

Friedrich-Alexander-Universität Erlangen-Nürnberg

**Lehrstuhl für Multimediakommunikation und
Signalverarbeitung**

Prof. Dr.-Ing. Walter Kellermann

Bachelorarbeit

**Kombination eines LCMV-Beamformers
mit einkanaligen Filtern zur
Unterdrückung von nicht punktförmigen
Rauschquellen**

von Sebastian Kram

April 2014

Betreuer: Stefan Meier

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Ort, Datum

Unterschrift

Inhaltsverzeichnis

| | |
|--|------------|
| Kurzfassung | V |
| Abkürzungsverzeichnis | VI |
| Formelzeichen | VII |
| 1 Einleitung | 1 |
| 2 Grundlagen der akustischen Signalverarbeitung | 3 |
| 2.1 Signalmodell | 3 |
| 2.2 Schätzung der Leistungsdichtespektren | 5 |
| 2.3 Modelle für räumliche Kohärenzfunktionen | 5 |
| 2.3.1 Inkohärentes Rauschfeld | 6 |
| 2.3.2 Sphärisch Isotropes Rauschfeld | 7 |
| 3 Grundlagen zur Geräuschreduktion | 8 |
| 3.1 Szenario | 8 |
| 3.2 Mehrkanaliges Wienerfilter | 9 |
| 3.3 Beamformer | 11 |
| 3.3.1 LCMV-Beamformer | 12 |
| 3.3.2 MVDR-Beamformer | 12 |
| 3.3.3 Generalized Sidelobe Canceller (GSC) | 13 |

| | | |
|----------|---|-----------|
| 4 | Nachfilter | 15 |
| 4.1 | Zelinski | 15 |
| 4.2 | Simmer | 16 |
| 4.3 | McCowan | 17 |
| 4.4 | Wolff | 17 |
| 4.5 | Begrenzung der Filterfunktionen | 19 |
| 5 | Bewertung der Algorithmen | 20 |
| 5.1 | Signal-Rausch-Verhältnis | 20 |
| 5.2 | Sprachverzerrung | 21 |
| 5.3 | Geräuschreduktion | 21 |
| 6 | Versuchsauswertung | 22 |
| 6.1 | Modellaufbau | 22 |
| 6.2 | Anwendung mit weißem Rauschen | 24 |
| 6.2.1 | Zelinski-Filter | 25 |
| 6.2.2 | Simmer-Filter | 26 |
| 6.2.3 | Mc Cowan-Filter | 27 |
| 6.2.4 | Wolff-Filter | 28 |
| 6.2.5 | Vergleich | 29 |
| 6.3 | Anwendung mit diffusem Rauschen | 30 |
| 6.3.1 | Zelinski-Filter | 31 |
| 6.3.2 | Simmer-Filter | 32 |
| 6.3.3 | Mc Cowan-Filter | 33 |
| 6.3.4 | Wolff-Filter | 34 |
| 6.3.5 | Vergleich | 34 |
| 6.4 | Abschließende Analyse | 35 |
| 7 | Zusammenfassung und Ausblick | 37 |
| | Abbildungsverzeichnis | 38 |

Tabellenverzeichnis

40

Literaturverzeichnis

42

Kurzfassung

In der folgenden Arbeit werden verschiedene Methoden zu Nachfilterung von Beamformer-Ausgangssignalen untersucht. Dazu werden zuerst theoretische Grundlagen zu räumlicher Filterung mit Mikrophonarrays, basierend auf dem MMSE (*Minimum Mean Square Error*)-Kriterium hergeleitet und in die Filterung von lokalisierbaren (durch Beamformer) und richtungslosen Störquellen (durch Nachfilter) zerlegt. Schließlich wird auf verschiedene Ansätze, das theoretisch hergeleitete Nachfilter zu schätzen, eingegangen. Diese Filter wurden mit Matlab[®]-Simulationen und in verschiedenen akustischen Situationen in Kombination mit einem LCMV- (*Linearly Constrained Minimum Variance*) Beamformer in GSC (*Generalized Sidelobe Canceller*)-Struktur getestet und nach verschiedenen Bewertungskriterien evaluiert, wobei sich das McCowan-Postfilter als geeigneter Ansatz, vor allem bei bekanntem Kohärenzmodell, hervorgetan hat.

