THEORETICAL NOISE REDUCTION LIMITS OF THE GENERALIZED SIDELOBE CANCELLER (GSC) FOR SPEECH ENHANCEMENT

Joerg Bitzer†, Klaus Uwe Simmer‡ and Karl-Dirk Kammeyer†

†University of Bremen, FB–1, Dept. of Telecommunications
P.O. Box 330 440, D–28334 Bremen, Germany, email: bitzer@comm.uni-bremen.de
‡ Aureca GmbH,
Mozartstrasse 26, D–28203 Bremen, Germany, e-mail: uwe.simmer@aureca.com

ABSTRACT

In this paper we present an analysis of the generalized sidelobe canceller (GSC). It can be shown that the theoretical limits of the noise reduction performance depend only on the auto- and cross-spectral densities of the input signals. Furthermore, we compute the limits of the noise reduction performance for the theoretically determined diffuse noise field, which is an approximation for reverberant rooms. Our results will show that the GSC cannot reduce noise further than 1dB. These results were verified by simulation of reverberant environments. Only in sound-proof rooms with a reverberation time less than 100ms the GSC performs well.

1. INTRODUCTION

Noise reduction in speech communication is still an unsolved problem in the signal processing society. The problem is that the desired speech signal and the environmental noise overlap in the time- and in the frequency domain. One way of separating the signals is to use different spatial characteristics, but to get spatial information we need multi microphone systems.

In the last decades different approaches to use spatial information for noise reduction were introduced. Many of them are based on the noise cancelling system introduced by Widrow [1]. For this scheme and some closely related systems theoretical studies for the noise reduction performance are given in [2, 3, 4]. Another closely related scheme is the generalized sidelobe canceller (GSC) introduced by Griffith and Jim [5]. Many modern concepts are derivated from this scheme [6, 7, 8]. Many authors have evaluated more or less good results of the performance under different conditions. In the single-source coherent noise field the adaptive beamformer performs very well, but the performance degrades in rooms, due to reverberation [9]. In cars as a possible application for hands–free devices the results differ for the reported noise reduction between four and twelve dB, as different cars were examined and different numbers of microphones were used [10, 11, 12, 13, 14]. Affes and Grenier [14] reported that adaptive beamformers are not the right choice for noise reduction in cars, as the GSC causes strong distortions to the desired speech, due to signal cancellation effects.

In this paper we derive new formulas to predict the optimal noise reduction performance in a given noise field. We will show that the auto- and cross-power spectral densities between the microphones are the only information we need. For theoretical studies we will show that only the complex coherence functions between the sensors determine the noise reduction performance. Secondly, we examine the noise reduction performance in three theoretically well defined noise fields (single-source coherent, completely incoherent, and diffuse noise field, as a good approximation of a reverberant environment [2]). Finally, results for simulated reverberant rooms are given.

2. THEORETICAL STUDIES

2.1. Structure

The generalized sidelobe canceller (GSC) consists of two parts, a fixed delay and sum beamformer and a sidelobe cancelling path (see figure 1). For our studies we are using an open-loop frequency domain implementation of the original Griffith et al. beamformer. The $(N-1\times N)$ blocking matrix B is set to

$$B = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \tag{1}$$

where N is the number of microphones.

2.2. Performance Analysis

The noise reduction performance (NR) for the fixed beamformer is given in [2, 15]. The NR of the adaptive sidelobe path can be described as the ratio of the power spectral density (PSD) of the fixed beamformer output $P_{Y_bY_b}$ and the output of the complete system P_{ZZ}.

$$\begin{aligned} NR(\omega) &= \frac{P_{Y_bY_b}(\omega)}{P_{ZZ}(\omega)} \\ &= \frac{P_{Y_bY_b}(\omega)}{P_{Y_bY_b}(\omega) - \dfrac{1}{N-1}\displaystyle\sum_{i=0}^{N-2} |H_i(\omega)|^2 \cdot P_{Y_iY_i}} \end{aligned} \tag{2}$$

where $P_{Y_iY_i}$ is the PSD of one output path of the blocking matrix and $|H_i(\omega)|^2$ is the transfer function of the adaptive filter. According to Wiener theory the optimal filter coefficients H_{opt} are given by the cross-PSD of the beamformer output $P_{Y_bY_i}$ and the single blocking output $P_{Y_iY_i}$

