ET4147: Signal Processing for Communications

Homework 3

Erik Hagenaars
4272404

Teun de Smalen
4321014

TODO LIST

I. Introduction

This report discusses the work done on the mini-project for the course "IN4182 - Digital Audio and Speech Processing". In the first chapter the overall system is described, followed by the method used for framing.

II. System

A. Signal model

$$Y_t[n] = S_t[n] + N_t[n] \quad \text{(time domain)} \tag{1}$$

$$Y_k[l] = S_k[l] + N_k[l] \quad \text{(frequency domain)} \tag{2}$$

$$R_{S_tN_t}(n,m) = 0 \quad \text{(uncorrelated)} \tag{3}$$

$$R_{Y_tY_t}(n,m) = R_{S_tS_t}(n,m) - R_{N_tN_t}(n,m) \tag{4}$$

$$R_{Y_tY_t}(n,m) = R_{Y_tY_t}(m-n) \quad \text{(wide-sense stationary)} \tag{5}$$

$$P_{YY,k} = \lim_{L\to\infty} \sum_{m=-L/2}^{L/2} R_{Y_tY_t}(m) e^{-j2\pi\frac{km}{K}} \tag{6}$$

$$= P_{SS,k} + P_{NN,k} \tag{7}$$

$$\hat{P}^P_{YY,k}(l) = \frac{1}{L}\left|Y_k(l)\right|^2 \tag{8}$$

$$\hat{P}^B_{YY,k}(l) = \frac{1}{M}\sum_{m=l-M+1}^{l} \hat{P}^P_{YY,k}(m) \tag{9}$$

B. System overview

III. Framing & Overlap Add

The first step, as described in Figure 1, is the framing of the audio file. This is done according to Equation 10. Where l is the frame index (the l-th frame), n is sample number, R is the hoplength. $w[n]$ is the window used to smoothen the signal in such a way that the signal does not become discontinues. For the window multiple options can be used. As in [], the square-root Hann window is used.

Paper omtrent windowing

$$y_l[n] = y[n+lR]w[n], \quad n = 0,\ldots,N-1 \tag{10}$$

The last step of the system is the Overlap Add-block. The windowing is removed after which the samples are added back together to one file.

$$y[n] = \sum_{l=1}^{k} y_l[n]/w[n] \tag{11}$$

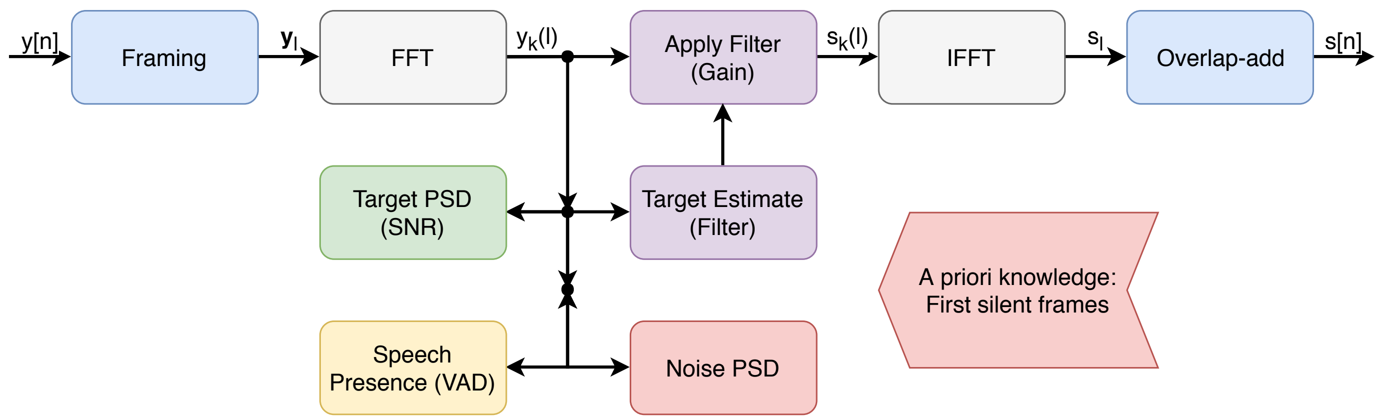

Fig. 1: Overview of the system.

