基于麦克风阵列的语音增强系统 V2.0 设计说明书

大连理工大学

北京华捷艾米科技有限公司

2021 年 4 月 1 日

目录

一、引言

1.1 编写目的

本说明书介绍了语音增强系统 V2.0 的设计方案，该版本对 V1.0 版本中的算法进行了改进，为读者详细说明了该系统的总体设计思路和各个模块的设计思路，旨在让每一位读者了解该系统的设计以及学会使用该系统。

1.2 背景

随着语音设备的更新换代和语音增强技术的快速发展，手机、智能音箱等语音设备对减少噪声干扰和降低语音失真的要求越来越高。为了满足语音设备的语音增强需求，大连理工大学通信与信号处理实验室和北京华捷艾米有限公司的员工共同开发了语音增强系统 V1.0，该版本的系统有一定的语音增强能力，但是在噪声干扰较大的情况下，增强后的语音仍然残留了较多的干扰，鉴于此，我们改进了语音增强的算法，开发了语音增强系统 V2.0，该系统能在各种平台上运行，为语音设备和语音识别引擎提供噪声干扰小且失真小的语音。

1.3 定义

AEC：Adaptive Echo Canceller，自适应回波消除
NLP：None Linear Processing，非线性处理
SRP：Steered Responser Power，可控功率响应
DoA：Direction of Arrival，波达方向
VAD：Voice Activity Dectection，语音活动检测
PMWF：Parameterized Multichannel Wiener Filter，参数化的多通道维纳滤波
ANC：Adaptive Noise Canceller，自适应噪声消除
AGC：Automatic Gain Control，自动增益控制
PCM：Pulse Code Modulation，脉冲编码调制
NLMS：Normalized Least Mean Square，归一化最小均方误差
SNR：Signal Noise Ratio，信号噪声比
SIR：Signal Interference Ratio，信号干扰比
kHz：kiloHertz，千赫兹
dB：deciBel，分贝
T60：Reverberation time，混响时间
ms：millisecond，毫秒

1.4 参考资料

(1) 韩纪庆，张磊，郑铁然. 语音信号处理.第 2 版[M]. 清华大学出版社，2013.
(2) 谭颖，殷福亮，李细林. 改进的 SRP-PHAT 声源定位方法[J]. 电子与信息学报，2006，28(7)：1223-1227.
(3) 王冬霞，殷福亮. 基于近场波束形成的麦克风阵列语音增强方法[J]. 电子与信息学报，2007，29(1)：67-70.
(4) Hansler E , Schmidt G U . Hands-free telephones – joint control of echo cancellation and postfiltering[J]. Signal Processing, 2000, 80(11):2295-2305.
(5) Hoshuyama O , Sugiyama A , Hirano A . A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters[J]. Proc IEEE Icassp96 May, 1996, 47(10):2677-2684.
(6) Herbordt W , Buchner H , Nakamura S , et al. Multichannel Bin-Wise Robust Frequency-Domain Adaptive Filtering and Its Application to Adaptive Beamforming[J]. IEEE Transactions on Audio, Speech and Language Processing, 2007, 15(4):1340-1351.

二、总体设计

2.1 需求规定

2.1.1 主要功能

对麦克风阵列采集到的信号做声学回声消除和语音增强，定位出主声源的方位角，增强主声源方向的语音信号，抑制噪声和其他方向的干扰，并且减少在语音处理过程中产生的语音失真。

2.1.2 系统输入

输入为六路语音信号以及一路背景回声参考信号，其中六路语音信号是由均匀分布的环形六元麦克风阵列采集的，阵列模型如图 1 所示，语音信号的采样频率为 16kHz，室内混响 T60 为 200ms，噪声选用白噪声，信噪比为 0dB，回声参考信号要接入外部扬声器以保证被麦克风阵列采集到。

2.1.3 系统输出

输出是经过语音增强后的单路语音信号以及系统定位出的主声源方位角信息，其中方位角的单位是度。

2.1.4 性能需求

要求经系统处理后的信号的回声比输入语音信号减弱 40dB 以上，主声源方向信号的信噪比 SNR 提高 30dB 以上，信干比 SIR 提高 20dB 以上，系统定位出的主声源方位角与实际主声源的方位角的偏差不超过 3°。

