Robust, time-variant design of MVDR Beamformers

Joerg Bitzer[a], K. Uwe Simmer[b] and Markus Kallinger[c]

[a]*Houpert Digital Audio, Anne-Conway-Str. 1, 28359 Bremen; Email: j.bitzer@hda.de*

[b]*ASP Acoustic Signal Processing GmbH, Wachtstr. 17, 28195 Bremen; Email: uwe.simmer@t-online.de*

[c]*Universität Bremen, Dept. of Telecommunication, Otto-Hahn-Allee 1, 28359 Bremen; Email: kallinger@ant.uni-bremen.de*

Abstract

In this contribution, we show a noise estimation procedure in order to design an adaptive Minimum Variance Distortionless Response (MVDR)-Beamformer in the frequency domain. This new algorithm uses a multi-channel minimum statistic algorithm. This leads to a very robust implementation that outperform ideal Voice Activity Detection (VAD) based algorithms in a non-stationary environment.

Introduction

The Problem of speech enhancement in noisy environment is still a research and development problem. One possible solution includes the use of several microphones. It has been shown that the optimal solution consists of a MVDR-Beamfomer which is independent from the desired signal in conjunction with a single channel Wiener-filter[1]

The implementation of the MVDR-beamformer part can be done in many different ways. We will focus on frequency-domain solutions, where the coefficients are estimated directly and plucked into an overlap-add filtering structure [2]. This leads to a so-called adaptive open-loop (AOL) implementation of the MVDR-Beamformer.

Algorithm

MVDR-Beamformer

The coefficients of a time-variant MVDR-Beamformer consisting of N microphones are given by

$$\mathbf{A}(n,l) = \frac{\mathbf{\Phi}^{-1}{}_{xx}(n,l)\mathbf{d}(n,l)}{\mathbf{d}^{H}(n,l)\mathbf{\Phi}^{-1}{}_{xx}(n,l)\mathbf{d}(n,l)} \qquad \text{eq. 1}$$

Where

$$\mathbf{\Phi}_{xx}(n,l) = \begin{bmatrix} \Phi_{X_0X_0}(n,l) & \Phi_{X_0X_1}(n,l) & \cdots & \Phi_{X_0X_{N-1}}(n,l) \\ \Phi_{X_1X_0}(n,l) & \Phi_{X_1X_1}(n,l) & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \Phi_{X_{N-1}X_0}(n,l) & \cdots & \cdots & \Phi_{X_{N-1}X_{N-1}}(n,l) \end{bmatrix}$$

is the power spectral density matrix of the input Signal, estimated by using a recursive Welsh-Periodogram for each element of the form

$$\Phi_{X_iX_j}(n,l) = \alpha\Phi_{X_iX_j}(n,l-1) + (1-\alpha)X_i^{*}(n,l)X_j(n,l) \qquad \text{eq. 2}$$

d represents the propagation vector of the desired source, l denotes the current time, and n the frequency index. In typical speech application scenarios we are using 129 frequency bins with 50% overlap in the overlap-add structure. The recursion constant α is set to 0.9 for a sampling frequency of 8kHz.

This design is close to the theoretically optimal coefficients if -and only if- the steering of the array is perfect and that there is no correlation between the desired signal and any other signals from any other directions. In real-world applications neither requirement is fulfilled. On the one hand, the angle of arrival estimation is not perfect and in most cases strong early reflections of the desired signal occur due to the reverberation effect. Both real-world problems will lead to strong signal cancellation. Therefore most adaptive array algorithms include a voice activity detection (VAD) algorithm to prevent adaptation during speech intervals.

Multi-channel Minimum Statistics

In order to overcome the need for VAD an extension of the minimum statistics (MS) algorithm for noise power spectral estimation[3] can be applied.

The problem of extending the MS-algorithm is that the inherent rules are not valid anymore, since the assumption that noise is always a minimum of the PSD is only valid for Auto-PSDs, but not for Cross-PSDs. Therefore, we suggest to use the channel with the best signal-to-noise ratio (SNR) as a master for all PSD estimators. If a new minimum occurs in the master channel all cross-channel and auto-channel PSDs are updated accordingly, including the correct time-frame, if an older minimum is selected as the valid minimum of the master channel. For speech enhancement typical parameters of the minimum statistics algorithm are a memory of 1-2s divided into 10 blocks.

The problem of underestimation which has been known and is solved for MS [4] is not an issue for this extension, since we are more interested in the spatial information, and the bias is present in all estimated elements and therefore, it is cancelled out in eq. 1.

Simulation Results

In order to demonstrate the performance and robustness of this new algorithm, we generated a time-variant environment. Figure 1 shows the actual setting of the environment, which consists of 5 microphones linearly spaced with a distance of 10 cm. Since we are choosing a fixed broadside configuration, the steering vector reduces to **d = 1**. Furthermore, two noise sources are present, which changes their behaviour over time. The first noise source reduces its power by 12dB after 6s and the second noise source starts after 2s. These dynamic changes happen during a phase, where a VAD detection algorithm will detect speech, if it is optimised to reject

speech under all circumstances. The VAD detection boundaries (set manually) are given in the topmost graph in Figure 2. In order to simulate a more realistic environment all signals were convolved with room impulse responses, generated by the image method[5]. The reverberation constant was set to τ_{60}= 300ms, which is a good approximation for typical office environments. Figure 3 shows the SNR-Enhancement (SNRE) for a varying input-SNR. We tested different algorithms, all based on the AOL-structure:

- **No NR**: no noise reduction applied
- **AOL-MVDR**: Direct implementation eq. 1 without any modification
- **AOL-MVDR VAD**: A VAD-based implementation of the algorithm
- **AOL-MVDR MS**: the new MS based algorithm.