$$H_{opt} = \frac{P_{Y_iY_b}(\omega)}{P_{Y_iY_i}(\omega)} \tag{3}$$

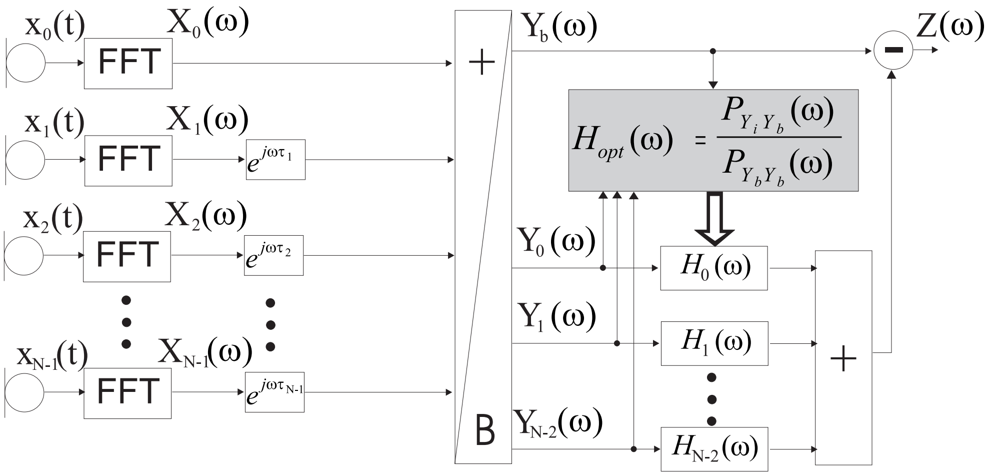

Figure 1: Block diagram of a GSC in an open-loop frequency domain implementation

Equation 2 for the optimal solution results in

$$NR(\omega) = \frac{1}{1 - \frac{1}{(N-1)P_{Y_bY_b}(\omega)} \sum_{i=0}^{N-2} \left| \frac{P_{Y_iY_b}(\omega)}{P_{Y_iY_i}(\omega)} \right|^2 \cdot P_{Y_iY_i}(\omega)} \tag{4}$$

As a next step we describe the dependence of equation 4 in terms of the input signals $X_i(\omega)$. The beamformer output is given by

$$Y_b(\omega) = \frac{1}{N} \sum_{j=0}^{N-1} X_j(\omega) \tag{5}$$

and the output of one sidelobe path is $Y_i(\omega) = X_i(\omega) - X_{i+1}(\omega)$. The cross-PSD $P_{Y_iY_b}$ then is

$$\begin{aligned} P_{Y_iY_b}(\omega) &= (X_i(\omega) - X_{i+1}(\omega)) \frac{1}{N} \sum_{j=0}^{N-1} X_j^*(\omega) \qquad (6) \\ &= \frac{1}{N}\Big(X_i(\omega) \sum_{j=0}^{N-1} X_j^*(\omega) - X_{i+1}(\omega) \sum_{j=0}^{N-1} X_j^*(\omega) \Big) \\ &= \frac{1}{N}\Big(\sum_{j=0}^{j<i} (X_j(\omega)X_i^*(\omega))^* + |X_i(\omega)|^2 \\ &\quad + \sum_{j=i+1}^{N-1} X_i(\omega)X_j^*(\omega) \\ &\quad - \sum_{j=0}^{(j<=i)} (X_j(\omega)X_{i+1}^*(\omega))^* \\ &\quad - |X_{i+1}(\omega)|^2 - \sum_{j=i+2}^{N-1} X_{i+1}(\omega)X_j^*(\omega) \Big) \end{aligned}$$

and the PSD $P_{Y_iY_i}$ is given by [16]

$$P_{Y_iY_i}(\omega) = |X_i(\omega)|^2 + |X_{i+1}(\omega)|^2 - 2\Re\left\{X_i(\omega)X_{i+1}^*(\omega)\right\} \tag{7}$$

where $\Re\{\cdot\}$ denotes that only the real part of the complex cross-PSD is taken into account. Equations 6 and 7 show that only the auto- and the cross-PSD of the inputs are necessary to predict the optimal performance.

