行变换；

[0072] 在这一阶段，前一阶段由两路麦克波束形成子模块与零陷波束形成子模块得到的两路输出信号（频域记为 $Y_1(k)$ 与 $Y_2(k)$，时域记为 $y_1(n)$ 与 $y_2(n)$）分别经过外围分析，通过模拟声音进入人耳的过程，将两路语音进行分频和变换。

[0073] 对于声音在空间路径中的传播的过程，我们直接由处于空间中特定位置的两路麦克采集到的语音信号自身体现，本实施例中不考虑外耳部分耳廓与外耳道对于声音信号的影响，则 HRTF 可以取 1。

[0074] 外围分析子模块 3023 包括以下单元：

[0075] 内耳耳蜗模拟单元 30231，用于将语音信号进行 gammatone 滤波分频，获取两路信号在 128 个不同子频带的语音信号；

[0076] 内耳非线性神经传导模拟单元 30232，用于将每个频率通道的子带信号进行非线性变换；

[0077] 中耳模拟单元 30233，用于通过对每个通道的 gammatone 滤波器的增益按照等响度曲线进行调整。

[0078] 这里所述的内耳耳蜗模拟单元 30231 主要起对语音信号进行分频的作用，它相当于一个带通滤波器组。具体来说，实际应用中采用 128 滤波器组成的 gammatone 滤波器组来进行模拟，不同的是它对于频带的划分是依据的人耳的听觉特性、采用等矩形带宽 ERB(equivalent rectangular bandwidth)，类似于 bark 频率，在低频有较小的带宽，在高频有较大的带宽。各个频率通道的带宽与中心频率值的关系如式 (17) 所示，其中 c 可以取 1 到 128，表示第 1 到 128 个频率通道，从而获取到两路信号在 128 个不同频带的语音信号。

[0079] $ERB(f_c) = 24.7(4.37f_c/1000+1)$ (17)

[0080] Gammatone 滤波器连续时域表达如式 (18) 所示。

[0081] $$g(c,t)=\begin{cases} t^{N-1}\exp(-2\pi b_c t)\cos(2\pi f_c t+\phi_c), & if \quad t>0 \\ 0, & else \end{cases} \tag{18}$$

[0082] 其中，c 为频率通道数，N 是滤波器的阶数，b_c 是与频率带宽相关的衰减因子，f_c 是第 c 个频率通道的中心频率，Φ_c 是相位值（取 0）. b_c 的计算如式 (19) 所示。

[0083] $b_c = 1.019ERB(f_c)$ (19)

[0084] 按式 (20) $y_1(n)$ 与 $y_2(n)$ 与对应的数字滤波器 g(c,n) 进行离散时间卷积处理后将分别得到两路信号经过 gammatone 滤波器滤波后的 128 个子带的信号：

[0085] $y_i(c,n) = y_i(n)*g(c,n)\ i = 1,2\ ;c = 1,2,\ldots,128$ (20)

[0086] 声音进入耳蜗分频后，两路信号的每个频率通道的子带信号将分别通过耳蜗内部的毛细胞感应转化为电信号由听神经传导的过程由 meddis 模型进行模拟（神经传导过程的非线性变换），其原理图如图 8 所示。其中，q(t) 是 t 时刻毛细胞内部的自由递质数；k(t) 是 t 时刻，递质从毛细胞穿透到突触间隙里的穿透率，即单位时间的穿透量；s(t) 是耳蜗毛细胞感受到的语音信号激励，对应于经过耳蜗滤波处理后得到的各个子带信号；c(t) 是突触间隙里的递质数。k(t) 与 s(t) 的关系如式 (21) 所示。

[0087] $$k(t)=\frac{g(s(t)+A)}{s(t)+A+B}\ k(t)=0 \qquad for \qquad s(t)+A<0 \tag{21}$$

[0088] 式 (21) 中的 g、A、B 是通过生物医学实验获得的参数。

[0089] 类似于扩散原理，递质在毛细胞与神经突触之间的运动有如下方面：毛细胞里边的递质可以由递质产生器不断得到新的补充，毛细胞里的递质可以以一定的穿透率被释放进入突触间隙，突触间隙里边的递质可以重新返回到毛细胞里边去，也可以损失掉或是以一定的概率激发神经元后突触，从而继续传递下去最终到大脑。由此得到 q(t) 和 c(t) 的微分方程，如式 (22) 和 (23) 所示。

[0090] $$\frac{dq(t)}{dt}=y[1-q(t)]+rc(t)-k(t)q(t) \tag{22}$$

[0091] $$\frac{dc(t)}{dt}=k(t)q(t)-lc(t)-rc(t) \tag{23}$$

[0092] 其中，y，l，r 是常数，表示对应运动过程单位时间的发生比率。突触间隙里边的递质以一定的概率激发神经元后突触，如式 (24) 所示，其中 h 是常数，与 y，l.r 单位相同。

[0093] Prob(event) = hc(t)dt(24)

[0094] 在处理离散的语音数字信号时，以上各式可以相应地转换成差分方程。dt 取采样周期值。这样在设定各个量的初值后，每来一个样点，则进行一次迭代，对应得到的各个样点由 (24) 式输出。

[0095] 此过程可等效为一个滤波器，记为 $g_{haircell}(n)$，则以两路不同子带的信号 $y_i(c,n)$ (i = 1,2) 经过毛细胞与神经纤维间的非线性传导后的输出 $h_i(c,n)$ (i = 1,2) 如式 (25) 所示：

