课程答疑

主讲人 宋 辉

清华大学电子工程系 博士
滴滴AI Labs 语音技术部

- AEC相关
- athena-signal
- 解混响
- 一些大家感兴趣的话题

AEC相关： 时延估计

将经过FFT后的频域信号，划分为32个子带。

对于参考信号，用75个32-bit数组存放参考信号历史信息。

对于近端信号，用16个32-bit数组存放近端信号历史信息。

“1”的数量越少，表明近端信号与参考信号越吻合，二者的时延越接近真实时延。

调整参考信号实现对齐。

Step1： spectrum → bitmap

$$\begin{cases} if\ |X(k,l)|^2 \geq thr_x, & then\ X(k,l) = 1 \\ if\ |X(k,l)|^2 < thr_x, & then\ X(k,l) = 0 \end{cases}$$

$$\begin{cases} if\ |Y(k,l)|^2 \geq thr_y, & then\ Y(k,l) = 1 \\ if\ |Y(k,l)|^2 < thr_y, & then\ Y(k,l) = 0 \end{cases}$$

Step2： XOR

Step3： count the number of 1

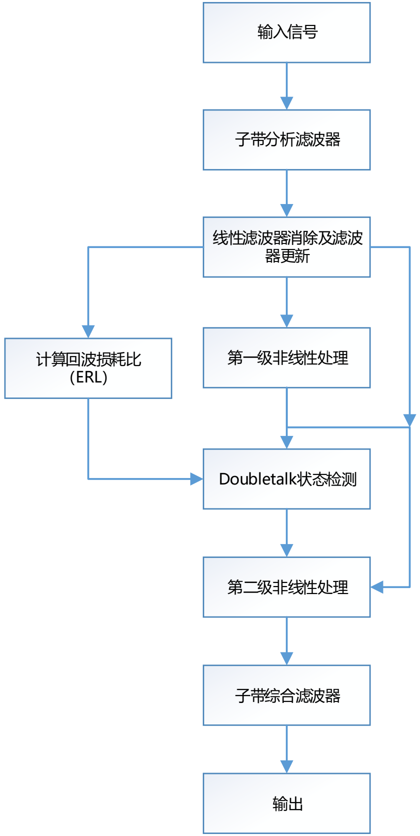
Step4： first-order markvo model to obtain final "delay sequence "

AEC相关： 回声消除的过程

需要注意两点：

1、非线性处理（残余回声抑制）分为两级，为了获取更为准确的DTD状态结果。在第一级非线性处理中，利用统一的标准对信号进行非线性处理，利用第一级的处理结果进行DTD。