If we assume that the PSD of the noise P_{NN} is the same at each sensor, we can rewrite equation 6 in terms of the complex coherence

$$\Gamma_{X_1X_2}(\omega) = \frac{P_{X_1X_2}(\omega)}{\sqrt{P_{X_1X_1}(\omega)P_{X_2X_2}(\omega)}} \tag{8}$$

The result is:

$$\begin{aligned} P_{Y_iY_b}(\omega) &= \frac{P_{NN}(\omega)}{N}\Big(\sum_{j=0}^{j<i} \left(\Gamma_{X_jX_i}(\omega)\right)^* \qquad (9) \\ &\quad + \sum_{j=i+1}^{N-1} \Gamma_{X_iX_j}(\omega) \\ &\quad - \sum_{j=0}^{(j<=i)} \left(\Gamma_{X_jX_{i+1}}(\omega)\right)^* \\ &\quad - \sum_{j=i+2}^{N-1} \Gamma_{X_{i+1}X_j}(\omega) \Big) \end{aligned}$$

and the PSDs $P_{Y_iY_i}$ and $P_{Y_bY_b}$ are given by [16]

$$P_{Y_iY_i}(\omega) = 2P_{NN}(\omega)\left(1 - \Re\left\{\Gamma_{X_iX_{i+1}}(\omega)\right\}\right) \tag{10}$$

and

$$P_{Y_bY_b}(\omega) = P_{NN}(\omega)\left(\frac{1}{N} + \left(1 - \frac{1}{N}\right)\bar{\Gamma}(\omega)\right) \tag{11}$$

where

$$\bar{\Gamma}(\omega) = \frac{2}{N^2 - N} \sum_{i=0}^{N-2} \sum_{j=i+1}^{N-1} Re\left\{\Gamma_{ij}(\omega)\right\} \tag{12}$$

denotes the average complex coherence function of the noise field for all sensor pairs $i \neq j$.

Putting eqs. 9-11 into eq. 4 shows that the noise reduction performance depends only on the spatial coherence of the noise field of all sensor pairs.

$$NR(\omega) = \frac{1}{1 - \frac{1}{(N-1)} \sum_{i=0}^{N-2} \frac{\left| \sum_{j=0}^{j<i} \left(\Gamma_{X_j X_i}(\omega)\right)^* + \sum_{j=i+1}^{N-1} \Gamma_{X_i X_j}(\omega) - \sum_{j=0}^{(j<=i)} \left(\Gamma_{X_j X_{i+1}}(\omega)\right)^* - \sum_{j=i+2}^{N-1} \Gamma_{X_{i+1} X_j}(\omega) \right|^2}{2N^2 \left(1 - \Re\left\{\Gamma_{X_i X_{i+1}}(\omega)\right\}\right) \left(\frac{1}{N} + \left(1 - \frac{1}{N}\right) \bar{\Gamma}(\omega)\right)}} \quad (13)$$

2.3. Theoretical Limits

In this section we will examine different theoretically well defined noise fields to compute the limit of the noise reduction performance of the GSC-structure. Only the sidelobe path is taken into account. The noise reduction of the complete system is always the result of the addition of the GSC part and the noise reduction of the fixed beamformer.

2.3.1. Coherent Noise Field

Assuming a single noise source in the far field of the sensor array the complex coherence function is given by

$$Re\{\Gamma_{X_1 X_2}(\omega)\} = \cos\left(\frac{\omega \cos(\theta) d}{c}\right) \quad (14)$$

$$Im\{\Gamma_{X_1 X_2}(\omega)\} = -\sin\left(\frac{\omega \cos(\theta) d}{c}\right) \quad (15)$$

where d denotes the microphone distance, θ the angle of arrival, and c the speed of sound. The noise reduction performance of the sidelobe path reaches infinity at all frequencies.

2.3.2. Incoherent Noise Field

The coherence in the incoherent noise field is zero at all frequencies. This may be caused by sensor noise, for example. In this case the noise reduction performance is zero dB at all frequencies.