[0096] $h_i(c,n) = y_i(c,n)*g_{haircell}(n) = y_i(n)*g(c,n)*g_{haircell}(n)$ (25) i = 1,2 ;c = 1,2,...,128

[0097] 中耳模拟单元 30233 通过对每个通道的 gammatone 滤波器的增益按照等响度曲线进行调整来实现模拟的，具体实现原理如式 (26)、(27)、(28) 所示。

[0098] phon = (loudnesslevelInphones(cf,loudFunc)-DB) (26)

[0099] DB = 60 (27)

[0100] $midEarCoeff = 10.0^{phon/200}$ (28)

[0101] 其中，loudnesslevelInphones(cf，loudFunc) 部分是一个函数，其值为对应一个等响度曲线 loudFunc 在中心频率 cf 处对应的一个响度值，单位为 phon. 由以上三式可以直接得到 128 个频率通道的中耳系数 midEarCoeff(c) (c = 1,2,...,128)，其作用方式可以直接将该系数附加到对应的频率通道的 gammatone 滤波器上，通过外围分析所有过程的两路信号的各子带信号仍以 $h_i(c,n)$ 表示，则最后的输出如式 (29)、(30) 所示：

[0102] $h_1(c,n) = midEarCoeff(c) \cdot y_1(n)*g(c,n)*g_{haircell}(n)$ c = 1,2,...,128(29)

[0103] $h_2(c,n) = midEarCoeff(c) \cdot y_2(n)*g(c,n)*g_{haircell}(n)$ c = 1,2,...,128(30)

[0104] 为便于后边的特征提取，需对上边两式中各个子带的信号进行分帧，得到第 c 个频率通道、第 m 个时间帧、第 n 个时间点的信号 $h_1(c,m,n)$ 与 $h_2(c,m,n)$。从而通过分频和分帧，获得两个二维的语音信号。

[0105] 特征提取子模块 3024，用于提取分频后两路语音信号的延时差和能量差信息；

[0106] 由两路麦克的空间拓扑结构图图 4 可以看出，噪声源特别是较远距离的噪声源同目标声源的延时差存在一定的差异，因此，我们需要提取的特征之一是两路麦克的延时差 (ITD)。此外，两路麦克的能量差别 (IID) 信息也能体现噪声源与目标声源之间的差异性，因而 ITD 与 IID 信息是我们需要提取出来的两个主要信息。

[0107] 耳间延时差 (Interaural Time Difference) 是指两路麦克的语音信号间的延时差值，由于两路麦克采集到的是带噪语音，这个延时差是各种环境噪声源发出的噪声与目标语音混合作用的结果。当噪声强度不大时，在目标声源发声的时候，它主要体现的是目标声源的延时值；如果与目标语音相比，噪声强度很大或是目标语音没有出现的时候，则它主要体现背景噪声的混合延时。由于不同频带的语音信号的延时会有微小的差异，对于同一时间帧，我们需要计算不同频带的延时差，前边的 gammatone 滤波等过程为这一过程做好了准备。

[0108] 通过求互相关的最大值可以求得延时差值。第 c 个频率通道、第 m 帧、延时为 τ 时的互相关可以通过式 (31) 求得，由式 (32) 可以求得第 c 个频率通道、第 m 时间帧的两路麦克信号的延时差值。

[0109]

$$Corr(c,m,\tau)=\frac{\sum_{n=0}^{L-1}(h_1(c,m,n)-\bar{h}_1)(h_2(c,m,n-\tau)-\bar{h}_2)}{\sqrt{\sum_{n=0}^{L-1}(h_1(c,m,n)-\bar{h}_1)^2}\sqrt{\sum_{n=0}^{L-1}(h_2(c,m,n-\tau)-\bar{h}_2)^2}} \tag{31}$$

[0110]

$$ITD(c,m)=\max_{\tau} Corr(c,m,\tau) \tag{32}$$

[0111] 其中，$h_1(c,m,n)$ 与 $h_2(c,m,n)$ 分别为外围分析输出后，在第 c 个频率通道、第 m 个时间帧的第 n 个采样点处的值，L 为 320，为一帧信号的样点数。$\bar{h}_1$ 与 $\bar{h}_2$ 为第 c 个频率通道、第 m 个时间帧的两路语音信号的平均值。

[0112] 耳间能量差 (Interaural Intensity Difference) 是指两路麦克语音信号间的能量比值，对于同一时间帧信号的不同频率通道，也需要分别计算 IID 值。第 c 个频率通道、第 m 时间帧的 IID 值可以通过式 (33) 求得。

[0113]

$$IID(c,m)=20\log_{10}\left(\frac{\sum_{n=0}^{L-1}h_1^2(c,m,n)}{\sum_{n=0}^{L-1}h_2^2(c,m,n)}\right) \tag{33}$$

[0114] 对于第 c 个频率通道、第 m 个时间帧的位置的语音信号，称为一个 T-F 单元 (时频单元)，ITD(c,m) 与 IID(c,m) 均是由两个通道语音信号的第 c 个频率通道、第 m 个时间帧对应时频单元的两帧信号计算出来的。对于每一个时频单元，对应有一个 ITD(c,m) 跟一个 IID(c,m) 信息，它们是我们后边选择出目标声源语音信息的依据。

