Background Noise Simulation in Cars based on Multiple Input – Multiple Output Equalization

Simon Grimm, Jürgen Freudenberger
Institute for System Dynamics, HTWG Konstanz, Germany, Email: {sgrimm, jfreuden}@htwg-konstanz.de

Abstract

For the evaluation of speech enhancement algorithms in conditions which contain background noise, often simulation environments are used that aim to accurately reproduce the proper noise signal conditions at the microphones. Besides the power spectral density (PSD), the spatial properties of the noise field should be reproduced correctly to verify algorithms that use more than one microphone. We consider a Multiple-Input Multiple-Output (MIMO) arrangement using more loudspeakers than microphones as a simulation environment. Using more loudspeakers than microphones decreased the probability of ill-conditioned acoustic transfer function matrices. Since the acoustic transfer function matrix is not square, the Moore-Penrose pseudo-inverse is used to equalize the MIMO system. In measurements taken in a car environment, the reproduced signals show an equal PSD as well as a similar magnitude squared coherence (MSC) compared with the input signals.

Introduction

Hands-free communication systems continue to be popular in a growing field of applications. However, they are often operated in noisy environments, e.g. in car communications. To achieve an augmentation of the desired signals, signal processing algorithms are used that aim to improve the signal-to-noise ratio (SNR) to ensure a better communication. Speech enhancement algorithms where developed, that take the spatial information of the sound field into account by using more than one microphone signal [1, 2, 3].

To evaluate these algorithms in noisy environments, background noise simulations are often used. In order to verify these algorithms in a real acoustic environment, the proper signal conditions between the recorded noise signals must be preserved. In the ETSI EG 202 396-1 standard [4], the background noise conditions are reproduced by several loudspeakers and one or two microphones. The acoustic transfer functions from the loudspeakers to the microphones are equalized in third octave bands and the time-difference-of-arrivals are compensated. However, for enhancement algorithms using multiple microphones this is not sufficient, since they take the spatial information of the noise field into account, which is not considered in this equalization approach.

In order to achieve a more accurate reproduction of the noise signals, the equalization must be capable to decorrelate the acoustic propagation paths of the simulation environment. This is considered in the ETSI TS 103 224 standard [5], where eight loudspeakers and microphones are used to recreate a sound field around a dummy head or a hand-held communication device. In the standard, the inverse of the acoustic transfer function (ATF) matrix is used for the equalization of the acoustic MIMO system. While this approach shows good results compared with ETSI EG 202 396-1 and other approaches [6], there are several problems if the inverse acoustic matrix is ill-conditioned. Therefore a regularization factor is introduced, which can be optimized in different manners [7, 8].

We propose an approach using more loudspeakers than microphones compared with [5]. Since we have access to the signals of the hands-free communication system, we aim to equalize the MIMO system at the actual microphones compared with the ETSI TS 103 224 standard, where measurement microphones around the actual hand-held device or dummy head are used for the equalization. By using a greater number of loudspeakers than microphones, no unique solution for the inverse equalization matrix exists. One solution, the Moore-Penrose pseudo-inverse of the ATF matrix, can be calculated and be used as a mixing matrix for the noise input signals. The measurements with this MIMO arrangement in a car environment show good results for the accurate reproduction of the PSDs of the input signals as well as the MSC. The frequency dependent condition number shows that the equalization matrix is better conditioned if more loudspeakers than microphones are used. Therefore less or even none regularization is required for the equalization of the MIMO system compared with [5].

In the following sections, the signal model and the notation of the MIMO acoustic environment is described. Further, the equalization approach that uses the pseudo-inverse of the acoustic transfer function matrix is derived. The results of the simulation as well as the measurements are presented afterwards.