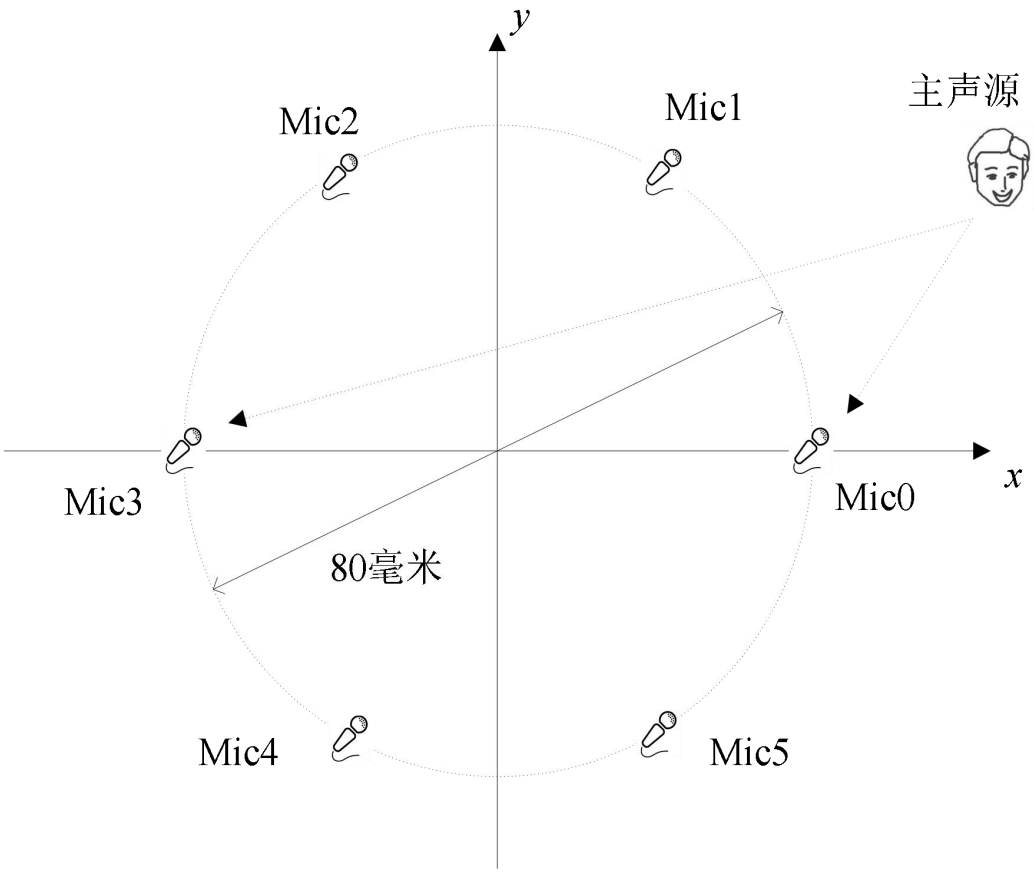

图 1. 环形六元麦克风阵列模型

2.2 运行环境

(1) 硬件环境：内存 1GB 以上，硬盘 1GB 以上，Yamaha Steinberg 声卡
(2) 软件环境：Windows XP/7/10 操作系统、Linux 操作系统
(3) C/C++编译器环境：MSVC 编译器或 GNU 编译器