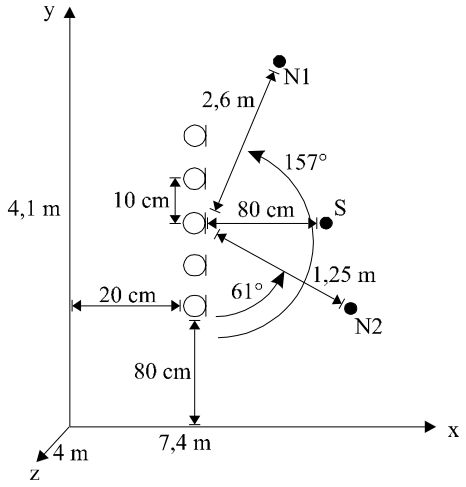

Figure 1: Configuration

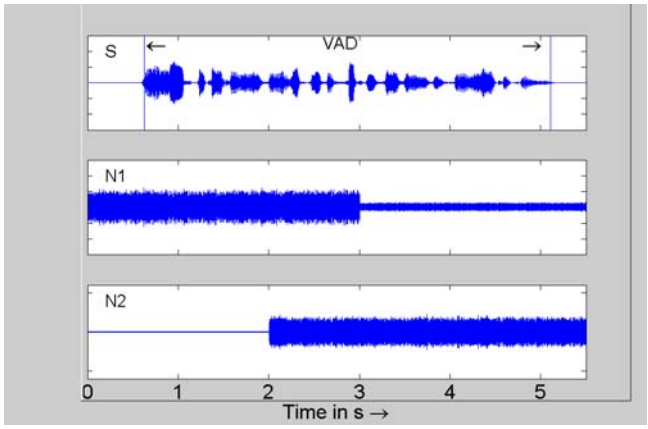

Figure 2: Signal description

The results shown in Figure 3 clearly indicate that an algorithm without any protection from signal cancellation (AOL-MVDR) will reduce the SNR, especially if the desired signal is dominant (Input SNR>10dB). Due to the non-stationary environment the VAD algorithm performs fair, but cannot follow the changes and is therefore restricted in its performance. In comparison, the MS algorithm is able to follow the changes and outperforms the other algorithms in terms of SNR-Enhancement.

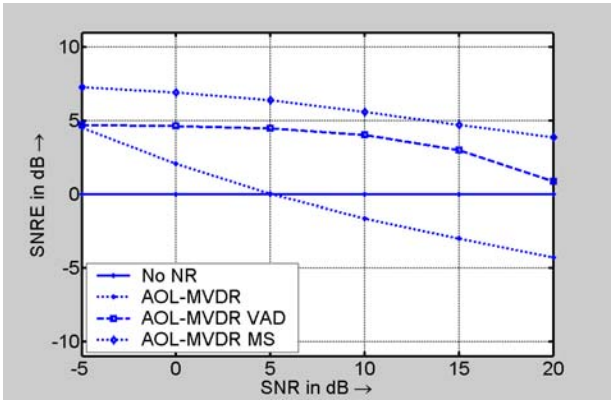

Figure 3: Simulation results SNRE

In order to analyse the signal cancellation problem in more detail, we computed a signal degradation (SD) measure shown in Figure 4. SD is a Log-Area-Ratio[6] (LAR) coefficient based comparison of the clean desired signal without reverberation to the enhanced speech signal after processing. If **No NR** is applied the signal is distorted by the reverberation. By using the standard algorithm the signal distortion is high and noticeable in the output signal. For VAD and MS the additional degradation is quite low and not noticeable.

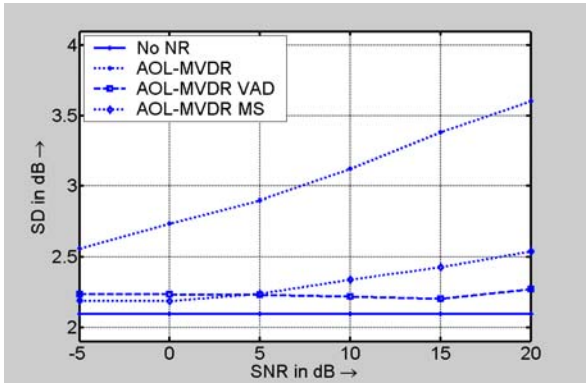

Figure 4: Simulation results for signal distortion (SD)

Conclusion

In this contribution we proposed a new multi-channel noise estimation algorithm based on minimum statistics. Due to the fact that it excludes all speech components and estimates noise only, a time-variant robust MVDR-Beamformer can be designed. Simulation results show that the algorithm is suitable for tracking time-variant noisy environments without any VAD-algorithm. Informal listening tests indicate that no artefacts are introduced in comparison to a VAD solution.

---

[1] Simmer, Bitzer, Marro, „Post-Filtering“ in „Microphone Arrays“ edited by Ward and Brandstein, Chapter 3, p. 39- 54

[2] Rabiner, Schafer, “Digital Processing of Speech Signals”, Englewood Cliffs , Prentice Hall, 1978

[3] Martin, “Spectral Subtraction based on Minimum Statistics”, EUSIPCO 94, Edinburgh, UK, p. 1182-1185

[4] Martin, „Noise Power Spectral Density Estimation based on optimal Smoothing and Minimum Statistics“, IEEE Trans. On Speech and Audio Signal Processing, Vol. 9, Juli 2001 p. 504-512

[5] Allen, Berkley, „Image Method for efficiently simulating small-rooms acoustic“, Journal of acoustical society of America JASA, 65,1979, p. 943-950

[6] Quackenbush, Barnwell, Clemens, “Objective Measures of Speech Quality”, Prentice-Hall, Englewood Cliffs, 1988