The Signal Model

In this section, the acoustic MIMO model as well as the corresponding notation is presented. We consider the MIMO system, which consists of M loudspeakers and N microphones ($M > N$) as time-invariant and linear. Also it is assumed, that the signal-to-noise ratio is sufficiently high, so noise influences can be neglected in the following. As a result, the signal $y_n(k)$ at the n-th microphone can be written as

$$y_n(k) = \sum_{m=1}^{M} h_{m,n}(k) * x_m(k) \tag{1}$$

where k is the discrete time index, $x_m(k)$ refers to the discrete input signal that is fed to the m-th loudspeaker and $h_{m,n}$ describes the acoustic impulse response from the m-th loudspeaker to the n-th microphone. $*$ denotes the convolution.

The acoustic transfer functions derived from the corresponding impulse responses as well as the MIMO input and output signals can be written in the frequency domain as the acoustic transfer function matrix

$$\mathbf{H}(\nu) = \begin{bmatrix} H_{11}(\nu) & H_{12}(\nu) & \cdots & H_{1N}(\nu) \\ H_{21}(\nu) & H_{22}(\nu) & \cdots & H_{2N}(\nu) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M1}(\nu) & H_{M2}(\nu) & \cdots & H_{MN}(\nu) \end{bmatrix} \tag{2}$$

ν denotes the frequency bin index. Furthermore, we have

$$\mathbf{X}(\nu) = [X_1(\nu), X_2(\nu), \cdots, X_M(\nu)] \tag{3}$$
$$\mathbf{Y}(\nu) = [Y_1(\nu), Y_2(\nu), \cdots, Y_N(\nu)] \tag{4}$$

where $\mathbf{X}(\nu)$ is the loudspeaker input signal vector and $\mathbf{Y}(\nu)$ is the microphone signal vector. In the following the index ν is omitted when possible.

As a result, the MIMO system equation as given in Eq.(1) can be written in a compact form in the frequency domain

$$\mathbf{Y} = \mathbf{XH} \tag{5}$$

Equalization of the MIMO system

In the following, the equalization of the MIMO acoustic environment is derived. Therefore, a new signal vector $\mathbf{S}(\nu)$ is introduced, which contains the input signals that need to be accurately reproduced at the microphones.

$$\mathbf{S}(\nu) = [S_1(\nu), S_2(\nu), \cdots, S_N(\nu)] \tag{6}$$

Furthermore an equalization matrix $\mathbf{W}(\nu)$ is introduced, that acts as a pre-equalization filter for the input signals before they are sent to the loudspeakers.

$$\mathbf{W}(\nu) = \begin{bmatrix} W_{11}(\nu) & W_{12}(\nu) & \cdots & W_{1M}(\nu) \\ W_{21}(\nu) & W_{22}(\nu) & \cdots & W_{2M}(\nu) \\ \vdots & \vdots & \ddots & \vdots \\ W_{N1}(\nu) & W_{N2}(\nu) & \cdots & W_{NM}(\nu) \end{bmatrix} \tag{7}$$

The whole system, from the input signals in vector $\mathbf{S}(\nu)$ to the microphone signals in vector $\mathbf{Y}(\nu)$, can be represented as a block diagram as shown in Figure 1.

The relation between $\mathbf{S}(\nu)$ and $\mathbf{Y}(\nu)$ can be written as

$$\mathbf{Y} = \mathbf{SWH} \tag{8}$$

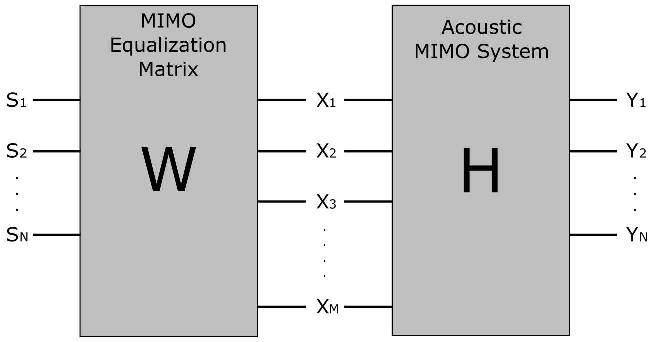

Figure 1: MIMO System - Block Diagramm

In order to equalize the acoustic system it must be assured that the following statement holds