IV. Noise Estimation

$$H_0 : Y_K(l) = N_k(l) \quad \text{(speech absence)} \tag{12}$$
$$H_1 : Y_K(l) = S_k(l) + N_k(l) \quad \text{(speech presence)} \tag{13}$$

$$f(x) = \begin{cases} \alpha\hat{\sigma}^2_{N,k}(l-1) + (1-\alpha)\left|y_k(l)\right|^2 & \text{when } H_0(l) \\ \hat{\sigma}^2_{N,k}(l-1) & \text{when } H_1(l) \end{cases} \tag{14}$$

$$\mathbf{Q} = \{P_{YY,k}(l-M+1) \quad \dots \quad P_{YY,k}(l)\} \tag{15}$$
$$\hat{\sigma}^2_{N,k}(l) = Q_{min} \tag{16}$$

$$\widehat{\sigma^2_N}(l) = \alpha\widehat{\sigma^2_N}(l-1) + (1-\alpha)E\left[\left|N_k(l)\right|^2 \left|y_k(l)\right|\right] \tag{17}$$
$$E\left[\left|N_k(l)\right|^2 \left|y_k(l)\right|\right] = P\left(H_{0,k}(l)|y_k(l)\right) E\left[\left|N_k(l)\right|^2 \left|y_k(l)\right|, H_{0,k}\right] + P\left(H_{1,k}(l)|y_k(l)\right) E\left[\left|N_k(l)\right|^2 \left|y_k(l)\right|, H_{1,k}\right] \tag{18}$$

$$P\left(H_{0,k}(l)|y_k(l)\right) = 1 - P\left(H_{1,k}(l)|y_k(l)\right) \tag{19}$$
$$E\left[\left|N_k(l)\right|^2 \left|y_k(l)\right|, H_{0,k}\right] = \left|y_k(l)^2\right| \tag{20}$$
$$E\left[\left|N_k(l)\right|^2 \left|y_k(l)\right|, H_{1,k}\right] = \widehat{\sigma^2_N}(l-1) \tag{21}$$

$$P\left(H_{1,k}(l)|y_k(l)\right) = \frac{P\left(H_{1,k}(l)\right)p_{Y|H_1}}{P\left(H_{1,k}(l)\right)p_{Y|H_1} + P\left(H_{0,k}(l)\right)p_{Y|H_0}} \tag{22}$$
$$p_{Y|H_0} = \frac{1}{\widehat{\sigma^2_N}\pi}\exp\left(-\frac{\left|y^2\right|}{\widehat{\sigma^2_N}}\right) \tag{23}$$
$$p_{Y|H_0} = \frac{1}{\widehat{\sigma^2_N}(1+\xi_{H_1})\pi}\exp\left(-\frac{\left|y^2\right|}{\widehat{\sigma^2_N}(1+\xi_{H_1})}\right) \tag{24}$$

V. SNR Estimation

$$\xi = \frac{\sigma_{S,k}(l)^2}{\sigma_{S,k}(l)^2} = \frac{P_{SS,k}}{P_{NN,k}} = \frac{E\left\{|S_k(l)|^2\right\}}{E\left\{|N_k(l)|^2\right\}} \tag{25}$$

$$\xi_k(l) = \frac{E\left\{|Y_k(l)|^2\right\}}{E\left\{|N_k(l)|^2\right\}} - 1 \tag{26}$$

$$= \frac{\hat{P}^B_{YY,k}(l)}{\frac{1}{L}E\left\{|N_k(l)|^2\right\}} \tag{27}$$

$$\xi_k(l) = \alpha \frac{E\left\{|S_k(l)|^2\right\}}{E\left\{|N_k(l)|^2\right\}} + (1-\alpha)\left(\frac{E\left\{|Y_k(l)|^2\right\}}{E\left\{|N_k(l)|^2\right\}} - 1\right) \tag{28}$$

$$|S_k(l)|^2 = \left|\hat{S}_k(l-1)\right|^2 \tag{29}$$

$$\frac{E\left\{|Y_k(l)|^2\right\}}{E\left\{|N_k(l)|^2\right\}} - 1 = \max\left[\left(\frac{|Y_k(l)|^2}{E\left\{|N_k(l)|^2\right\}} - 1, 0\right)\right] \tag{30}$$

VI. Voice Activity Detection

$$T(l) = \frac{1}{L}\sum_{k=1}^{k=L} log(\Lambda_k(l)) \gtrless^{H_1}_{H_0} \lambda \tag{31}$$

$$\Lambda_k(l) = P_{YY,k} \tag{32}$$

VII. Target Estimation

$$P_{SS,k}(l) = P_{YY,k}(l) - P_{NN,k}(l) \tag{33}$$

$$\widehat{|S_k(l)|}^2 = |Y_k(l)|^2 - |N_k(l)|^2 \tag{34}$$

$$\widehat{|S_k(l)|}^2 = \left(\max\left\{1 - \frac{E\left[|N_k(l)|^2\right]}{|y_k(l)|^2}, \epsilon\right\}\right)^{\frac{1}{2}} |y_k(l)| \tag{35}$$

$$S_k(l)^2 = \widehat{|S_k(l)|}^2 \dot{e}^{j\measuredangle y_k(l)} \tag{36}$$

$$\hat{S}_k = H_k \dot{Y}_k \tag{37}$$

$$H_k = \frac{P_{SY,k}}{P_{PYY,k}} \tag{38}$$

$$= \frac{SNR_k}{SNR_k + 1} \tag{39}$$

VIII. Conclusion

References

[1] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement:A Survey of the State of the Art*. Morgan & Claypool, 2013. [Online]. Available: https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6813348

APPENDIX A
MATLAB CODE