2.3.3. Diffuse Noise Field

In a diffuse noise field the coherence is real valued only and given by

$$Re\{\Gamma_{X_1 X_2}(\omega)\} = \frac{\sin\left(\frac{\omega d}{c}\right)}{\frac{\omega d}{c}} \quad Im\{\Gamma_{X_1 X_2}(\omega)\} = 0 \quad (16)$$

This noise field is a good approximation of noise fields in reverberant rooms with highly reflective walls, if the reverberation time τ_{60} is larger than $300ms$ and the noise sources in the room have no direct path to the microphone array. In this case the noise reduction performance depends only on the microphone distance d, and the number of microphones N. Figures 2 and 3 show an example of the noise reduction as function of the frequency for different distances d and different numbers N of microphones. We can see that the noise reduction performance is always $> 1dB$. For $N = 2$ the noise reduction is zero dB at all frequencies; an analysis can be found in [17].

3. SIMULATION RESULTS

For the simulation we used the image method described by Allen and Berkley [18]. The simulated office room is $7m \times 5m \times 3.5m$.

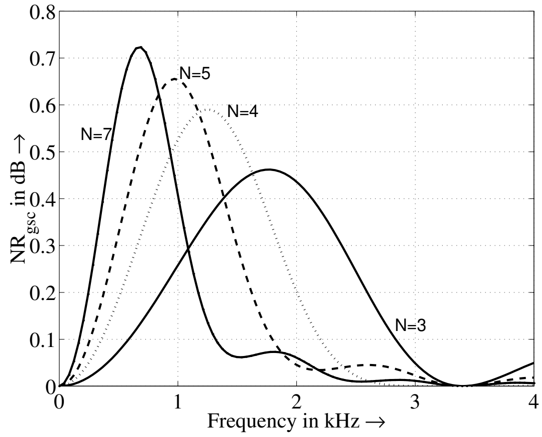

Figure 2: Noise reduction performance in a diffuse noise field for different numbers of microphones ($d = 5cm$)

The microphone array consists of five microphones in a distance of $5cm$. The white gaussian noise source was placed at 30 degrees out of the look-direction in distances of $2.5m$. Figure 4 shows the noise reduction as a function of the reverberation time τ_{60}. We can see that the GSC has a great potential for anechoic rooms. In contrast, for highly reverberant rooms the noise reduction is $> 1dB$. Similar results for implemented adaptive beamformers are given in [19]. The theoretical study of the GSC structure predicts exactly this behaviour. The noise reduction in the simulated noise field at $\tau_{60} = 300ms$ is only $1.2dB$ above the theoretical noise reduction for a perfectly diffuse noise field.

4. CONCLUSION

Noise reduction with microphone arrays is one possibility to enhance degraded speech. A well-known solution is the generalized sidelobe canceller. In this study we gave the theoretical limits for this structure. The results show that the GSC is capable of suppressing noise in anechoic rooms, but it does not work well in reverberant environments. For the usage of the GSC for speech enhancement and broadband noise reduction we recommend to analyse the noise field first to decide whether the GSC is the right choice. An overview of possible alternative algorithms and their limits are given in [20, 17].

5. REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, 1985.

[2] M. M. Goulding and J. S. Bird, "Speech Enhancement for

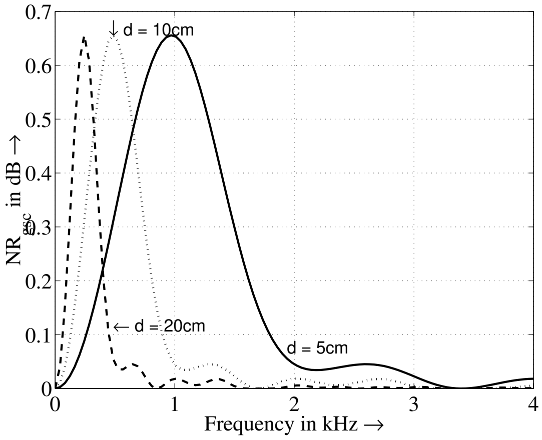

Figure 3: Noise reduction performance in a diffuse noise field for different distances of microphones ($N = 5$)