[0115] 生成掩蔽子模块 3025，根据提取的延时差和能量差信息获得不同时频区域对应的掩蔽值；

[0116] 当声音信息最终由神经传递至大脑后，由人脑根据声音的有关信息，如 ITD，IID，

IED(耳间信号包络差别)等 cue 信息,对声音进行选择性分离处理。

[0117] CASA 在模拟完声音进入人耳的过程之后,通过两路语音信号的 T-F 二维语音信息计算提取出各 T-F 单元的 ITD 与 IID 等 cue 信息,然后利用声学掩蔽效应进行语音分离。

[0118] 声学掩蔽效应是一种心理声学现象,它是指在一个较强的声音附近,相对较弱的声音将不被人耳察觉,即被强音所掩蔽。声学掩蔽分为同时掩蔽与异时掩蔽,我们采用同时掩蔽。

[0119] 我们先是着眼于每一个 T-F 单元进行掩蔽处理。对于某一个 T-F 单元,如果目标声源的强度大于背景噪声的强度,则认为在这个 T-F 单元目标声源能够将背景噪声掩蔽掉,我们保留这个 T-F 单元信息或是给予一个较大的权值;反之,则认为不能掩蔽掉背景噪声,我们去掉这一块儿语音信息或是给予其一个较小的权值。为尽可能地保留目标声源语音信息,我们采用加权而不采用或取或舍的形式。对于目标语音与噪声强度比越大的 T-F 单元,我们给越大的权值;反之,如果越小,给越小的权值。

[0120] 每一个 T-F 单元的信号中的目标语音与噪声的强度比是无法直接得到的,正如前面分析,我们可以根据 ITD 和 IID 信息间接得到。当目标语音强度占优时,ITD 与 IID 主要体现的是目标语音到两路麦克的 ITD 与 IID 值,即实际的 ITD 与 IID 值会偏向目标语音单独作用时的 ITD 与 IID 值;反之,ITD 与 IID 会偏向背景噪声单独作用时的 ITD 与 IID 值。据此关系,我们给定加权掩蔽值的原则是,对于某个 T-F 单元,若它的 ITD 或是 IID 越接近目标语音单独作用时的 ITD 或是 IID 值,我们给予较大的权值;若是 ITD 或是 IID 值越远离目标语音单独作用时的 ITD 或是 IID 值时,我们给予较小的权值。

[0121] 如果每一个 T-F 单元都按如上进行掩蔽处理,这样的总体效果就是目标声源得到保留或是增强,噪声得到了抑制,从而将目标语音从带噪语音中分离出来,达到了去噪效果。

[0122] 对于不同的频率通道使用 ITD 与 IID 值进行掩蔽值处理的错误率是不相同的,ITD 在低频带(较小的频率通道数)有较低的错误率,IID 在高频带(较大的频率通道数)会有较低的错误率。经过实验,我们选取了 1500HZ(对应第 67 个频率通道)作为 ITD 与 IID 使用的分界线。

[0123] 优选的,所述双麦克语音增强模块 302 还包括以下子模块:

[0124] 语音激活检测子模块,用于检测纯环境噪声语音段。

[0125] 为进一步提高无目标语音段的噪声抑制效果,我们需要使用 VAD 模块的判决结果,对于由 VAD 模块判定为无语音段的语段,这些语段的每一帧的每个频率通道,即这些语段的每个 T-F 单元,我们都直接给予较小的掩蔽权值,实现对噪声段的直接抑制。

[0126] 这里,VAD 模块充分利用了两通道语音信息,克服了普通 VAD 无法适用于复杂多变的噪声环境的局限性,对于各种类型的噪声的带噪语音均能得到有效检测,可以与去噪算法很好地配合应用。在去噪处理时引入 VAD 模块可以在检测出无目标语音说话语段时对带噪信号进行直接的抑制,使总体的信噪比大大提高。外围分析原理如前所述,采用计算机算法的形式模拟实现人耳对声音的外围分析过程。特征提取主要提取的是两路语音的延时差信息 (Interaural Time Difference, ITD) 与能量差信息 (Interaural IntensityDifference, IID)。生成掩蔽过程是对于已得到的 ITD 与 IID 信息,进行计算得到掩蔽值。掩蔽值平滑模块是对已得到的各个通道的掩蔽值进行滤波,以去除掩蔽估计值

的野点，获取更好的听觉效果。语音重建是对于各个频带的语音信号进行重建，以得到处理后的时域语音信号。

[0127] 此 VAD 模块采用两路语音信号进行 VAD 检测，其原理如图 7 所示，第 1 路与第 2 路分别为离目标声源的近端和远端麦克风。其中 $x_1(n)$ 与 $x_2(n)$ $(n \geq 0)$ 是由采集部分采集到的两路数字语音信号，采样率为 16KHZ，采样精度为 16bit，两路信号 VAD 判断时以帧为单位进行处理，每一帧时长是 20ms，帧移是 10ms. $x_1(m)$ 与 $x_2(m)$ $(m \geq 0)$ 是两路带噪语音第 m 帧语音信号，每一帧帧长是 L，即 $x_1(m) = x_1[mL, \ldots, mL+L-1]$，$x_2(m) = x_2[mL, \ldots, mL+L-1]$。$\sigma_1(m)$ 与 $\sigma_2(m)$ 是经过迭代更新、平滑处理后的两路带噪语音第 m 帧的能量谱。$\lambda_1(m)$ 与 $\lambda_2(m)$ 是估算的两路带噪语音信号中的噪声能量谱。$r_1(m)$ 与 $r_2(m)$ 是第 m 帧的两路后验信噪比。