$$\mathbf{Y} \overset{!}{=} \mathbf{S} \tag{9}$$

which can be achieved by choosing $\mathbf{W}$ to fulfil

$$\mathbf{WH} = \mathbf{I} \tag{10}$$

where $\mathbf{I}$ denotes the unity matrix. In case the number of loudspeakers is equal to the number of microphones ($M = N$), this obviously results in

$$\mathbf{W} = \mathbf{H}^{-1} \tag{11}$$

if $\mathbf{H}$ has full rank and is well conditioned. In the ETSI TS 103 224 standard this is implemented as

$$\mathbf{W} = (\mathbf{H}^\dagger\mathbf{H} + \beta\mathbf{I})^{-1}\mathbf{H}^\dagger \tag{12}$$

where † is the conjugate transpose and β is a regularization factor to avoid problems due to an ill-conditioned matrix $\mathbf{H}$.

A matrix is ill-conditioned if it has a very high condition number κ, which is defined as the ratio between the maximal and minimal singular values σ_{max} and σ_{min}, respectively

$$\kappa(\nu) = \frac{\sigma_{max}(\nu)}{\sigma_{min}(\nu)} \tag{13}$$

The condition number also relates the matrix norms of the matrices $\mathbf{H}$ and $\mathbf{W}$, i.e.

$$\kappa(\nu) = ||\mathbf{H}|| \cdot ||\mathbf{W}||. \tag{14}$$

From which we observe that an ill-conditioned matrix $\mathbf{H}$ results in a matrix $\mathbf{W}$ with large matrix norm. For large condition numbers, the regularization factor β in Eq.(12) limits the norm of $\mathbf{W}$. However, the regularization prevents perfect equalization of the MIMO system.

We propose to use more loudspeakers than microphones ($M > N$). This reduces the probability of ill-conditioned matrices, as will be shown in the following section. Hence, a smaller regularization factor β can be chosen or the regularization is completely avoided.

With $M > N$ the matrices are not square and no unique inverse exists. By using the Moore-Penrose pseudo-inverse, an inverse of the acoustic system MIMO matrix $\mathbf{H}$ can be calculated, analogous to Eq.(12). This pseudo-inverse minimizes the matrix norm of the equalization matrix $\mathbf{W}$. The minimized norm also results in reduced power of the signals that are played back by the loudspeakers, which follows from the submultiplicative property

$$||\mathbf{X}|| \leq ||\mathbf{S}|| \cdot ||\mathbf{W}||. \qquad (15)$$

Simulation and Measurement Results

In order to verify the proposed equalization approach for the proper reproduction of noise signals at the microphones, measurements in a car environment where taken. The simulation scenario consisting of four loudspeakers and two microphones is shown in Figure 2.

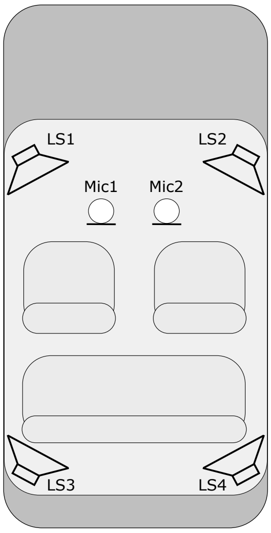

Figure 2: Loudspeaker-Microphone Arrangement in the Car Environment

The input signals are two noise signals that are recorded in a car at a driving speed of 100 km/h, which should be reproduced at the microphones. For the equalization approach, the acoustic transfer functions in the simulation environment from each loudspeaker to each microphone where measured using a logarithmic sine sweep. Based on that measurement, the mixing matrix $\mathbf{W}$ was calculated. Since the elements in $\mathbf{W}$ contain non causal filters, $\mathbf{W}$ is transformed to the time domain and delayed to obtain a causal filter. Then the input signals are pre-equalized with the corresponding filters of the mixing matrix in the time domain. The filter length for the MIMO equalization was chosen to 8192 samples and a sampling frequency of fs = 16 kHz was used. The pre-equalized signals are sent to the loudspeakers and the results are recorded with the two microphones. For comparison, further measurements where also taken with the signals pre-equalized as suggested in the ETSI EG 202 396-1 standard.