2、在得到doubletalk状态之后，在第二级非线性处理中，针对不同的状态，利用不同的非线性处理等级，达到更好的消除效果以及尽可能降低语音信号的失真。



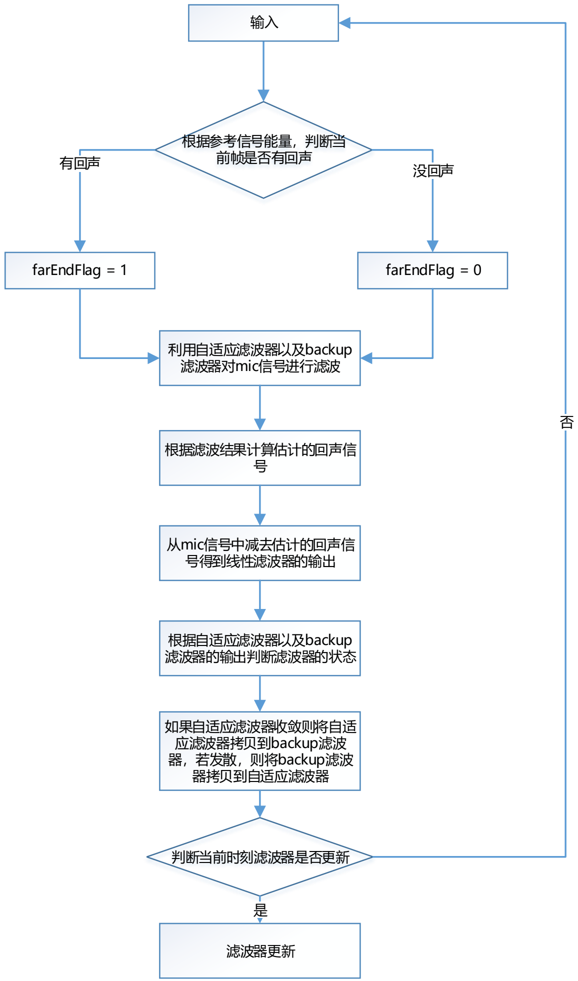
AEC处理流程

AEC相关： 线性回声消除

为了保证滤波器的稳定工作，引入了backup滤波器（fir_coef）与自适应滤波器（adf_coef）之间相互拷贝的机制。

滤波器系数更新时，要求回波信号的能量的信噪比要足够高，而且此时参信号不能存在截幅的风险。

- dios_ssp_aec_common.c
- dios_ssp_aec_common.h
- dios_ssp_aec_doubletalk.c
- dios_ssp_aec_doubletalk.h
- dios_ssp_aec_erl_est.c
- dios_ssp_aec_erl_est.h
- dios_ssp_aec_firfilter.c
- dios_ssp_aec_firfilter.h
- dios_ssp_aec_macros.h



AEC相关： 线性回声消除

滤波器收敛、发散状态判断条件是通过对比滤波器的输出与原始mic的能量大小。

（1）如果自适应滤波器的输出能量远大于mic能量，则滤波器已经发散，应该把自适应滤波器系数重置为0。

（2）如果自适应滤波器的输出能量大于backup滤波器的能量，则应该将backup滤波器系数拷贝到自适应滤波器。

（3）反之亦然。

```c
/* filter convergence detection */
if (srv->mse_adpt[ch] > srv->mse_mic_in[ch] * MSE_RATIO_OUT_IN)
{
    for (i_ref = 0; i_ref < srv->ref_num; i_ref++)
    {
        for (i = 0; i < srv->num_main_subband_adf[ch]; i++)
        {
            srv->adf_coef[i_ref][ch][i].r = 0.0;
            srv->adf_coef[i_ref][ch][i].i = 0.0;
        }
    }
    srv->mse_mic_in[ch] = 0.0;
    srv->mse_adpt[ch] = 0.0;
    srv->mse_main[ch] = 0.0;
}
else if ((srv->mse_mic_in[ch] > srv->mse_adpt[ch] * MSE_RATIO_OUT_IN)
        && (srv->mse_adpt[ch] < FILTER_COPY_FAC * srv->mse_main[ch]))
{
    for (i_ref = 0; i_ref < srv->ref_num; i_ref++)
    {
        for (i = 0; i < srv->num_main_subband_adf[ch]; i++)
        {
            srv->fir_coef[i_ref][ch][i] = srv->adf_coef[i_ref][ch][i];
        }
    }
    srv->mse_mic_in[ch] = 0.0;
    srv->mse_adpt[ch] = 0.0;
    srv->mse_main[ch] = 0.0;
}
```

```c
if (srv->mse_main[ch] > srv->mse_mic_in[ch] * MSE_RATIO_OUT_IN)
{
    for (i_ref = 0; i_ref < srv->ref_num; i_ref++)
    {
        for (i = 0; i < srv->num_main_subband_adf[ch]; i++)
        {
            srv->fir_coef[i_ref][ch][i].r = 0.0;
            srv->fir_coef[i_ref][ch][i].i = 0.0;
        }
    }
    srv->mse_main[ch] = 0.0;
    srv->mse_adpt[ch] = 0.0;
    srv->mse_mic_in[ch] = 0.0;
}
else if ((srv->mse_mic_in[ch] > srv->mse_main[ch] * MSE_RATIO_OUT_IN)
        && (srv->mse_main[ch] < FILTER_COPY_FAC * srv->mse_adpt[ch]))
{
    for (i_ref = 0; i_ref < srv->ref_num; i_ref++)
    {
        for (i = 0; i<srv->num_main_subband_adf[ch]; i++)
        {
            srv->adf_coef[i_ref][ch][i] = srv->fir_coef[i_ref][ch][i];
        }
    }
    srv->mse_mic_in[ch] = 0.0;
    srv->mse_adpt[ch] = 0.0;
    srv->mse_main[ch] = 0.0;
    srv->err_adf[ch] = srv->err_fir[ch];
}
```