2.3 系统总体框架

系统总体框架图如图 2 所示：

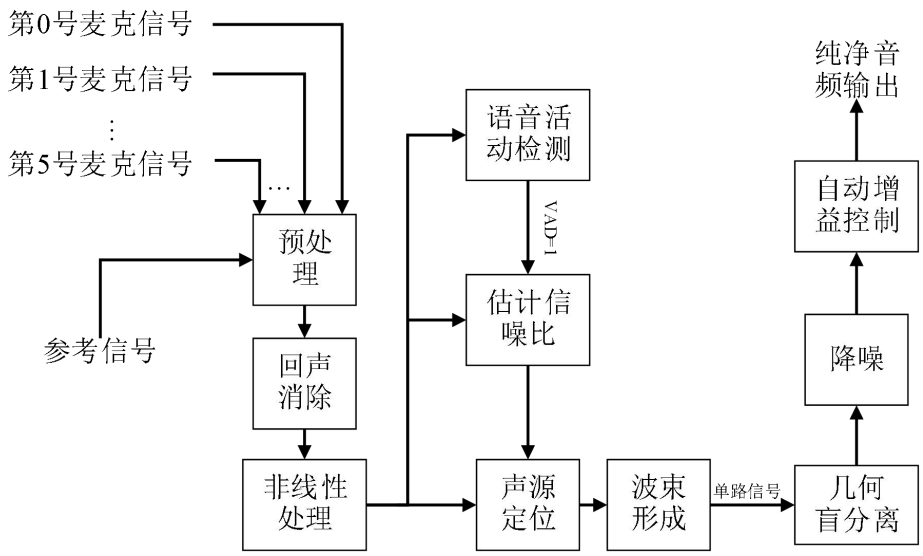

图 2. 系统总体框架图

2.4 系统处理流程

本系统包括九个模块：预处理、回声消除、非线性处理、语音活动检测、信噪比估计、声源定位、波束形成、几何盲分离、噪声抑制、自动增益控制。基本设计思路如下：

1、对六路麦克风采集信号及一路参考信号进行预处理（分帧、加窗、子带分解）

2、对预处理后的六路麦克风信号采用子带 NLMS 算法做回声消除，衰减麦克风信号中的回声参考信号成分

3、对回声消除后的信号进行非线性处理，进一步衰减非线性部分的回声参考信号

4、对回声消除后的帧信号用长短时能量法进行语音活动检测，标记该信号帧是否为语音帧，若为语音帧，估计当前帧的信噪比

5、若当前帧的信噪比大于 15dB，则采用 SRP 算法实现声源定位，找到声源相对于麦克风阵列的方位角

6、利用声源定位出的方位角并采用 PMWF 算法对回波消除后的信号做波束形成，六路信号经过波束形成后合成为单路信号

7、波束形成输出的单路信号利用积累的位置信息进行几何盲分离

8、几何盲分离后的信号采用维纳滤波法进行噪声抑制处理

9、降噪后的语音再通过自动增益控制使信号的增益保持一致，得到最终的输出。

2.5 算法改进说明

(1) 与 V1.0 相比，V2.0 在回声消除之后增加了非线性处理模块，能够进一步消除非线性部分的残留回声，提高了系统的回声消除能力

(2) 与 V1.0 相比，V2.0 改进了语音活动检测模块的算法，提高了检测正确率

(3) 与 V1.0 相比，V2.0 在声源定位之前增加了信噪比估计模块，当信号满足一定的信噪比条件时进行定位，提高了定位结果的准确率

(4) 与 V1.0 相比，V2.0 改进了波束形成模块的算法，用 PMWF 算法代替 GSC 算法，提高了干扰的消除量

2.6 系统输入输出说明

(1) 输入：麦克风阵列采集到的语音信号通过声卡转换为数字信号传入 PC 机中，作为系统的输入，声卡对语音信号的采样频率为 16kHz，采样精度为 16 位。回声参考信号要由外部放音设备输出，同时接入声卡和扬声器。

(2) 输出：系统的输出是采样频率为 16kHz，采样精度为 16 位的单路数字信号，由 PC 机通过声卡传出。

2.7 主要宏定义

Word16：C 语言中的短整型类型
Word32：C 语言中的长整型类型
Float32：C 语言中的单精度浮点类型
FRM_LEN：待处理信号的帧长
NUM：麦克风的数量
DIAMETER：麦克风阵列的直径
SPEED：声音在空气中的传播速度
FS：输入信号的采样频率
T60：混响时间
D：子频带的数量
ORD2：原型低通滤波器的阶数
AZ_NUM：声源方位角的数量
AZ_STEP：方位角的粗搜索步长
PI：保留 15 位有效数字的圆周率
UPDATE_THLD：VAD 结果更新阈值

DEV_THLD：频谱偏差阈值

2.8 主要结构体定义

AEC_SRP_ST：包含回声消除、声源定位、波束形成模块所需变量的结构体
NS_STRUCT：包含噪声抑制模块所需变量的结构体
Agc_t：包含自动增益控制模块所需变量的结构体

2.9 主要函数接口说明

(1) aec_srp_pmwf_init

函数功能	参数初始化
函数原型	void aec_srp_pmwf_init(AEC_SRP_ST *st,Float32 appointed_pitch)
输入参数	st：结构体 AEC_SRP_ST 的地址 appointed_pitch：固定俯仰角，取值（-PI/2，PI/2）
返回值	无

(2) aec_srp_pmwf

函数功能	实现回声消除、声源定位和波束形成
函数原型	Void aec_srp_pmwf(AEC_SRP_ST *st,Word16 mic_sp[NUM][FRM_LEN],Word16 spk_sp[FRM_LEN], Word16 out_sp[FRM_LEN])
输入参数	mic_sp[NUM][FRMLEN]：6 路麦克风输入信号 spk_sp[FRM_LEN]：参考信号 out_sp[FRM_LEN]：输出信号
返回值	无

(3) NS_run

函数功能	噪声抑制的执行函数

函数原型	void NS_run(NS_STRUCT* st, short* speechFrame, short* outFrame);
输入参数	st：待处理结构体的地址 speechFrame：输入信号帧的首地址 outFrame：输出信号帧的首地址
返回值	无

(4) get_position

函数功能	输出主声源方位角的函数
函数原型	float get_position(AEC_SRP_ST *st）
输入参数	st：结构体 AEC_SRP_ST 的地址
返回值	单精度的方位角弧度

(5) fft_320

函数功能	实现 320 点的快速傅里叶变换
函数原型	void fft_320(Float32 *tmp1, Float32 *tmp2)
输入参数	tmp1：输入序列的实数部分 tmp2：输入序列的虚数部分
返回值	无

(6) Matrix_Mul

函数功能	实现两个实数矩阵相乘
函数原型	void Matrix_Mul(double a[NUM1][NUM1], double b[NUM1][NUM1], double c[NUM1][NUM1], int n)
输入参数	a[NUM1][NUM1]:左乘矩阵 b[NUM1][NUM1]:右乘矩阵 n:相乘矩阵的阶数
返回值	无

(7) INV

函数功能	实现一个实数矩阵的求逆

函数原型	void INV(double A[NUM][NUM], int n)
输入参数	A[NUM][NUM]:待求逆矩阵 n:求逆矩阵的阶数
返回值	无

(8) inverse

函数功能	实现一个复数矩阵的求逆
函数原型	void inverse(double A[NUM][NUM], double B[NUM][NUM], double C_re[NUM][NUM], double C_im[NUM][NUM], int n)
输入参数	A[NUM][NUM]:待求逆矩阵的实部 B[NUM][NUM]:待求逆矩阵的虚部 C[NUM][NUM]:逆矩阵的实部 D[NUM][NUM]:逆矩阵的实部 n:求逆矩阵的阶数
返回值	无

(9) AGC_Init

函数功能	自动增益控制的初始化函数
函数原型	int AGC_Init(void *agcInst, int32_t minLevel, int32_t maxLevel, int16_t agcMode, int32_t fs);
输入参数	minLevel：信号的最小增益 maxLevel：信号的最大增益 agcMode：AGC 处理模式 fs：输入信号的采样频率
返回值	无