Abkürzungsverzeichnis

| | |
|-------|---|
| STFT | Kurzzeit-Fourier-Transformation |
| GSC | <i>Generalized Sidelobe Canceller</i> |
| MMSE | Minimaler quadratischer Fehler |
| LCMV | <i>Linearly Constrained Minimum Variance</i> |
| MVDR | <i>Minimum Variance Distortionless Response</i> |
| IC | <i>Interference Canceller</i> |
| VAD | <i>Voice Activity Detection</i> |
| SNR | Signal-Rausch-Verhältnis |
| (F)BF | <i>(Fixed) Beamformer</i> |

Formelzeichen

| | |
|------------------|--|
| $\text{sinc}(x)$ | Sinc-Funktion ($\frac{\sin(x)}{x}$) |
| $\Re\{x\}$ | Realteil |
| \mathbf{x}^* | Komplex konjugierte |
| \mathbf{x}^H | Adjungierte (Transponierte der komplex konjugierten) |
| $E\{x\}$ | Erwartungswert-Operator |
| ∇ | Gradient |
| argmin | bestimmt die Stelle des Minimums |

Kapitel 1

Einleitung

Mit der wachsenden Popularität von Mobilkommunikationssystemen, vor allem in Kombination mit Spracherkennungsanwendungen, steigt der Qualitätsanspruch in diesem Bereich zunehmend. So soll auch bei lauten Umgebungsgeräuschen und ohne Mikrofon am Körper störungsfreie Interaktion mit technischen Geräten möglich sein. Zum Beispiel muss die Spracherkennungssoftware von Navigationssystemen auch bei lauten Verkehrsgeräuschen problemlos funktionieren oder Benutzer von Hörgeräten sogar in einem vollen Restaurant in der Lage sein, sich auf einen Sprecher zu konzentrieren. Eindimensionale Systeme erfüllen die gestellten Ansprüche schon lange nicht mehr, deshalb ist die Verwendung von Arrays aus Mikrofonen heute üblich, da sie die Lokalisation von Schallquellen und eine Verwendung der akustischen Eigenschaften des Raumes ermöglicht. Das beste Beispiel für ein solches System ist der Mensch selbst, der, mit zwei Ohren ausgestattet, in der Lage ist, die Richtung von Geräuschen zu ermitteln. Die nötige Signalverarbeitung übernimmt das Gehirn. In der Praxis werden bei der Geräuschreduktion zwei Typen von Störsignalen unterschieden: Gerichtete Rauschquellen, die eine bestimmbare Position besitzen (z.B.: Störsprecher) und nicht-punktförmige Rauschquellen wie z.B.: die Hintergrundgeräusche in einem Raum voller sprechender Menschen (*Babble Noise*). Somit sind zur Verbesserung des Sprachsignals zwei Signalverarbeitungsschritte notwendig: Erst müssen die gerichteten Schallquellen gedämpft werden, danach sollen noch die Hintergrundgeräusche herausgefiltert wer-

den. Die räumliche Selektion geschieht durch ein mehrkanaliges Filter, zur Nachfilterung wird beispielsweise ein einkanaliges Wiener-Filter verwendet, das allerdings in der Praxis nur als Schätzung realisierbar ist. Beim Entwurf ist hierbei zu beachten, dass bei der Unterdrückung der Störsignale die Natürlichkeit des Nutzsignals nicht zu sehr beeinflusst wird. In dieser Arbeit werden deshalb verschiedene Postfilter in Kombination mit einem vorhandenen Beamformer hergeleitet und untersucht. Außerdem wurde verschiedene Nachfilter als Matlab[®]-Code implementiert und in verschiedenen akustischen Szenarien getestet.

In **Kapitel 3** wird erst das vorliegende Szenario erläutert und danach einige theoretische Grundlagen erklärt. **Kapitel 2** führt den Ansatz der Geräuschreduktion basierend auf einem mehrkanaligen Wienerfilter ein und enthält außerdem eine Beschreibung von verschiedenen Beamformer-Typen. **Kapitel 4** beschäftigt sich mit der Herleitung und Realisation verschiedener Nachfilter. Anschließend wird in **Kapitel 5** auf die verwendeten Bewertungskriterien für die Filter eingegangen. In **Kapitel 6** wird danach der Simulationsaufbau eingeführt und die Ergebnisse der Simulation behandelt. Der Matlab[®]-Code zu den Postfiltern befindet sich als CD im Anhang. **Kapitel 7** schließt die Arbeit mit einer Zusammenfassung der Ergebnisse und einem Ausblick auf mögliche zukünftige Projekte ab.

