Variable Speech Distortion Weighted Multichannel Wiener Filter based on Soft Output Voice Activity Detection for Noise Reduction in Hearing Aids

Kim Ngo[1], Ann Spriet[1,2], Marc Moonen[1], Jan Wouters[2] and Søren Holdt Jensen[3]

[1]Katholieke Universiteit Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

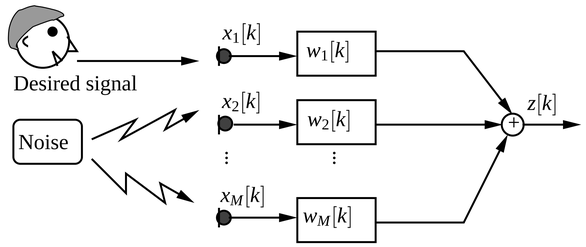
[2]Katholieke Universiteit Leuven, ExpORL, O.& N2, Herestraat 49/721, B-3000 Leuven, Belgium

[3]Aalborg University, Dept. Electronic Systems, Niels Jernes Vej 12, DK-9220 Aalborg, Denmark

kim.ngo@esat.kuleuven.be

1 Introduction

- Background noise (multiple speakers, traffic etc.) is a significant problem for hearing aid users and is especially damaging to speech intelligibility
- Difficulty hearing conversations and understanding speech
- The objective of noise reduction algorithms is to maximally reduce the noise while minimizing speech distortion



- A known multichannel noise reduction technique is the Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF) [1] [2] that allows for a trade-off between noise reduction and speech distortion

2 Multichannel Wiener Filter

- Let $X_i(f), i = 1, ..., M$ denote the frequency-domain microphone signals

$$X_i(f) = X_i^s(f) + X_i^n(f) \tag{1}$$

- f is the frequency domain variable and the superscripts s and n are used to refer to the speech and the noise contribution of a signal, respectively.
- Let $\mathbf{X}(f) \in \mathbb{C}^{M\times 1}$ be defined as the stacked vector

$$\mathbf{X}(f) = [X_1(f)\ X_2(f)\ ...\ X_M(f)]^T \tag{2}$$
$$= \mathbf{X}^s(f) + \mathbf{X}^n(f) \tag{3}$$

- The superscript T denotes the transpose
- Assume that the speech and the noise signals are statistically independent, then the optimal SDW-MWF that provides an estimate of the speech component in the first microphone is given by

$$\mathbf{W}(f) = \left(\mathbf{R}^s(f) + \mu\mathbf{R}^n(f)\right)^{-1}\mathbf{R}^s(f)\mathbf{e}_1 \tag{4}$$

- The output $Z(f)$ of the SDW-MWF can then be written as

$$Z(f) = \mathbf{W}^H(f)\mathbf{X}(f). \tag{5}$$

- Traditionally, the trade-off parameter μ is set to a fixed value and the improvement in noise reduction comes at the cost of a higher speech distortion
- Furthermore, the speech+noise segments and the noise-only segments are weighted equally

3 MWF with Soft Output VAD

- This work presents a variable SDW-MWF based on soft output VAD which allows for a variable trade-off between noise reduction and speech distortion in the SDW-MWF procedure
- Variable SDW-MWF aims at improving noise reduction without increasing speech distortion
- A soft output VAD [3] is used to distinguish between speech and noise and to incorporate a variable trade-off
- In speech dominant segments it is desirable to have less noise reduction to avoid speech distortion
- In noise dominant segments it is desirable to have as much noise reduction as possible

3.1 Variable SDW-MWF

- The variable SDW-MWF is derived from the MSE criterion as (The frequency parameter f is omitted in the sequel for the sake of conciseness)

$$\mathbf{W} = \arg\min_{\mathbf{W}}\ \varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2\} \tag{6}$$
$$\mathbf{W} = \arg\min_{\mathbf{W}}\ \varepsilon\{p\cdot|X_1^s - \mathbf{W}^H\mathbf{X}|^2 + (1-p)\cdot|\mathbf{W}^H\mathbf{X}^n|^2\} \tag{7}$$

- p is the probability that speech is present in a given signal segment
- The solution is then given by

$$\mathbf{W} = (p\cdot\varepsilon\{\mathbf{X}^s\mathbf{X}^{s,H}\} + p\cdot\varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\} + (1-p)\cdot\varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\})^{-1}p\cdot\varepsilon\{\mathbf{X}^sX_1^{s,H}\} \tag{8}$$

$$\mathbf{W} = (p\cdot\varepsilon\{\mathbf{X}^s\mathbf{X}^{s,H}\} + \varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\})^{-1}p\cdot\varepsilon\{\mathbf{X}^sX_1^{s,H}\} \tag{9}$$

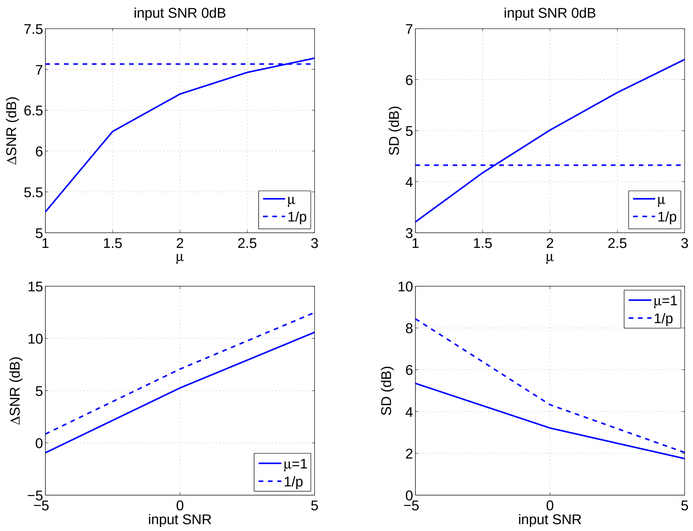
- The variable SDW-MWF can then be written as

$$\mathbf{W} = \left(\mathbf{R}^s + \left(\tfrac{1}{p}\right)\mathbf{R}^n\right)^{-1}\mathbf{R}^s\mathbf{e}_1 \tag{10}$$

3.2 Concept

- Compared to Eq. 4 with the fixed μ the term $\frac{1}{p}$ is now changing based on the soft output VAD
- If $p = 0$, i.e. the probability that speech is presence is zero, the variable SDW-MWF will attenuate the noise by applying $\mathbf{W} \leftarrow 0$
- If $p = 1$ the variable SDW-MWF solution corresponds to the MWF solution
- If $0 < p < 1$ there is a trade-off between noise reduction and speech distortion

4 Experimental Results



5 Conclusion

- Experimental results with a variable SDW-MWF show a SNR improvement with a lower speech distortion compared to a traditional SDW-MWF
- The proposed approach is able to distinguish between speech+noise and noise-only segments
- The concept can be extended to an integrated approach of noise reduction and dynamic range compression [4]

References

[1] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction," *Speech Communication*, vol. 7-8, pp. 636–656, July 2007.

[2] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient based implementation of spatially pre-processed speech distortion weighted multi-channel wiener filtering for noise reduction in hearing aids," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 911–625, Mar. 2005.

[3] S. Gazor and W. Zhang, "A soft voice activity detector based on a laplacian-gaussian model," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 498–505, Sept. 2003.

[4] K. Ngo, S. Doclo, A. Spriet, M. Moonen, J. wouters, and S.H. Jensen, "An integrated approach for noise reduction and dynamic range compression in hearing aids," *accepted for publication in Proc. 16th European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland*, Aug. 2008.