(10) AGC_Process

函数功能	自动增益控制的实时处理函数
函数原型	int AGC_Process(Agc_t *st, const int16_t *in_near, const int16_t *in_near_H, int16_t samples, int16_t *out, int16_t *out_H, int32_t inMicLevel, int32_t *outMicLevel, int16_t echo, uint8_t *saturationWarning);
输入参数	st：待处理 AGC 结构体 in_near：输入语音帧数据 samples：输入数据长度 out：处理后的语音帧数据

返回值	无

三、各模块设计

3.1 预处理

3.1.1 模块功能

本系统预处理模块的功能是对输入信号重叠分帧、加窗、缓存，然后进行子带分解，将时域信号转换为包含 320 个子频带的子带信号。

3.1.2 处理流程

如图 3 所示，该模块的处理流程如下：

(1) 初始化分析滤波器系数，并且初始化各个子带的复指数权重向量

(2) 输入六路麦克风信号和一路参考信号，更新信号的缓存区

(3) 将缓存区的信号通过分析滤波器进行滤波然后对滤波后的信号加权求和，得到 2D 个子带信号

(4) 输出前 D 个子带信号

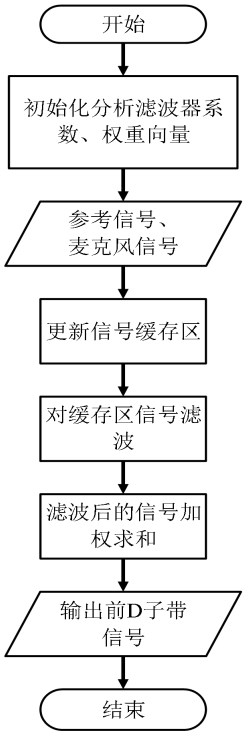

图 3. 预处理模块流程图

3.1.3 参数设置

输入信号的帧长为 160，分析滤波器的阶数为 1920，子带的个数为 320，每个子带输出一个信号，相当于在每个子带对输入信号做了 160 倍下采样。

关键变量在程序中的变量名：

(1) 输入子带信号：mic_ana_re、mic_ana_im

(2) 参考子带信号：spk_ana_re、spk_ana_im

(3) 分析滤波器系数：prototype_filter

3.2 回声消除

3.2.1 模块功能

本系统回声消除模块的功能是抑制麦克风采集信号中的背景音乐回声干扰，在保证主声源信号较小失真的条件下使用子带域 NLMS 算法自适应地消除回声，为后续的处理提供无回声的语音信号。

3.2.2 处理流程

如图 4 所示，该模块的处理流程如下：

(1) 初始化各个子带的滤波器系数

(2) 输入子带分解后的麦克风信号和参考信号，并且更新参考子带信号缓存区

计算麦克风信号和参考信号的幅度谱和帧能量，估计麦克风信号和参考信号之间的皮尔森相关系数。

(3) 将参考信号通过子带自适应滤波器得到模拟回波信号，用预处理后的 0 号麦克信号减去模拟回声信号得到第一路回波消除后的误差信号。

(4) 根据皮尔森相关系数的数值选择自适应滤波器的更新步长，然后利用误差信号和参考信号更新自适应滤波器的系数。

(5) 由于模拟回波信号由参考信号通过自适应滤波器所得，故其更接近麦克风接收到的参考信号，因此，将模拟回波信号信号作为第二至六路的参考信号，对预处理后的 1-5 号麦克信号按同样的方法做回声消除以及自适应滤波器的更新。