Kapitel 2

Grundlagen der akkustischen Signalverarbeitung

Diese Kapitel behandelt einige für das Verständnis späterer Zusammenhänge notwendige Grundlagen. Außerdem wird auf verschiedene verwendete Modelle eingegangen.

2.1 Signalmodell

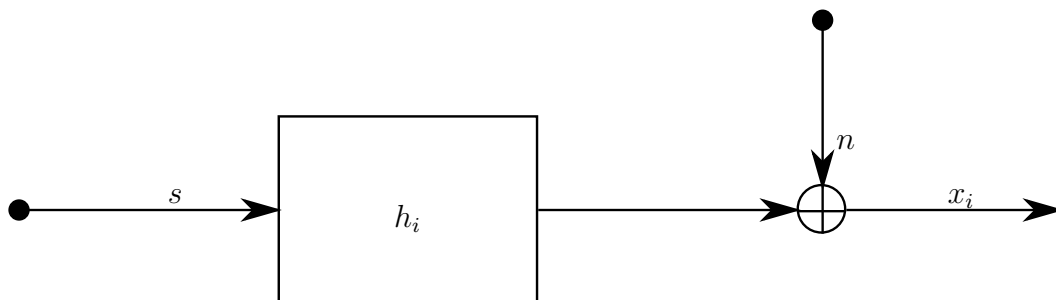


Abbildung 2.1: Blockdiagramm für ein Mikrophonsignal

Die zu bearbeiteten Signale in dieser Arbeit liegen werden mit Mikrofonen im Zeitbereich abgetastet und dann in den STFT (*Short Time Fourier Transform*)-Bereich transformiert. Theoretisches zur STFT findet sich u. a. in [5]. Die Signale werden

klein geschrieben und fett wenn es sich um Vektoren handelt, im Fall von Matrizen groß und fett. Die Vektoren \mathbf{x} , \mathbf{n} und \mathbf{h} besitzen hierbei P Elemente (P entspricht der Zahl der Mikrophone), sodass $\mathbf{x} = [x_1, x_2, \dots, x_P]$ gilt. Sowohl der Zeitframeindex als auch die Frequenzabhängigkeit werden, wenn nicht anders vermerkt, der Übersichtlichkeit wegen weggelassen. Der Ansatz für den Vektor \mathbf{x} der Mikrophone signale im STFT-Bereich ist:

$$\mathbf{x} = \mathbf{s}\mathbf{h} + \mathbf{n}, \quad (2.1)$$

mit dem Rauschen \mathbf{n} und dem Nutzsignal s .

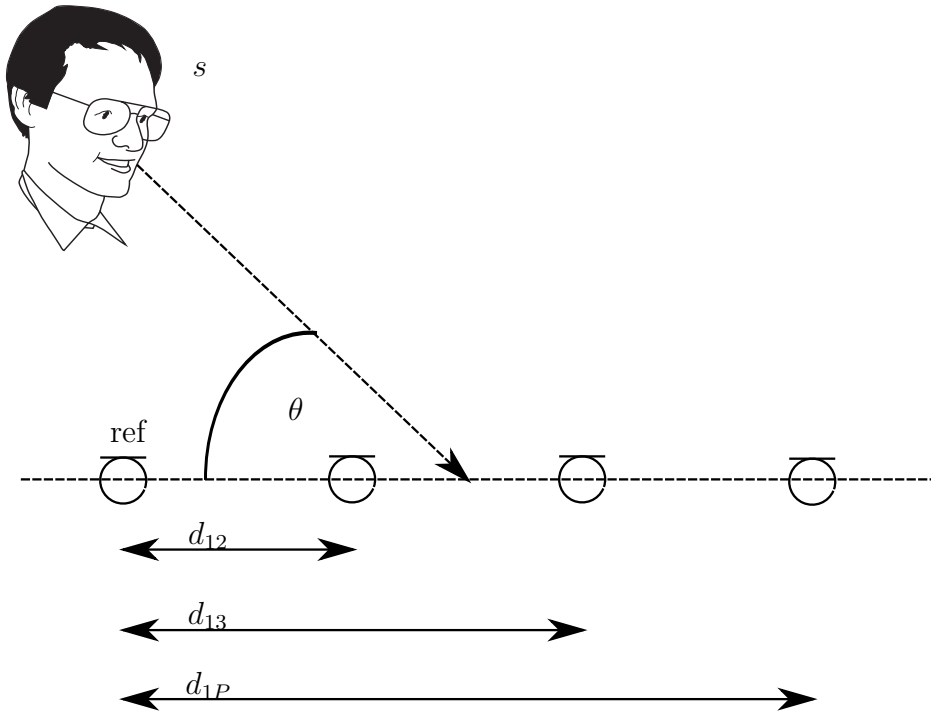


Abbildung 2.2: Laufzeitunterschiede zwischen den Mikrofonen am Beispiel eine vierkanaligen Mikrophonarrays