[0128] 具体的，(a) 初始化噪声能量谱值，前 5 帧认为是噪声，即 VAD 的值设为 0。将离目标声源稍远的麦克采集到的语音信号的前 5 帧能量平均值作为两路噪声能量谱初值。初始化两路带噪语音的能量谱为第 6 帧两路带噪语音的能量谱值。

[0129] (b) 迭代更新。从第 6 帧开始每来一帧都分别对两路带噪语音能量谱 $\sigma_1(m)$ 和 $\sigma_2(m)$ 进行迭代更新，更新方法如式 (1) 所示；如果当前帧的前一帧的 VAD 判决结果为噪声，即判决结果为 0，则对当前帧的两路噪声能量谱 $\lambda_1(m)$ 与 $\lambda_2(m)$ 进行迭代更新，更新方法如式 (2) 所示：

[0130] $\sigma_i(m) = \alpha |x_i(m)|^2 + (1-\alpha)\sigma i(m-1)$，$i = 1,2$ (1)

[0131] $$\lambda_i(m) = \begin{cases} \beta |X_i(m)|^2 + (1-\beta)\lambda_i(m-1), & VAD(m-1) = 0 \\ \lambda_i(m-1) & else \end{cases} \qquad i = 1,2 \qquad (2)$$

[0132] 两个迭代因子 α 和 β 应分别取 0.9 与 0.01，分别取较大值与较小值是用于跟踪快速变化的带噪语音的变化趋势与噪声的缓变趋势。

[0133] (c) 计算后验信噪比。两路带噪语音能量谱 $\sigma_1(m)$ 与 $\sigma_2(m)$ 中均包含两部分能量，即目标语音能量部分与背景噪声能量部分，$\sigma_1(m)$ 与 $\sigma_2(m)$ 能量分解表达式如式 (3) 与式 (4) 所示。其中，$\lambda_d(m)$ 与 $\lambda_x(m)$ 分别表示第 1 路语音信号中的噪声能量部分与目标声源能量部分。g 是由两麦克不同的性能差异造成的，p 是由两路麦克到目标声源不同的距离引起的目标声源能量的差异，l_1 与 l_2 是两路麦克到目标声源的距离如式 (5) 所示。

[0134] $\sigma_1(m) = \lambda_d(m) + \lambda_x(m)$ (3)

[0135] $\sigma_2(m) = g(\lambda_d(m) + p\lambda_x(m))$ (4)

[0136] $$p = \left(\frac{l_1}{l_2}\right)^2 \qquad (5)$$

[0137] $\lambda_1(m)$ 与 $\lambda_2(m)$ 是对 $\sigma_1(m)$ 与 $\sigma_2(m)$ 中噪声能量部分的估计，由于迭代速度的不同等原因会与 $\sigma_1(m)$ 与 $\sigma_2(m)$ 中的噪声能量部分略有差异，表示成如下式 (6) 与式 (7) 所示。

[0138] $$\lambda_1(m) = \overline{\lambda_d(m)} \qquad (6)$$

[0139] $$\lambda_2(m) = g\overline{\lambda_d(m)} \qquad (7)$$

[0140] 两路后验信噪比计算如下式 (8)、(9) 所示。

[0141] $$r_1(m)=\frac{\sigma_1}{\lambda_1}=\gamma(m)+\xi(m) \quad (8)$$

[0142] $$r_2(m)=\frac{\sigma_2}{\lambda_2}=\gamma(m)+p\xi(m) \quad (9)$$

[0143] 其中，$$\gamma(m)=\frac{\lambda_d(m)}{\lambda_d(m)} \quad (10)$$

[0144] $$\xi(m)=\frac{\lambda_x(m)}{\lambda_d(m)} \quad (11)$$

[0145] (d) 作差比较。将两后验信噪比作差，再设定一个合适的阈值（如可设为 1）并与之进行比较，若差值大于该阈值，则认为第 m 帧有目标语音，否则认为第 m 帧是纯噪声段。

[0146] $u(m) = r_1(m) - r_2(m) = (1-p)\,\xi(m)$ (12)

[0147] 掩蔽值平滑子模块 3026，用于对提取的掩蔽值进行平滑处理；

[0148] 由于我们的掩蔽处理是对于一个个 T-F 单元分别进行处理的，因而，相邻的 T-F 单元之间处理后会存在一定的不连续性，所以我们需要对掩蔽值进行平滑处理。

[0149] 通过对掩蔽值进行频谱分析，其频率成份主要集中在 10HZ 以下。我们设计了一个数字滤波器，其 3db 带宽为 10HZ，截止频率为 50HZ，在 50HZ 以后幅值下降到 80db。

[0150] 语音信号重建子模块 3027，用于对由波束形成模块得到的参考信号进行掩蔽值处理，并合成增强后的语音信号作为输出。

[0151] 信号的重建过程框图如图 9 所示，用于对波束形成模块的输出信号进行掩蔽值处理，并合成增强后的语音信号作为输出。主要步骤如下：

[0152] 依次进行 gammatone 滤波器组滤波、各个频率通道信号反折、再一次 gammatone 滤波、再一次各个频率通道信号的反折。这个过程的主要作用是对语音信号进行 gammatone 滤波，分频为 128 个频率通道的同时，增加两次反折处理与一次 gammatone 处理来消除 gammatone 滤波器组对于不同频通道的延时的影响。