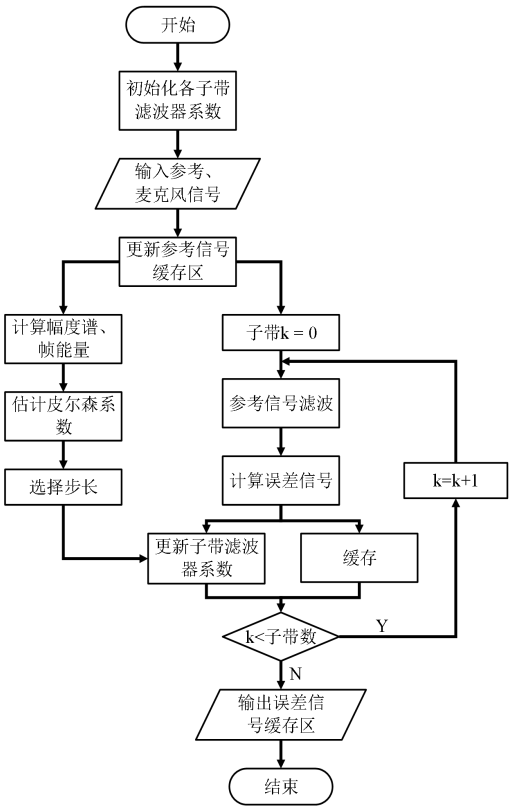

图 4. 回声消除模块流程图

3.2.3 参数设置

子带自适应滤波器的阶数设置为 20，根据混响时间得出，可以抵抗高达 200ms 的混响。

关键变量在程序中的变量名：

(1) 皮尔森系数：pearson

(2) 滤波器更新步长：aec_step

(3) 自适应滤波器系数：h_r、h_i，初始化为 0

(4) 误差子带信号：e_r、e_i

3.3 非线性处理

3.3.1 模块功能

本系统非线性处理模块的功能是检测回声消除后信号中非线性部分的残留回声，利用相关系数法判断信号的单双讲情况并抑制残留的回声信号。

3.3.2 处理流程

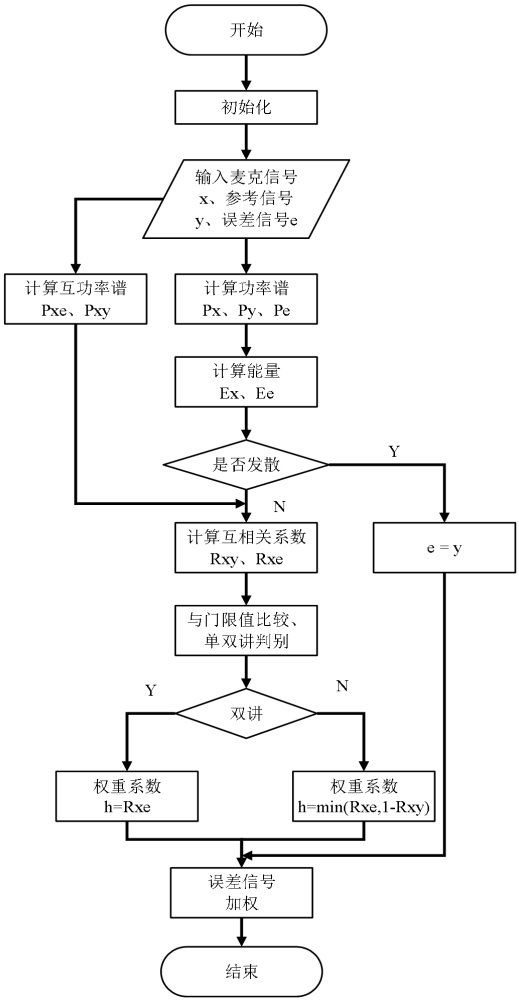

图 5. 非线性处理模块流程图

如图 5 所示，该模块的处理流程如下：

(1) 初始化抑制权重系数、功率谱和互功率谱

(2) 计算麦克风信号、参考信号、误差信号的功率谱和互功率谱

(3) 计算麦克风信号和参考信号、误差信号之间的互相关系数

(4) 在感兴趣的子带上对互相关系数求平均，得到平均互相关系数

(5) 用平均互相关系数与门限值比较，得到信号的单双讲情况

(6) 计算各子带的抑制权重系数，然后对误差子带信号加权得到输出信号

3.3.3 参数设置

更新功率谱的权重因子设置为 0.8，相当于 5 帧求平均，感兴趣的子带设置为频率为 300-3400Hz（人耳能够听到的频率范围）对应的子带。

关键变量在程序中的变量名：

(1) 输入信号和参考信号的互功率谱：mic0_spk_power_re 、

mic0_spk_power_im，初始化为 0

(2) 输入信号和误差信号的互功率谱：mic0_e0_power_re、mic0_e0_power_im，初始化为 0

(3) 更新功率谱的权重因子：ptrGCoh

(4) 平均互相关系数：hNl_mic0_spk_Avg、hNl_mic0_e0_Avg

(5) 子带抑制系数：hNl

3.4 语音活动检测

3.4.1 模块功能

本系统语音活动检测模块的功能是根据信号在 16 个临界频带上的长时能量和短时能量判断信号是否为语音信号。

3.4.2 处理流程

如图 6 所示，该模块的处理流程如下：

(1) 计算信号在临界频带上的总信噪比、总能量和对数能量偏差

(2) 初始化 update_flag 为 False，若总信噪比 vm_sum 小于阈值，则更新 update_flag 为 True，并将 update_cnt 清零，当前帧判为噪声帧，否则，进入(3)

(3) 若总能量 tce 大于噪声能量门限并且对数能量偏差 ch_enrg_dev 小于阈值（即：当前帧为能量变化较小的语音帧），累加 update_cnt，若 update_cnt 大于 50（即：连续 50 帧能量变化较小），则更新 update_flag 为 True，当前帧判为噪声帧，否则，不更新 update_flag，当前帧为语音帧

