Speech Information Processing Final Project

Speech Dereverberation

Based on the paper of: "Speech dereverberation based on variance-normalized delayed linear prediction," by Tomohiro Nakatani, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi, and Biing-Hwang Juang, in " IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 7, pp. 1717-1731, Sep. 2010.

Made By: Rohit Kumar

MTech SP(16100)

1 Fundamentals

If we consider one source, than signal at the m-th microphone at time n is :

$$x_{m(}n) = \sum_{m=1}^{M} r_{m(}n)s(n) + e_{m(}n)$$

where s(n) is Clean Speech Signal

$r_{m(}n)$ is Room Impulse response between source and m-th microphone

Now s(k,n) is STFT of clean Speech

where n $\in 1, ..., N$

where k $\in$ $1, ..., K$, then reverberant speech signal at m-th microphone is

$$x_{m(}k, n) = \sum_{l=0}^{L_h-1} h_{m(}k, l)s(k, n-l) + e_{m(}k, n)$$

where the first part:$d_{m(}k, n)$ is the desired speech which is composed of direct speech component and early reverberation part.

and second term is $r_{m(}k, n)$ is the late reverberation which we need to remove

$$x_{m(}k, n) = \sum_{l=0}^{\tau-1} h_{m(}k, l)s(k, n-l) + \sum_{l=\tau}^{L_g-1} h_{m(}k, l)s(k, n-l)$$

Considering the response of all M microphone

$$d_{m(}k, n) = \sum_{i=1}^{M} \sum_{l=0}^{L_g-1} x_{i(}k, n-\tau-l)g_m, i_{(}k, n-l) + x_{m(}k, n)$$

$$D(k) = X(k) - X_{\tau(}k)G(k),$$

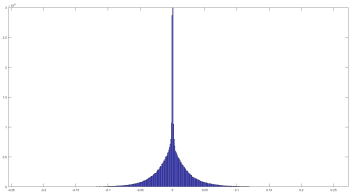

Figure 1: Histogram of reverb speech

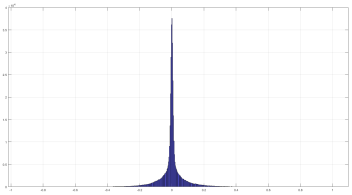

Figure 2: Histogram of original speech

where $D(k) = [d_{1(}k), ..., d_{M(}k)] \in (N * M)$

$d_{m(}k) = [d_{m(}k, 1), ..., d_{m(}k, N)]' \in (N * 1)$

$X(k) = [x_{1(}k), ..., x_{M(}k)] \in (N * M)$

$x_{m(}k) = [x_{m(}k, 1), ..., x_{m(}k, N)]' \in (N * 1)$

$X_{\tau(}k) = [X_{\tau}, 1_{(}k), ..., X_{\tau}, M_{(}k)] \in (N * ML_g)$ convolution matrix

$G(k) = [g_{1(}k), ..., g_{M(}k)] \in (ML_g * M)$

2 Concept of Sparsity

Idea Behind for Using Sparse Concept in Dereverberation

1.Aim:to obtain sparse residual rather than a residual with a min variance.

$$x(n) = \sum_{m=j}^{J} a_j x(n-j) + e(n)$$

2.L1 norm outperform the traditional L2 norm for Linear Prediction Algo.

3.Originally L2 norm $\Rightarrow$ for an excitation signal with iid and Gaussian assumptions.

4.However speech is quasi periodic.

5.Therefor L2 norms suffers from an overemphasis on peaks, which is avoided by using l1 norm which give less emphasis on the outliers of the spiky excitation associated with speech.

2.1 Maths behind the idea

The prediction coefficient matrix G is solved by

$$G = argmin_G ||X - X_|||^q_{pq} + \alpha ||G||^s_{rs}$$

Different type of p,q,r,s will result is different solution.

In my experiment the form of the equation is(when cvx toolbox is used)

$$G = argmin_G ||X - X_|||_1 + \alpha ||G||_1$$

while using the IRLS, value of

$$\alpha = 0$$

$$G = argmin_G ||X - X_|||^2_{22}$$

3 Alogorithm:For dereverbertaion of speech

for each k

input:Reverberant speech $x^m_n{}_k \forall n, m$

set parameters:$\tau, L_g, \epsilon, \delta$

initialize the variances $\lambda_{nk} \Leftarrow |x^1_{nk}|^2$

repeat

$\mathrm{A}_k \Leftarrow \sum_{n=1}^{N} \frac{(x_n - \tau_k)(x_n^H - \tau_k)}{\lambda_{nk}}$

$\mathrm{b}_k \Leftarrow \sum_{n=1}^{N} \frac{(x_n - \tau_k)(x_{nk})^*}{\lambda_{nk}}$