[0153] 分帧、加窗。分帧时每帧 20ms 时长，帧叠 10ms，加窗时采用汉明窗。

[0154] 掩蔽值处理。对于分频、分帧后得到的语音信号的每一个 T-F 单元，采用前边得到的对应于这个 T-F 单元的掩蔽值进行相乘处理。

[0155] 重叠相加，再累加。先将各个频率通道的不同的时间帧信号进行重叠相加，得到该频率通道的重建信号，再对 128 个频率通道的语音信号对应样点累加，得到最后语音频带的重建信号，作为最后的输出结果。

[0156] 以上对本发明所提供的一种语音增强的系统进行了详细介绍，本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

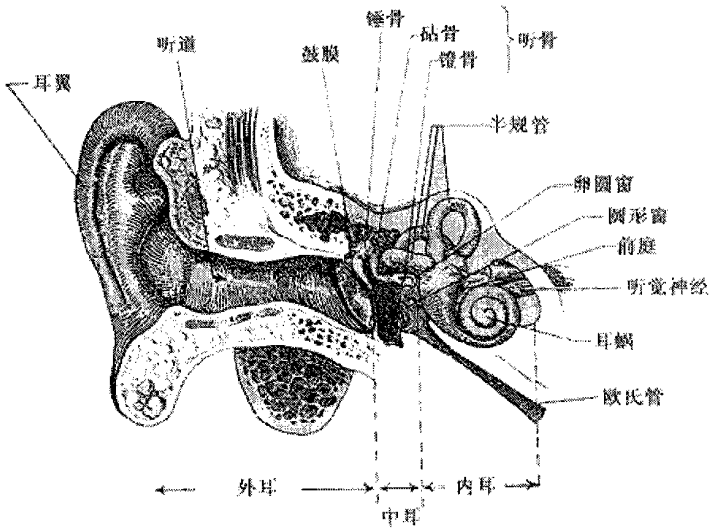

耳翼
听道
鼓膜
锤骨
砧骨
镫骨
听骨
半规管
卵圆窗
圆形窗
前庭
听觉神经
耳蜗
欧氏管
外耳
中耳
内耳

图 1

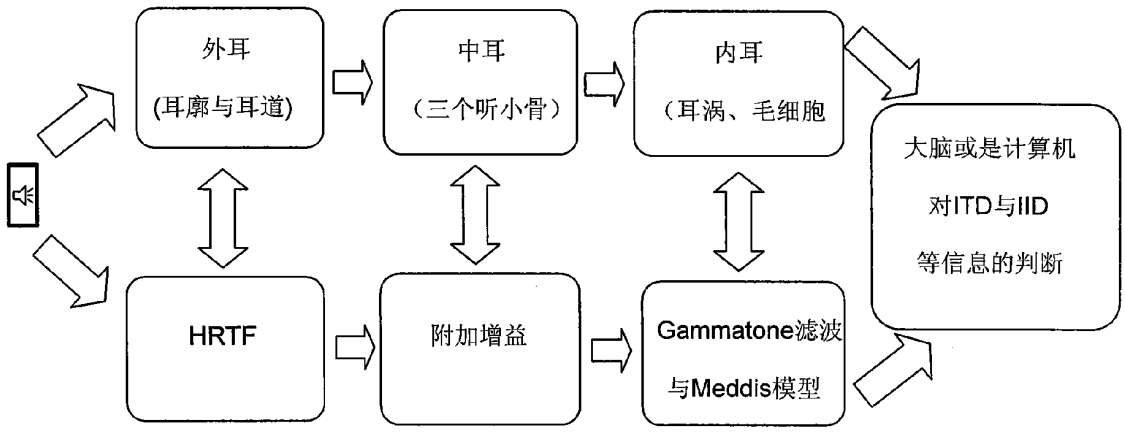

外耳
(耳廓与耳道)
中耳
（三个听小骨）
内耳
（耳涡、毛细胞
大脑或是计算机
对ITD与IID
等信息的判断
HRTF
附加增益
Gammatone滤波
与Meddis模型

图 2

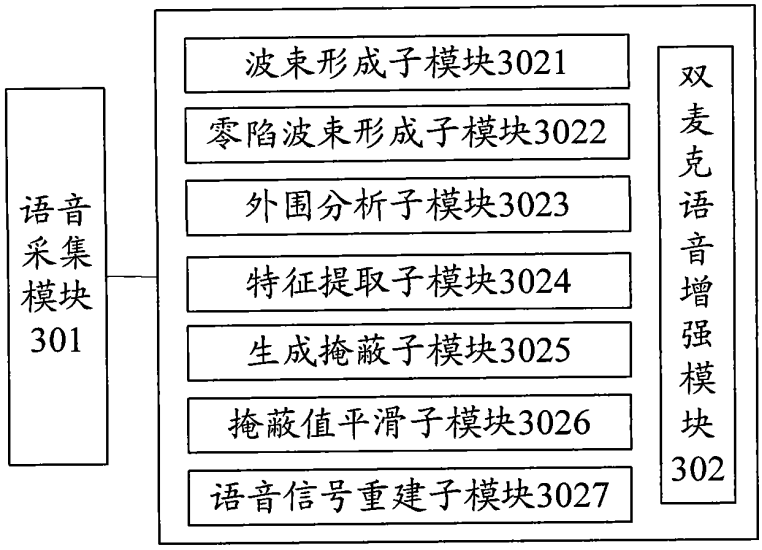

语音采集模块301
波束形成子模块3021
零陷波束形成子模块3022
外围分析子模块3023
特征提取子模块3024
生成掩蔽子模块3025
掩蔽值平滑子模块3026
语音信号重建子模块3027
双麦克语音增强模块302