(4) 若连续 6 帧 update_cnt 不变，则将 update_cnt 清零

3.4.3 参数设置

更新平均能量的平滑系数设置为 0.55，总信噪比的阈值为 35，总能量的阈值为 1.0，对数能量偏差的阈值为 28.0。

关键变量在程序中的变量名：

(1) 输入信号能量：ch_enrg，对数能量：ch_enrg_db，长时对数能量：ch_long_enrg_db

(2) 信噪比：ch_snr，量化后的总信噪比：vm_sum，信号总能量：tce

(3) 状态更新标志：update_flag（True 代表噪声、False 代表语音），不更新帧的计数：update_cnt

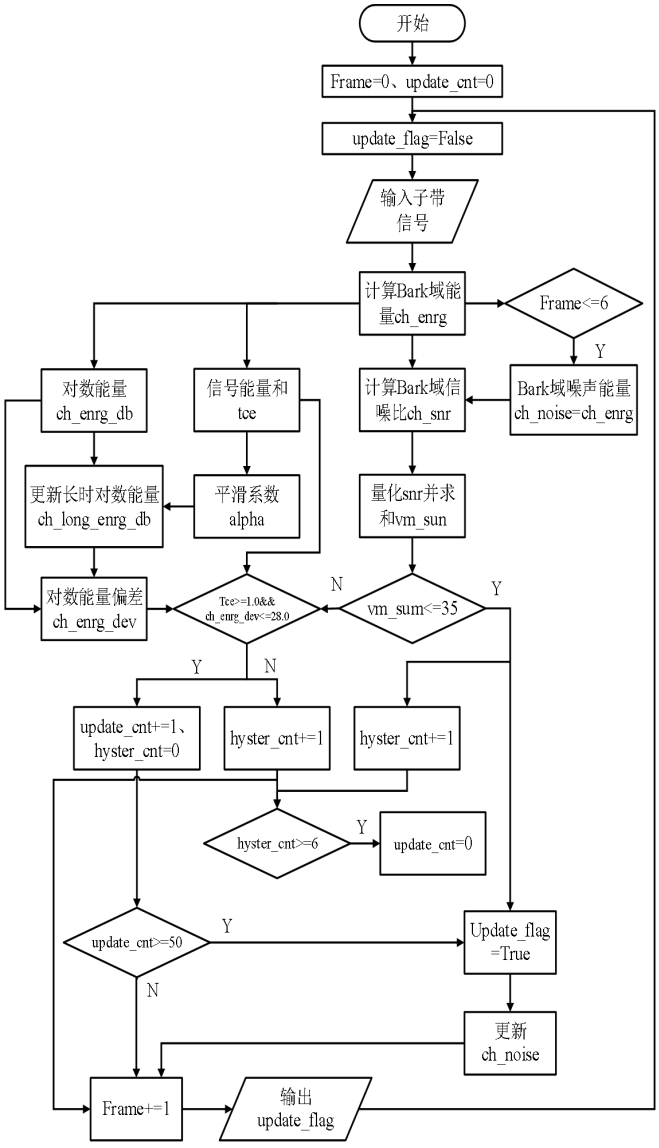

图 6. 语音活动检测模块流程图

3.5 信噪比估计

3.5.1 模块功能

本系统信噪比估计模块的功能是根据连续多帧信号的功率和 VAD 检测结果估计出噪声功率，进而计算出信号的信噪比。

3.5.2 处理流程

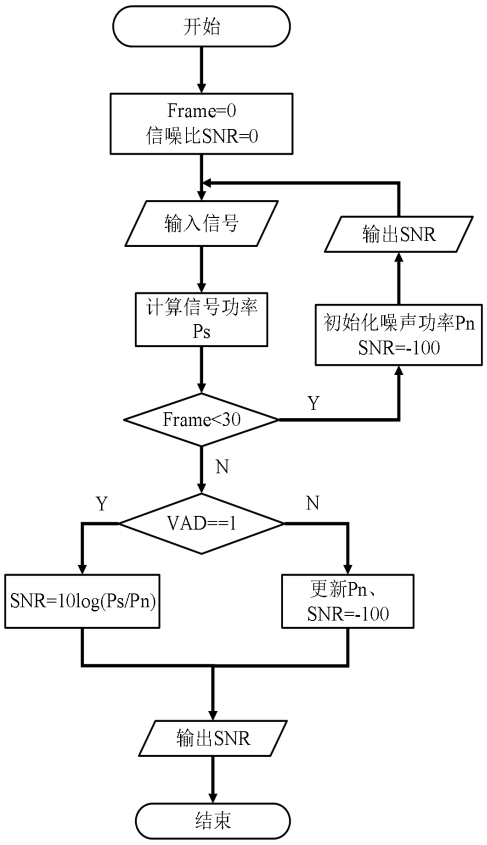

图 7. 信噪比估计模块流程图

如图 7 所示，该模块的处理流程如下：

(1) 初始化信噪比 SNR 为零，输入非线性处理后的第一路子带信号

(2) 计算信号功率，用前 30 帧信号初始化噪声功率并且令 SNR=-100

(3) 若当前帧检测为语音信号，计算 SNR，否则，令 SNR=-100 并更新噪声功率

(4) 输出当前帧信号的 SNR

3.6 声源定位

3.6.1 模块功能

本系统声源定位模块的功能是根据各个方向的波束输出功率搜索出声源相对于麦克风阵列的方位角，此方位角在几何上定义为声源和麦克风阵列中点的连线与 mic0 和 mic3 连线的夹角。