void dios_ssp_aec_firfilter_detect(objFirFilter *srv)

AEC相关： IPNLMS

AN IMPROVED PNLMS ALGORITHM

Jacob Benesty and Steven L. Gay

Bell Laboratories, Lucent Technologies
700 Mountain Avenue
Murray Hill, NJ 07974, USA
E-mail: {jbenesty, slg}@bell-labs.com

```c
void ipnlms_complex(int ch, objFirFilter *srv, int i_ref)
{
    int ii_spk;
    int m, M = srv->num_main_subband_adf[ch];
    float aec_ns_alpha = 0;
    float myu = srv->weight[ch * 2];
    xcomplex delta, z;
    float Padf = 0.0;
    float kl[NUM_MAX_BAND];
    float ip_alpha = 0.5;
    float x2_kl = 0.0;
    float norm_aec = 0.0;
    for (m = 0; m < M; m++)
    {
        kl[m] = complex_abs2(srv->adf_coef[i_ref][ch][m]);
        Padf += kl[m];
    }

    for (ii_spk = 0; ii_spk < srv->ref_num; ii_spk++)
    {
        x2_kl = 0.0;
        for (m = 0; m < M; m++)
        {
            kl[m] = (1 - ip_alpha) / (2 * M) + (1 + ip_alpha)*kl[m] / (Padf * 2 + 1e-5f);
            x2_kl += complex_abs2(srv->stack_sigIn_adf[ii_spk][ch][m])*kl[m];
        }
        norm_aec += x2_kl;
    }

    aec_ns_alpha = myu / (norm_aec + 0.01f);
    delta = complex_real_complex_mul(aec_ns_alpha, complex_conjg(srv->err_adf[ch]));
    for (m = 0; m < M; m++)
    {
        z = complex_mul(srv->stack_sigIn_adf[i_ref][ch][m], delta);
        z = complex_real_complex_mul(kl[m], z);
        srv->adf_coef[i_ref][ch][m] = complex_add(srv->adf_coef[i_ref][ch][m], z);
    }
}
```

athena-signal： 声源定位

```
float dios_ssp_doa_process_api(void* ptr, float* in, int vad_result, int dt_st)
```

加窗

↓

STFT

↓

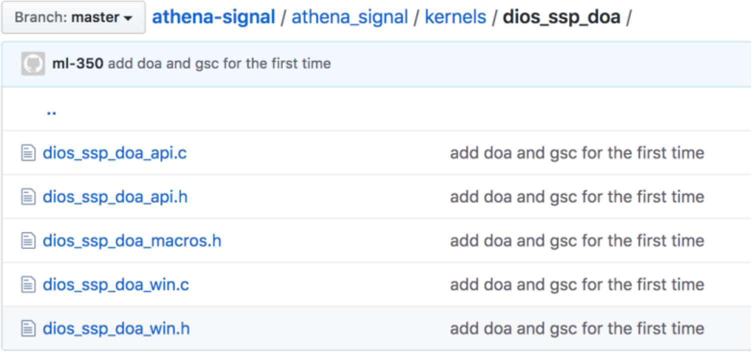
计算相关矩阵：
dios_ssp_doa_cal_rxx(ptr_doa);