图 3

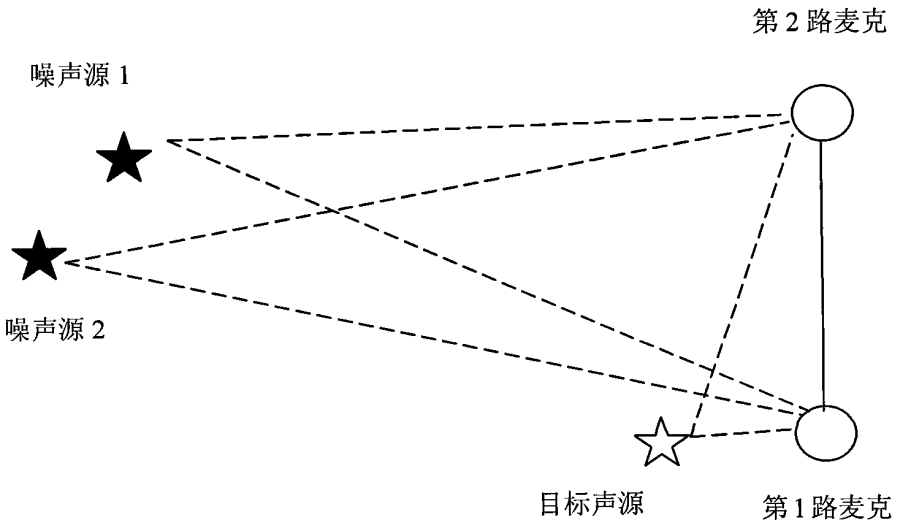

第 2 路麦克
噪声源 1
噪声源 2
目标声源
第 1 路麦克

图 4

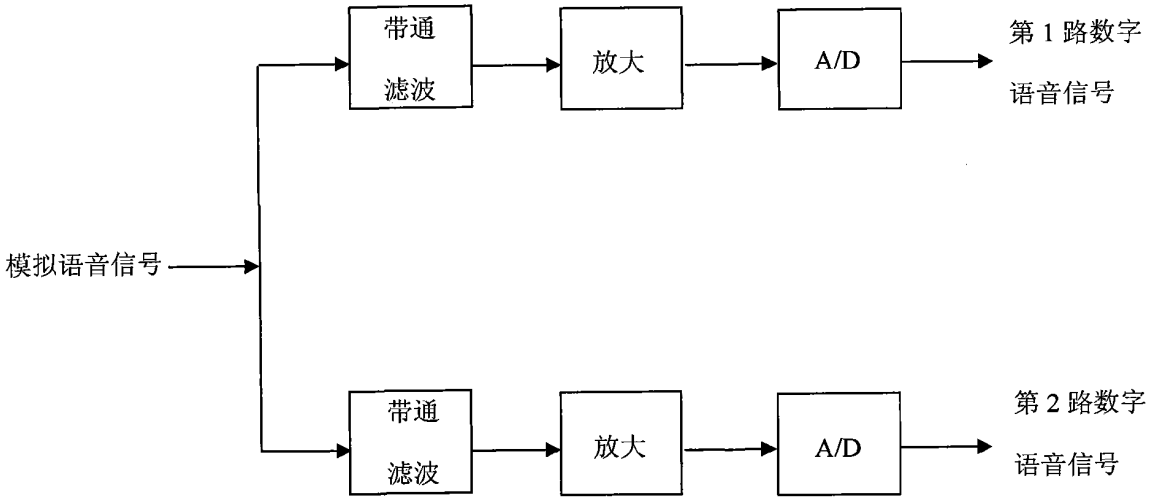

带通
滤波
放大
A/D
第 1 路数字
语音信号
模拟语音信号
带通
滤波
放大
A/D
第 2 路数字
语音信号

图 5

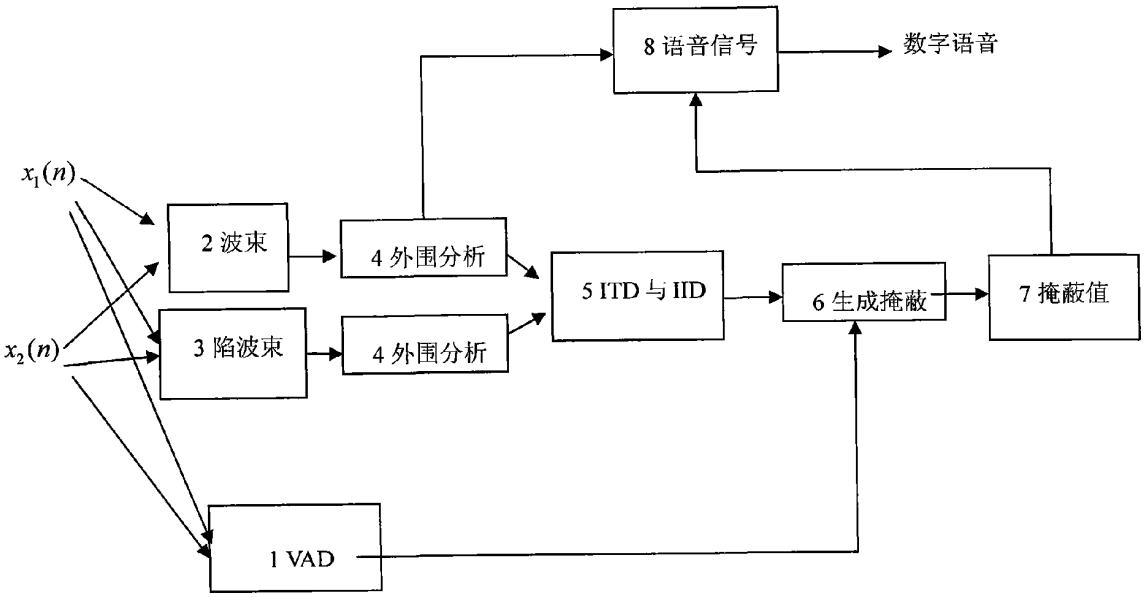

8 语音信号
数字语音
$x_1(n)$
2 波束
4 外围分析
5 ITD 与 IID
6 生成掩蔽