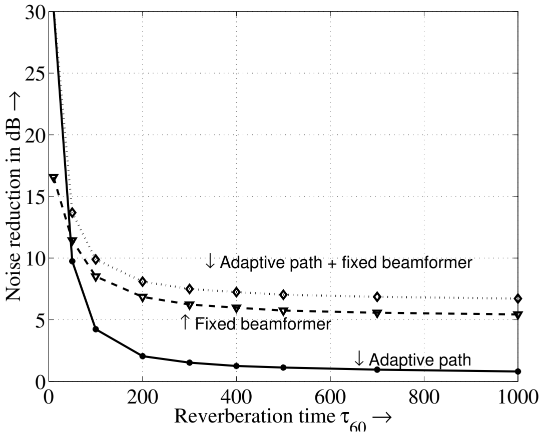

Figure 4: Noise reduction performance as function of the reverberation time τ_{60} ($N = 5, d = 5cm$)

Mobile Telephony," *IEEE Trans. on Vehicular Technology*, vol. 39, pp. 316–326, Nov 1990.

[3] W. Armbrüster, R. Czernach, and P. Vary, "Adaptive noise cancelling with reference input- possible applications and theoretical limits," in *Proc. EURASIP European Signal Proc. Conference (EUSIPCO)*, pp. 391–394, 1986.

[4] G. Faucon, S. Tazi Mezalek, and R. Le Bouquin, "Study and coparison of three structures for speech enhancement of noisy speech," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, (Glasgow, UK), pp. 385–188, 1989.

[5] L. J. Griffiths and C. W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Trans. on Antennas Propagation*, vol. 30, pp. 27–34, 1982.

[6] Y. Kaneda, "Adaptive microphone array system for noise reduction and its performance studies," in *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS)*, pp. 2395–2398, 1988.

[7] Y. Cao, S. Sridharan, and M. Moody, "Speech-seeking microphone array with multi-stage processing," in *Proc. ESCA European Conference on Speech Communication and Technology (EUROSPEECH)*, (Madrid, Spain), pp. 1991–1994, Sep 1995.

[8] O. Hoshuyama and A. Sugiyama, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, pp. 925–928, 1996.

[9] J. Greenberg and P. Zurek, "Adaptive beamformer performance in reverberation," in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, (Mohonk, USA), 1991.

[10] D. van Compernolle, W. Ma, F. Xie, and M. van Diest, "Speech Recognition in Noisy Environments with the Aid of Microphone Arrays," *EURASIP Speech Communication*, vol. 9, no. , pp. 433–442, 1990.

[11] S. Nordebo and S. Nordholm, "Noise reduction using an adaptive microphone array in a car - a speech recognition evaluation," in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, (Mohonk, USA), 1993.

[12] Y. Grenier, "A Microphone Array for Car Environments," *Speech Communication*, vol. 12, pp. 25–39, 1993.

[13] S. Nordholm, I. Claesson, and I. Bengtsson, "Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars," *IEEE Trans. on Vehicular Technology*, vol. 42, pp. 514–518, November 1993.

[14] S. Affes and Y. Grenier, "Test of Adaptive Beamformers for Speech Acqusition in cars," in *Int. Conference on Signal Processing Applications and Technology, ICSPAT 94*, (Dallas, Texas), pp. 154–159, October 1994.

[15] J. Meyer (Bitzer) and K. U. Simmer, "Multi-Channel Speech Enhancement in a Car Environment using Wiener Filtering and Spectral Subtraction," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, vol. 2, (Munich, Germany), pp. 1167–1170, April 1997.

[16] J. Meyer (Bitzer), K. U. Simmer, and K. D. Kammeyer, "Comparison of One- and Two-Channel Noise-Estimation Techniques," in *Proc. Int. Workshop on Acoustic Echo and Noise Control*, (London, Great Britain), pp. 17–20, September 1997.

[17] J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multichannel Noise Reduction –Algorithms and Theoretical Limits–," in *Proc. EURASIP European Signal Proc. Conference (EUSIPCO)*, (Rhodes, Greece), Sep 1998.

[18] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[19] S. Fischer and K. U. Simmer, "Beamforming Microphone Arrays for Speech Acquisition in Noisy Environments," *EURASIP Speech Communication*, vol. 20, pp. 215–227, December 1996.

[20] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of Noise Reduction and Dereverberation Techniques Based on Microphone Arrays with Postfiltering," *IEEE Trans. on Speech and Audio Processing*, vol. 6, pp. 240–259, May 1998.