RNNoise代码解读 - 特征提取

RNNoise是一个采用音频特征抽取+深度神经网络结合的降噪方案. 更多相关基本信息, 请查看 [RNNoise学习和翻译系列](https://blog.csdn.net/danteLiujie/article/details/102769708)

特征提取部分的代码其实也很清爽, 主要有两个函数需要理一理, 因为代码行数太多

# 1.读取文件生成特征的主循环

Denoise.c 中的main函数是特征提取部分的主流程.

循环之外的内容有:

* 数据对象管理
* 命令行参数分析
* 文件操作
* 跳过噪音开头的帧

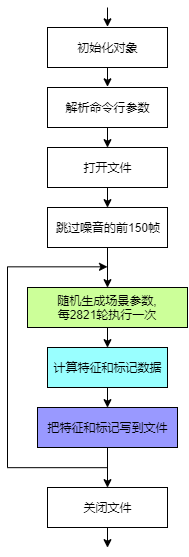
主循环主要做了三件事:

* 定期生成随机参数(参考第二节)
* 特征和标记提取运算(参考第三节)
* 把特征数据和标记数据写到文件

这部分代码逻辑来说就是: 循环读取语音和噪音, 混成训练数据, 并产生对应的特征和标记数据.

其中, 数据片大小是20ms, 步进速度是10ms, 有10ms的重叠数据.

其框图如下:



# 2. 构造不同场景和条件的训练

一个有效的模型, 为了能在现实场景下工作的很好, 必须要刷题, 也要实战, 这意味着训练用的数据必须花样百出. 一种最基本的考虑是使用不同SNR的语音来进行训练, 实验表明使用SNR范围较广的训练数据得到的训练模型对不同的SNR有更好的支持度(参考[链接](https://arxiv.org/abs/1808.10620)的3.1节 SNR维度)

本模型中采用了以下几种变化参数:

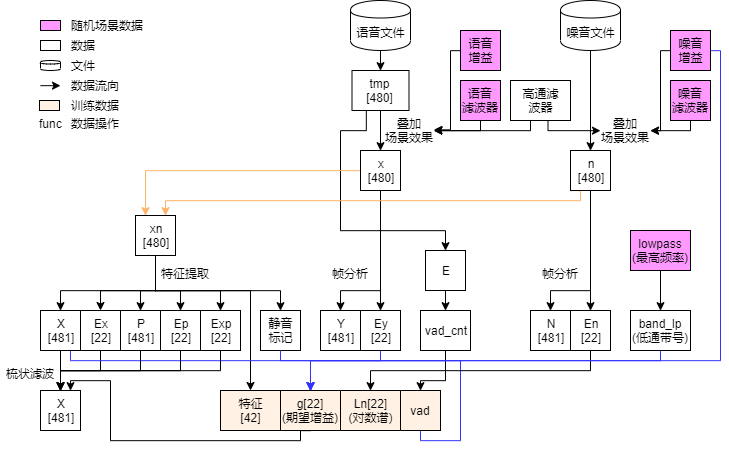
* 语音增益 和 噪声增益, 这两者合起来可以表示不同的SNR还可以表示不同的音量范围
* 两个随机的二阶滤波器, 模拟各种特殊的声学场景
* 最高截止频率(lowpass), 用来模拟各种采用频率的音频

各取值的范围:

* 语音增益: -40~20dB, 10%概率变为静音
* 噪音增益: 语音增益\*-30~20dB, 10%的概率不叠加噪音(这个10%和前面的10%乘数叠加的)
* 滤波器系数: 0~0.75
* Lowpass: FREQ\_SIZE\*50rand(1)/8, 当FREQ\_SIZE为481时, 取值范围大概是60~3000

# 3. 特征和标记提取代码

该部分代码的数据流向图如下, 从中我们可以发现, 执行梳状滤波那个函数是没必要的:



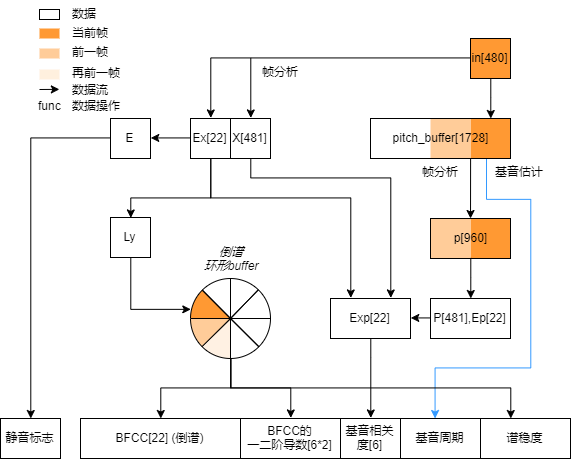
其中比较关键的子函数有:

* 特征提取(参考第四节)
* 帧分析: 计算fft和频带能量
* 梳状滤波: (本文先不涉及)
* 叠加场景效果: 根据随机产生的场景参数来叠加效果.

其中的帧分析和特征提取部分都会缓存部分数据. 比如帧分析会缓存一帧的数据, 加上当前帧, 共20ms的数据用.

# 4. 特征提取代码

该部分代码的数据流向图如下:



其中比较关键的子函数有:

* 特征计算: 此处使用了一个环形buffer来计算导数和谱稳度
* 基音估计: (本文先不涉及)
* 帧分析: 如上节