7 掩蔽值
$x_2(n)$
3 陷波束
4 外围分析
1 VAD

图 6

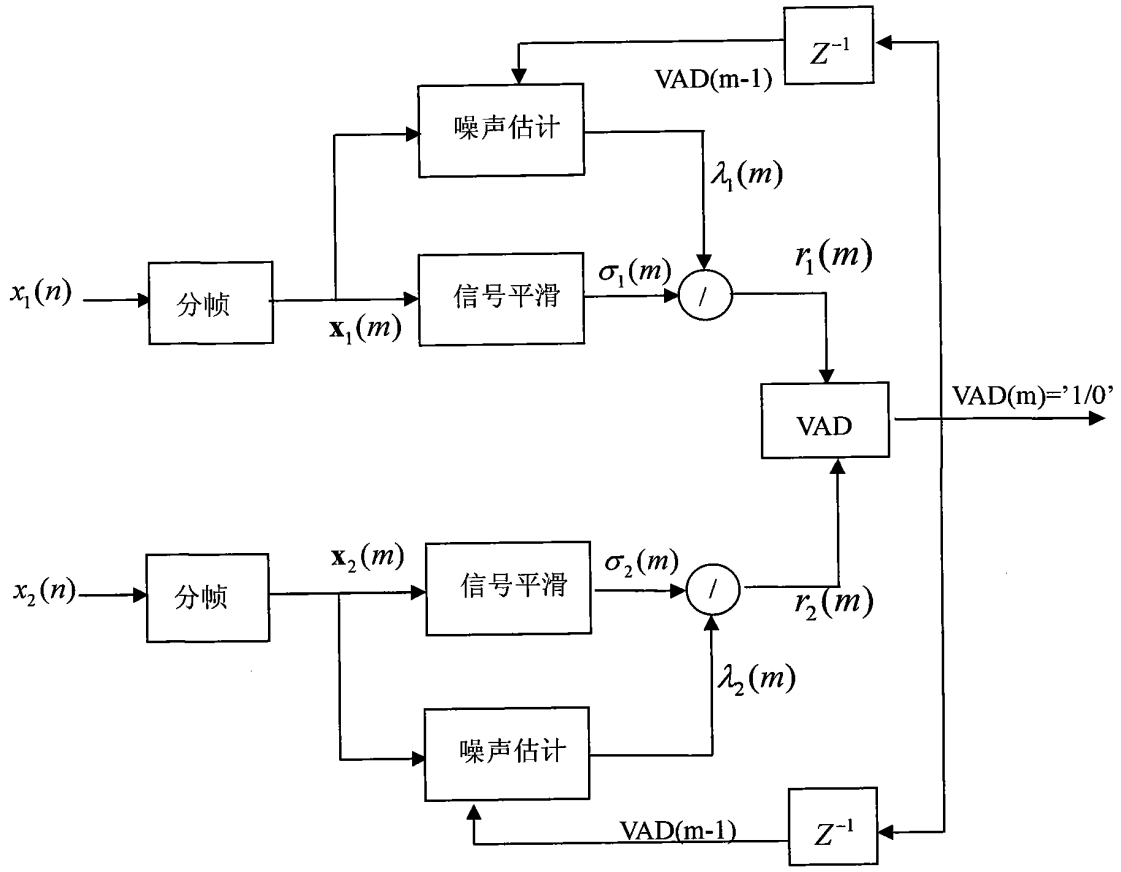

Z^{-1}
VAD(m-1)
噪声估计
$\lambda_1(m)$
$x_1(n)$
分帧
$\mathbf{x}_1(m)$
信号平滑
$\sigma_1(m)$
/
$r_1(m)$
VAD
VAD(m)='1/0'
$x_2(n)$
分帧
$\mathbf{x}_2(m)$
信号平滑
$\sigma_2(m)$
/
$r_2(m)$
$\lambda_2(m)$
噪声估计
VAD(m-1)
Z^{-1}