$\mathrm{g}_k \Leftarrow (pinvA)_k b_k$

$\mathrm{d}_{nk} \Leftarrow (g^H)_k x_x - \tau_k \lambda_{nk} \Leftarrow max|d_{nk}|^2, \epsilon$

until g_k converges as$||g^(i+1)_k - g^(i)_k||/||g^(i)_k|| < \delta$,pr a maximum number of iterations

4 Algorithm:Dereverberation using the Sparsity Constraint

for each k input:Reverberant speech $x^m_n{}_k \forall n, m$

set parameters:τ, L_g, ϵ

initialize the desired signal matrix D=x and the correlation structure $\theta = I$

repeat

$\mathrm{w}^(i)_n \Leftarrow ||d^(i-1)_n + \epsilon||_1$

$\mathrm{G}^(i) \Leftarrow (X^H \tau \mathrm{W(i)X} \tau) \mathrm{X}^H \tau \mathrm{W(i)X}$

D$\Leftarrow$X-XτG

until g_k converges or a maximum number of iterations

5 Experimental Setup

1.Dataset: The Dataset has been taken from Reverb Challenge Site, where in the project tackle the problem all type of reverberation that can occur namely very far field ,medium field and near the microphone(considered both male and female.).

2.Sampling Frequency=16Khz

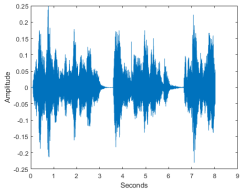

Figure 3: Reverberated Speech Signal

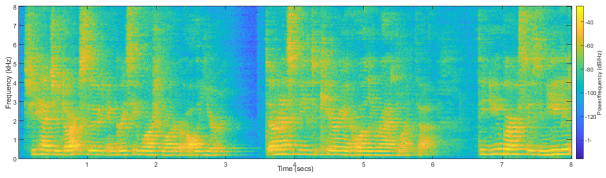

Figure 4: spectrogram of Reverberated sound

3.Window Size=512 sample

4.Window Shift=256

5.Window Used=Hann window

6.τ=2 samples

7.L_g=30

8.ϵ=1e-4

9.max iterations=2

10.CVX toolbox is used for obtaining the L1 minimisation of the prediction coefficient.

11.$\alpha = 0.1$

6 Observation and Results

Figure 3 and 4 of the reverberated speech,figure 5 and 6 of the de revereberated speech, and the figure 7 and 8 is the spectrogram of reverberated speech and histogram of sparse residual de reverb speech

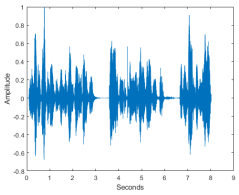

Figure 5: Dereverbrated SPeech Signal

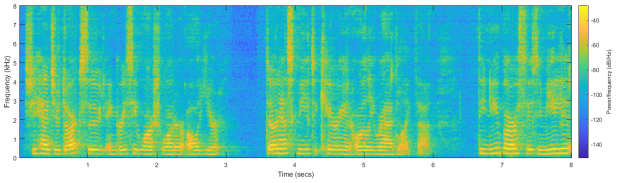

Figure 6: Dereverbrated SPeech spectrogram

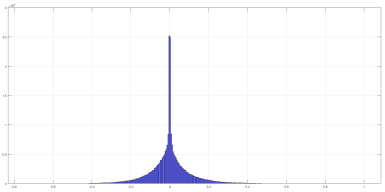

Figure 7: Sparse Signal Histogram

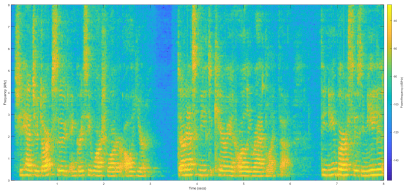

Figure 8: Sparse Signal Spectrogram

7 Performance Analysis

group sparse Linear Prediction by Daniele Giacobello and Tobias Lindstrøm Jensen.

7.1 Performance of Dereverb sound using WPE

Performance Parameters	Reverb Speech	Dereverb Speech
1.Cepstral Distance	2.72	1.23
2.Spectral Distance	1.062	0.96
3.Frequency Weighted SNR	8.29	32.02
4.LPC based Log Likelihood Ratio	0.416	0.0182

7.2 Performance of Dereverb sound using sparsity constraint in wpe

Performance Parameters	Reverb Speech	Dereverb Speech
1.Cepstral Distance	2.72	2.017
2.Spectral Distance	1.062	0.727
3.Frequency Weighted SNR	8.29	13.05
4.LPC based Log Likelihood Ratio	0.416	0.2099

8 References

1.Tomohiro Nakatani, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi, and Biing-Hwang Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 7, pp. 1717-1731, Sep. 2010.

2.P. C. Loizou, Speech Enhancement: Theory and Practice. CRC Press, 2013.

3.P. Naylor and N. D. Gaubitch, Speech Dereverberation. Springer Science Business Media, 2010. 4.Speech Dereverberation Based on Convex Optimisation Algorithm for