↓

计算Capon谱

↓

波达方向搜索

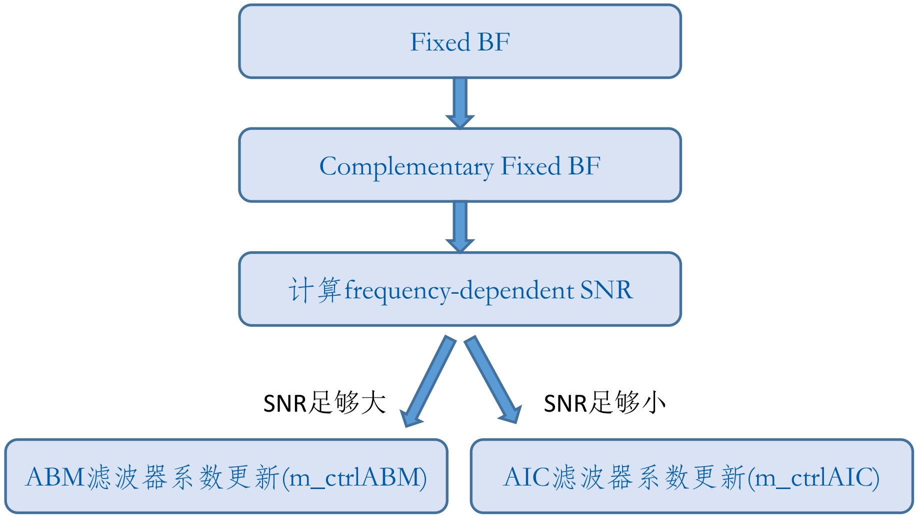


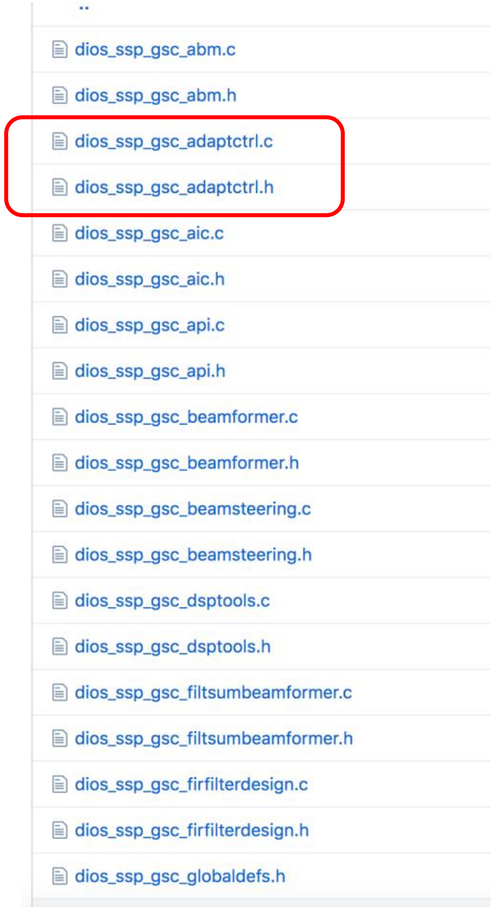
athena-signal： GSC

核心模块:

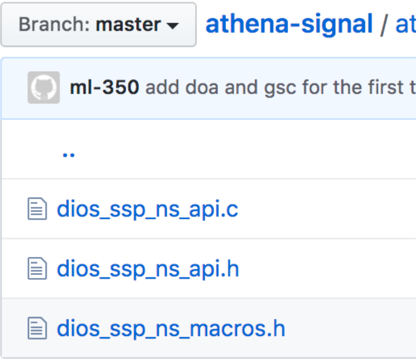
beamsteering – filtsumbeamformer – abm – aic

int dios_ssp_gsc_gscadaptctrl_process(...)





athena-signal ： NS 与 MVDR



MCRA+MMSE

MVDR

解混响： Weighted Prediction Error （WPE）

信号模型（subband）：

$$\boldsymbol{y}_l(t) = \sum_{\tau=0}^{J_l-1} \boldsymbol{H}_l^*(\tau)\boldsymbol{s}_l(t-\tau) + \boldsymbol{v}_l(t),$$