3.6.2 处理流程

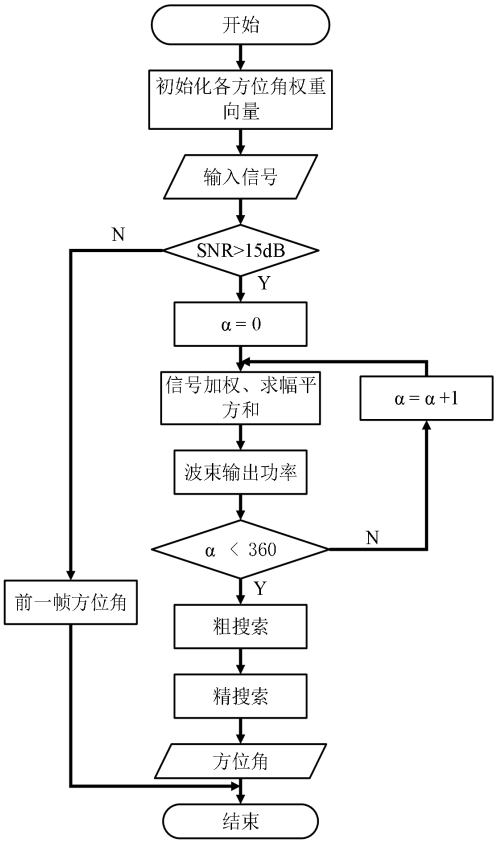

图 8. 声源定位模块流程图

如图 8 所示，该模块的处理流程如下：

(1) 初始化方位角为整数（1°-360°）时的权重矩阵

(2) 若当前帧的信噪比大于 15dB，对回波消除后的 6 路麦克子带信号加权求和得到各个指向各个方向的单路信号，否则，保留上一帧的定位结果。

(3) 对单路信号的每个子带求幅平方然后求和得到总的波束输出功率。

(4) 以 18°为步长对方位角在 0-360°范围内进行粗搜索，搜索出波束输出功率最大时的角度为β。

(5) 以 1°为步长在（β-9，β+9）范围内进行精搜索得到波束输出功率最大的角度，该角度即为估计出的声源相对于麦克风阵列的方位角。

3.6.3 参数设置

粗搜索的步长设置为 18°，精搜索的步长设置为 1°，这样可以覆盖 0-360°范围内的所有整数角度。

关键变量在程序中的变量名：

(1) 各个方位角对应的权重系数：wr、wi

(2) 波束形成输出信号功率的最大值：peak

(3) 最大功率对应的方位角：tot_az

3.7 波束形成

3.7.1 模块功能

本系统波束形成模块采用 PMWF 算法进行波束形成，增强波达方向的语音信号，抑制其他方向上的噪声和干扰。

3.7.2 处理流程

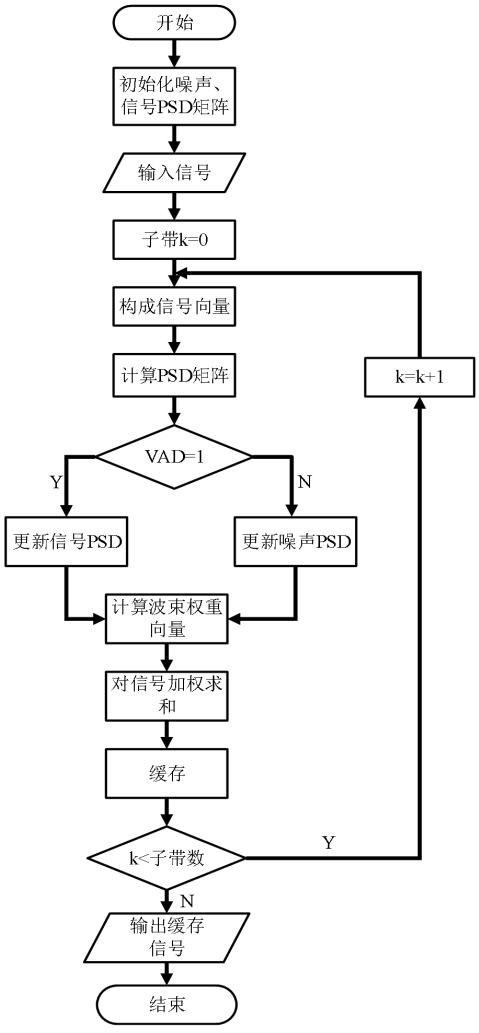

图 9. 波束形成模块流程图

如图 9 所示，该模块的处理流程如下：

(1) 初始化每个子带的噪声 PSD 矩阵和信号 PSD 矩阵。

(2) 将回声消除后的六路信号逐子带组成信号向量，计算当前帧信号的子带 PSD 矩阵

(3) 若当前帧检测为语音帧，则更新每个子带的信号 PSD 矩阵，若当前帧检测为噪声帧，则更新每个子带的噪声 PSD 矩阵

(4) 利用估计出的噪声 PSD 矩阵和信号 PSD 矩阵求出权重向量，对输入信号向量逐子带加权求和得到波束输出信号

3.7.3 参数设置

更新协方差矩阵的权重因子α设置为 0.8，相当于 5 帧求平均。

关键变量在程序中的变量名：

(1) 噪声协方差矩阵：qnn_re、qnn_im，初始化为单位矩阵

(2) 输入信号协方差矩阵：qxx_re、qxx_im，初始化为零矩阵

(3) 更新协方差矩阵的权重因子：alpha

3.8 噪声抑制

3.8.1 模块功能