图 7

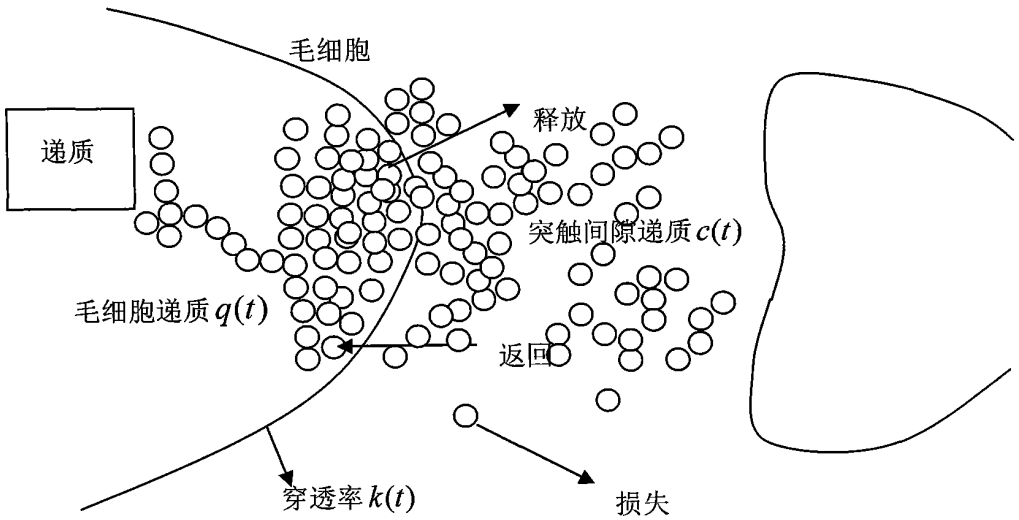

毛细胞
递质
释放
突触间隙递质 $c(t)$
毛细胞递质 $q(t)$
返回
穿透率 $k(t)$
损失

图 8

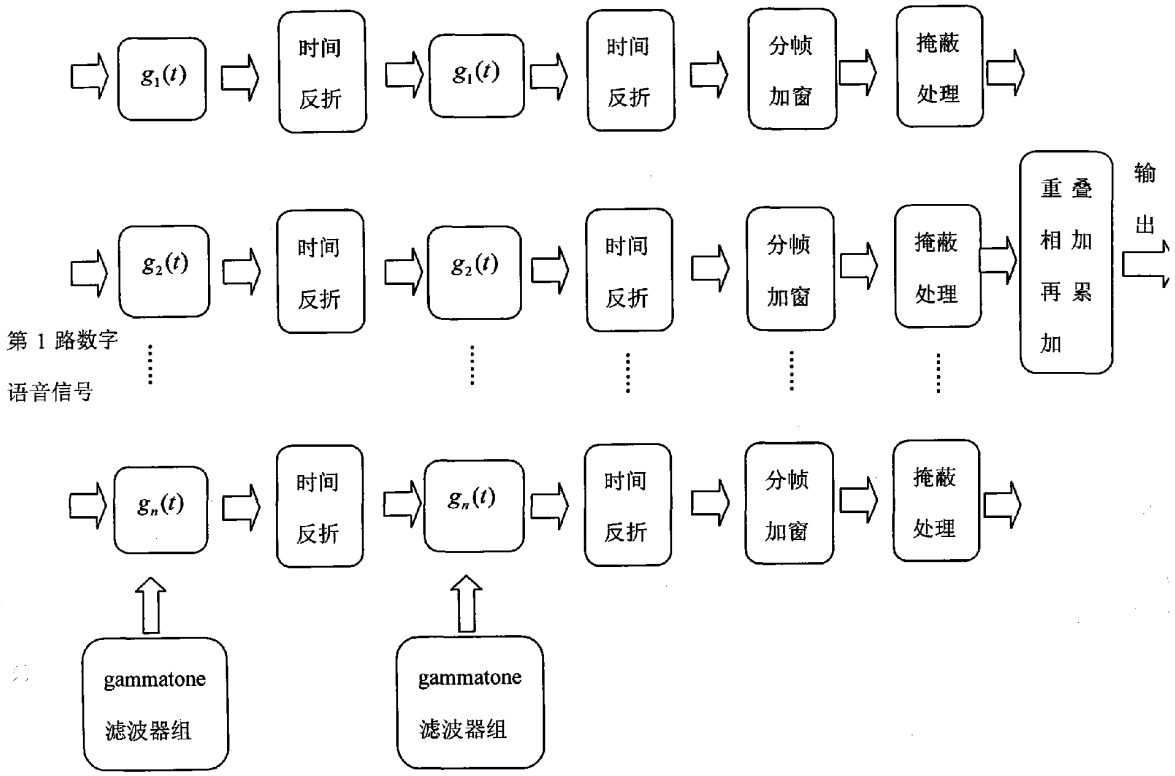

$g_1(t)$
时间反折
$g_1(t)$
时间反折
分帧加窗
掩蔽处理
$g_2(t)$
时间反折
$g_2(t)$
时间反折
分帧加窗
掩蔽处理
重叠相加再累加
输出
第 1 路数字语音信号
$g_n(t)$
时间反折
$g_n(t)$
时间反折
分帧加窗
掩蔽处理
gammatone 滤波器组
gammatone 滤波器组

图 9