\mathbf{h} bezeichnet den Vektor der Impulsantworten der Pfade zwischen der Sprachquelle und den Sensoren und vereinfacht sich im einfachsten Fall zu einem Steuervektor \mathbf{d} , der nur die Phasenverschiebungen der Mikrophone enthält [12]:

$$\mathbf{d}(f, \theta) = [1, \exp(-j\omega\tau_{12}), \dots, \exp(-j\omega\tau_{1P})], \quad (2.2)$$

die sich abhängig von dem Einfallswinkel θ und der Mikrophondistanz d_{ij} folgendermaßen berechnen lassen [2]:

$$\omega\tau_{ij} = \frac{2\pi f}{c} \cos(\theta). \quad (2.3)$$

Das erste Mikrophon wird hierbei als Referenzmikrophon verwendet.

Eine weitere Annahme, die in dieser Arbeit getroffen wurde, ist, dass das Nutz- und Störsignal unkorreliert sind. Daraus folgt für die Kreuzleistungsdichten dieser beiden Signale [11]:

$$\phi_{ns} = \phi_{sn} = 0 \quad (2.4)$$

2.2 Schätzung der Leistungsdichtespektren

Für die Schätzung von Kreuz- und Autoleistungsdichten im STFT-Bereich wird üblicherweise die Methode der rekursiven Mittelung angewendet. Diese beruht auf der exponentiellen Gewichtung von vergangenen Kurzzeit-Fourier-Spektren durch Rekursion (m stellt hier den Zeitframeindex dar) [11]:

$$\hat{\phi}_{x_i,j}(m) = \alpha \hat{\phi}_{x_i,j}(m-1) + (1-\alpha)x_i(m)x_j^*(m). \quad (2.5)$$

Dabei gilt bei Autoleistungsdichten $i = j$. Ein geeigneter Glättungsfaktor α lässt sich, abhängig von der Nachhallzeit T_{60} des Raumes berechnen [12]:

$$\alpha = \exp\left(-\frac{1}{T_{60}f_s}\right). \quad (2.6)$$

Wobei T_{60} die Zeit darstellt, nach der die Energie des nachhallenden Signals um 60 dB geringer als die des Ursprungssignals ist [11].

2.3 Modelle für räumliche Kohärenzfunktionen

Die räumliche Kohärenzmatrix ist eine Möglichkeit, das vorliegende Rauschfeld zu charakterisieren. Diese Matrix enthält die normierten spektralen Korrelationen zwischen

einer definierten Anzahl von Rauschsignalen n_i und n_j an diskreten Punkten, also an den Mikrofonen:

$$\mathbf{\Gamma}_{n,n} = \begin{pmatrix} 1 & \Gamma_{n_1 n_1} & \dots & \Gamma_{n_1 n_P} \\ \Gamma_{n_2 n_1} & 1 & \dots & \Gamma_{n_2 n_P} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{n_P n_1} & \Gamma_{n_P n_2} & \dots & 1 \end{pmatrix}, \quad (2.7)$$

wobei die Elemente der Matrix durch die komplexe Kohärenzfunktion

$$\mathbf{\Gamma}_{n_i n_j} = \frac{\phi_{n_i n_j}}{\sqrt{\phi_{n_i n_i} \phi_{n_j n_j}}} \quad (2.8)$$

gegeben sind [12]. Da für die Berechnung der Leistungsdichtespektren des Rauschanteils allerdings die Rauschsignale getrennt vorhanden sein müssen, was in der Praxis nicht der Fall ist, begnügt man sich oft mit Modellannahmen. Im folgenden sollen zwei verschiedene Ansätze zur Modellierung des Rauschfelds vorgestellt werden. Es existieren weitere Modelle, die hier nicht erwähnt werden sollen. Beschreibungen finden sich z.B.: in [12] oder [2].