To verify, that the correct power spectral densities of the input signals are reproduced at the microphones, the PSDs where analyzed in third octave bands for our equalization approach as well as for the ETSI EG 202 396-1 standard and compared with the PSDs of the original input noise signals. The results are shown in Figure 3. The reference (A) shows the PSD for the input signal, which is aimed to be reproduced at microphone 1. The PSDs for the equalization after ETSI EG 202 396-1 (B) as well as for our proposed MIMO equalization approach (C) are shown. As can be seen, both equalization approaches match quite well in terms of the PSD compared with the reference signal.

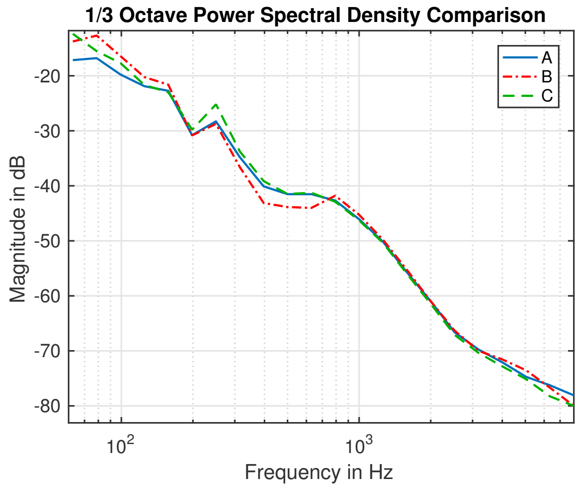

Figure 3: Power Spectral Density Comparison - A: Input Signal (Reference); B: ETSI EG 202 396-1 Equalization; C: proposed MIMO Equalization

In Figure 4, a comparison for the correct reproduction of the magnitude squared coherence between the two input signals is depicted. The MSC for the input signals is denoted as the reference (D). In comparison, the MSC for the MIMO equalization approach is shown (F). As can be observed, the MSC is reproduced quite well over the whole spectrum. In low frequencies, the performance decreases due to the not sufficient filter length for the equalization filters, but the overall performance can be considered good. The ETSI EG 202 396-1 standard (E) does not consider spatial reproduction at all, so the MSC is close to one for the observed frequencies.

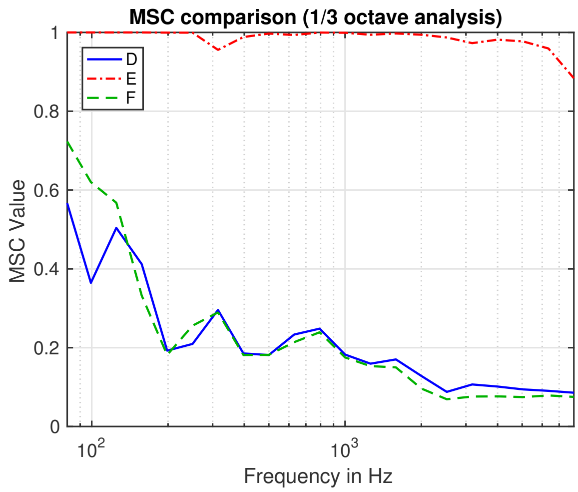

Figure 4: Magnitude Squared Coherence Comparison - D: Input Signals (Reference); E: ETSI EG 202 396-1 Equalization; F: proposed MIMO Equalization

To verify if the estimated acoustic transfer function matrices are well-conditioned, the condition number over all frequencies for the matrices are plotted in Figure 5. The acoustic transfer function matrices are derived using two microphones and a varying number of loudspeakers ($N = 2$ and $M \in \{1, 2, 3\}$). The frequency dependent condition numbers for two (2×2), three (3×2) and all four loudspeakers (4×2) are shown. As can be observed, the frequency dependent condition number values are the highest for the two loudspeaker case and decrease as more loudspeakers are used. Hence, the measurements show that the inverse / pseudo-inverse pre-equalization matrix is better conditioned using more loudspeakers than microphones.