本系统噪声抑制模块的功能是利用维纳滤波法抑制语音信号中的加性噪声，在保证语音失真小的条件下提高语音质量。

3.8.2 处理流程

如图 10 所示，该模块的处理流程如下：

(1) 初始化维纳滤波器的系数以及噪声模型，对输入信号做离散傅里叶变换得到频域信号

(2) 用前 50 帧信号构建噪声模型，50 帧以后的信号估计信号的先验信噪比、后验信噪比

(3) 计算信号似然比并更新平均对数似然比，计算信号谱平坦度和信号、噪声频谱差异

(4) 根据对数似然比和谱平坦度计算语音/噪声概率，当噪声概率超过阈值时，更新噪声谱估计值

(5) 计算维纳滤波器的系数，然后对输入信号进行滤波得到噪声抑制后的频

域信号，最后将频域信号经过离散傅里叶反变换得到时域输出信号。

3.8.3 参数设置

离散傅里叶变换的点数设置为 256，对数似然比的平滑因子设置为 0.9，相当于 10 帧求平均。

关键变量在程序中的变量名：

(1) 输入信号幅度谱：magn

(2) 语音概率：probSpeechFinal

(3) 维纳滤波器系数：theFilter

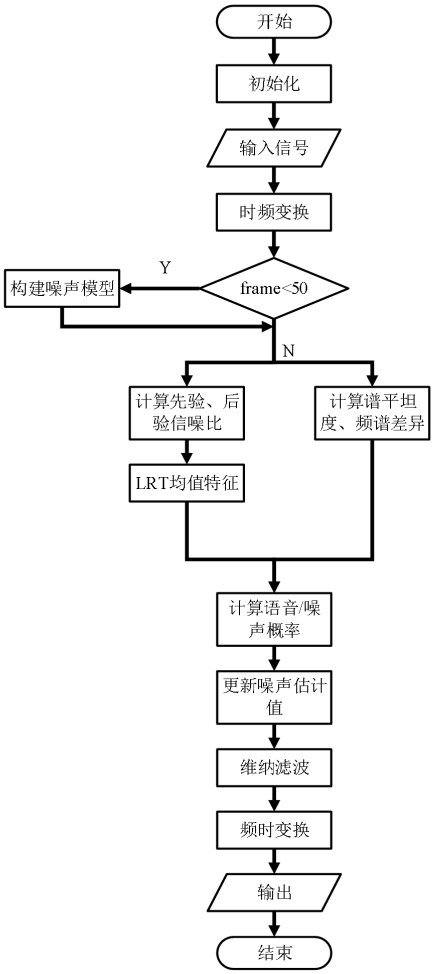

图 10. 噪声抑制模块流程图

3.9 自动增益控制

3.9.1 模块功能

本系统自动增益控制模块的功能是使输出语音信号的强度保持一致，提高语

音质量。

3.9.2 处理流程

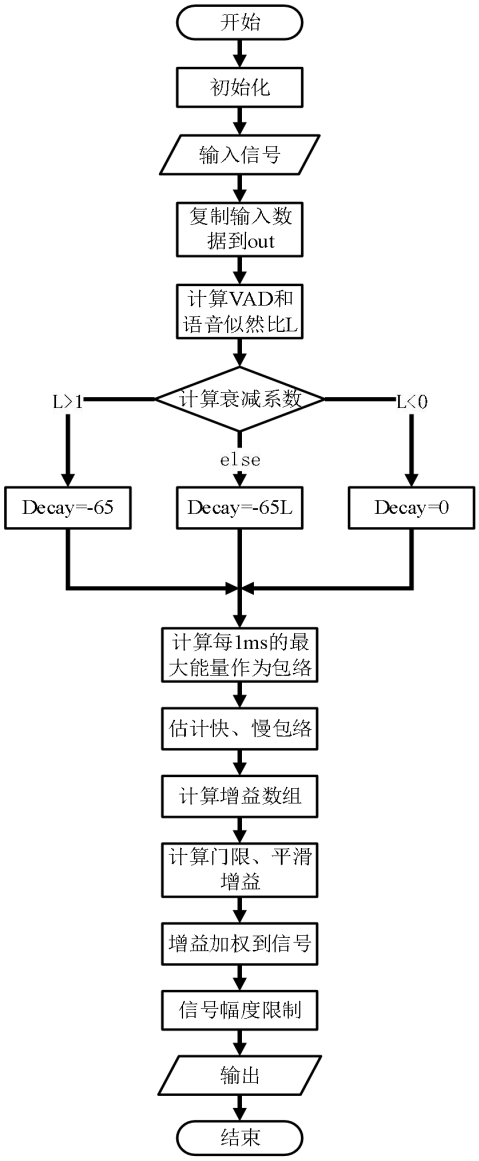

图 11. 自动增益控制模块流程图

如图 11 所示，该模块的处理流程如下：

(1) 初始化增益系数，复制输入数据到 out

(2) 计算语音 VAD 参数和语音似然比，通过似然比确定衰减系数 decay

(3) 计算每 1ms（10 个采样点）的最大能量作为能量包络，确定快谱、慢谱包络，通过快慢包络和增益表计算出增益数组 gain

(4) 计算并平滑门限 gate，用 gate 修正增益数组

(5) 对输入信号的每个采样点使用 gain，用当前点和上一点 gain 的差值加权

到输入信号

四、操作说明

(1) 用本系统要求的麦克风阵列和声卡进行音频信号采集。

(2) 用 C/C++编译器对源代码进行编译，得到可执行文件，运行可执行文件即可执行本系统，对麦克风阵列采集到的语音进行语音增强。

(3) 系统的输出可由 PC 机通过声卡直接传输到外部，也可以保存为本地文件