Impulse response shortening result:

$$\boldsymbol{x}_l(t) = \sum_{\tau=0}^{\Delta-1} \boldsymbol{H}_l^*(\tau)\boldsymbol{s}_l(t-\tau) + \text{noise},$$

线性预测和误差:

$$\tilde{\boldsymbol{y}}_l(t) = \sum_{\tau=\Delta}^{\Delta+K_l-1} \boldsymbol{G}_l^*(\tau)\boldsymbol{y}_l(t-\tau)$$

$$\boldsymbol{x}_l(t) = \boldsymbol{y}_l(t) - \tilde{\boldsymbol{y}}_l(t),$$

目标函数的选择:

$$F_{\mathrm{PE}}(\mathcal{G}_l) = \sum_{t\in\mathcal{T}} \left\| \boldsymbol{y}_l(t) - \sum_{\tau=\Delta}^{\Delta+K_l-1} \boldsymbol{G}_l^*(\tau)\boldsymbol{y}_l(t-\tau) \right\|^2,$$

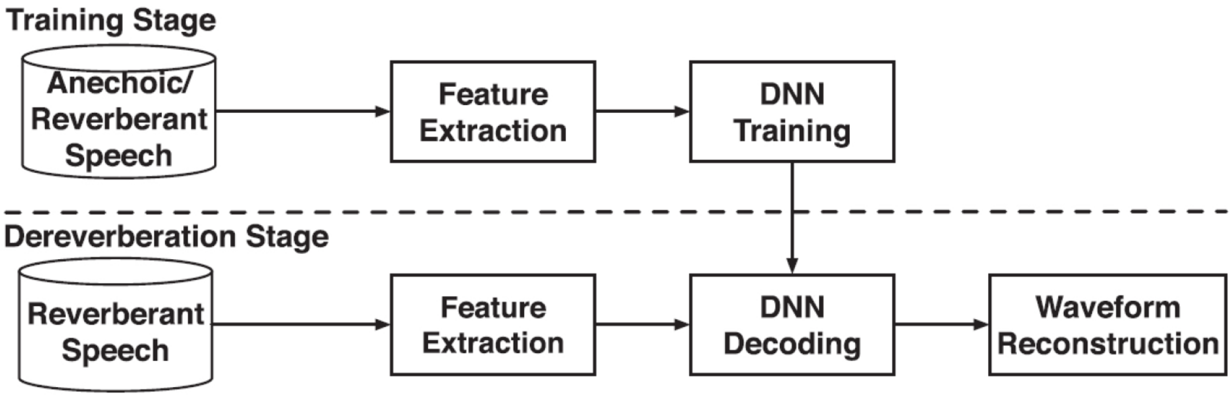
$$F_{\mathrm{WPE}}\left(\mathcal{G}_l^1\right) = \sum_{t\in\mathcal{T}} \frac{\left\| y_l^1(t) - \sum_{\tau=\Delta}^{\Delta+K_l-1} \boldsymbol{g}_l^{1*}(\tau)\boldsymbol{y}_l(t-\tau) \right\|^2}{\lambda_l^1(t)}.$$

“Generalization of Multi-Channel Linear Prediction Methods for Blind MIMO Impulse Response Shortening”, Takuya.

“Linear Prediction-based Dereverberation with Advanced Speech Enhancement and Recognition Technologies for the Reverb Challenge”, NTT.

“Acoustic Modeling for Google Home”, Google.

解混响： mapping-method



"Learning spectral mapping for speech dereverberation", K. Han.

"A reverberation-time-aware approach to speech dereverberation based on deep neural networks", B. Wu

其它一些大家感兴趣的话题

传统信号处理 Vs 深度学习。

信号处理方向，实际工作流程是什么?

麦克风阵列+深度神经网络用于语音识别任务。

两个人自由对话，间距1M，采集设备在一个人胸前佩戴，如何处理?

行业的发展前景和个人规划。

感谢各位聆听!

Thanks for Listening