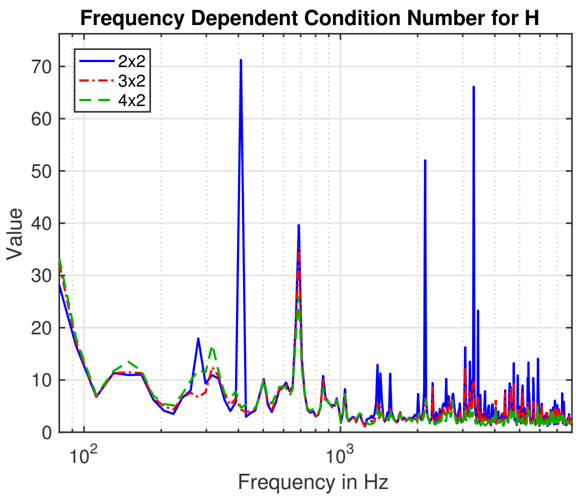

Figure 5: Condition Number Comparison - 2×2: two loudspeakers; 3×2: three loudspeakers, 4×2: four loudspeakers

Conclusion

In this paper, we proposed an approach for the reproduction of noise signals in a car environment. In contrast to the ETSI TS 103 224 standard, we proposes a setup with more loudspeakers than microphones. The MIMO system is equalized using the Moore-Penrose pseudo-inverse of the acoustic transfer function matrix. Using more loudspeakers than microphones decreased the probability of ill-conditioned acoustic transfer function matrices. The measurements in a car environment show that the noise PSD is reproduced well at the microphones, where an acoustic MIMO system consisting of four loudspeakers and two microphones was considered. Similarly, the MSC between the input signals is reproduced with high precision. For a system with a small number of target microphones, the proposed approach simplifies the measurement and reproduction setup for the noise simulation compared to the ETSI TS 103 224 standard.

References

[1] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, Wiley & Sons, 2006.

[2] Jacob Benesty, Jingdong Chen, and Yiteng Huang, Eds., *Microphone Array Signal Processing*, Springer Berlin Heidelberg, 2008.

[3] Yiteng Huang, Jacob Benesty, and Jingdong Chen, Eds., *Acoustic MIMO Signal Processing*, Springer Berlin Heidelberg, 2006.

[4] European Telecommunication Standards Institute, "Speech processing, transmission and quality aspects - part 1: Background noise simulation technique and background noise database, vol. 4," *ETSI EG 202 396-1*, pp. 1–58, 2011.

[5] European Telecommunication Standards Institute, "A sound field reproduction method for terminal testing including background database, vol. 1," *ETSI TS 103 224*, pp. 1–36, 2014.

[6] Juan David Gil Corrales, Marton Marschall, Torsten Dau, Wookeun Song, Claus Blaabjerg, Michael Hoby Andersen, and Soren W. Christensen, "Simulation of realistic background noise using multiple loudspeakers," in *Danish Sound Innovation Network*, 2014.

[7] Scott G. Norcross and Martin Bouchard, "Multichannel inverse filtering with minimal-phase regularization," in *Audio Engineering Society Convention 123, 1-8*, 2007.

[8] Juan David Gil Corrales, Wookeun Song, and Ewen Macdonald, "Reproduction of realistic background noise for testing telecommunications devices," in *Audio Engineering Society Convention 182, 1-10*, 2015.