2.3.1 Inkohärentes Rauschfeld

Das einfachste Modell ist das unkorrelierte Rauschfeld. Hierbei wird davon ausgegangen, dass für die Kohärenzmatrix gilt [2]:

$$\mathbf{\Gamma}_{n_i n_j} = 0 \text{ für } i \neq j, \quad (2.9)$$

wodurch sich die Matrix zu einer Einheitsmatrix vereinfacht. Dieses einfache Modell gilt in realen Schallfeldern nur näherungsweise in bestimmten Frequenzen und bei hohen Mikrofonabständen, hat allerdings den Vorteil, das es die Kohärenzmatrix stark vereinfacht.

2.3.2 Sphärisch Isotropes Rauschfeld

Die andere Annahme ist die des isotropen oder diffusen Rauschfelds. Man geht hierbei davon aus, dass die Sensorpunkte (also die Mikrophone i und j) von einer Kugel- fläche mit einem unendlich hohen Radius umgeben sind, auf der sich unendlich viele punktförmige Schallquellen befinden. Die Kohärenz dieser beiden Punkte lässt sich somit in Abhängigkeit der Distanzen im Mikrophonarray (vgl. Abb. 2.2) $d_{i,j}$ berechnen [12]:

$$\Gamma_{n_i n_j}(f) = \text{sinc}\left(\frac{2\pi f d_{i,j}}{c}\right), \quad (2.10)$$

wobei c die Schallgeschwindigkeit darstellt. Dieses Modell gilt näherungsweise für zahlreiche Umgebungen (z.B.: Büroräume, Autos) [8]. Abb. 2.3 zeigt diesen Zusammenhang, die Näherung ist dabei der Sinus mit abfallender Amplitude, die stark unstetige Funktion stellt gemessene Kohärenz dar.

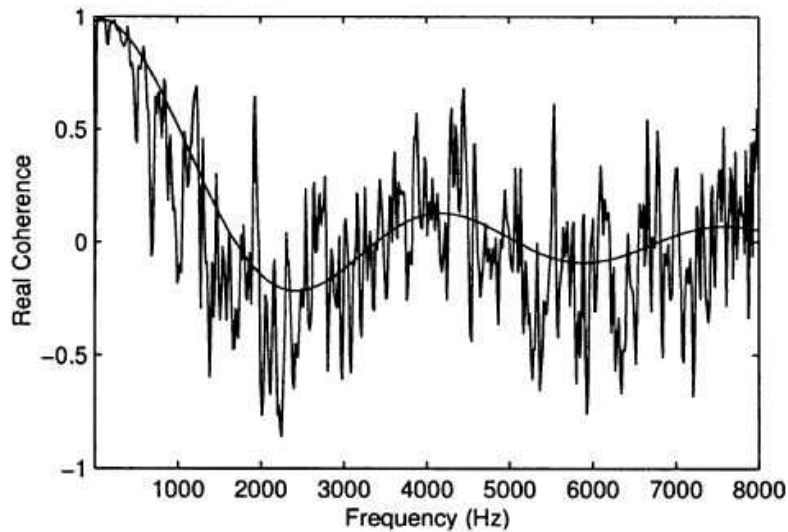


Abbildung 2.3: Kohärenzfunktion eines Büroraums im Vergleich zum Modell [8].

Kapitel 3

Grundlagen zur Geräuschreduktion

In diesem Kapitel werden die theoretischen Grundlagen der Geräuschreduktion erläutert. Dazu wird, ausgehend von der Beschreibung des Anwendungsszenarios, eine mathematische Herleitung des Ansatzes zur Minimierung des quadratischen Fehlers (MMSE) präsentiert. Dieser wird anschließend in zwei aufeinanderfolgende Strukturen geteilt: Einen Beamformer, um gerichtete Störquellen zu entfernen, und ein Nachfilter für Hintergrundrauschen. Danach wird auf verschiedene Entwurfsverfahren für Beamformer eingegangen.

3.1 Szenario

Abbildung 3.1 zeigt eine typische Problemstellung z.B. für eine Freisprechkommunikationsanlage. Das Sprachsignal des Sprechers ist hierbei sowohl durch einen oder mehrere Störsprecher, als auch durch Umgebungsrauschen überlagert, hinzu kommt die Mehrwegausbreitung (Nachhall) .

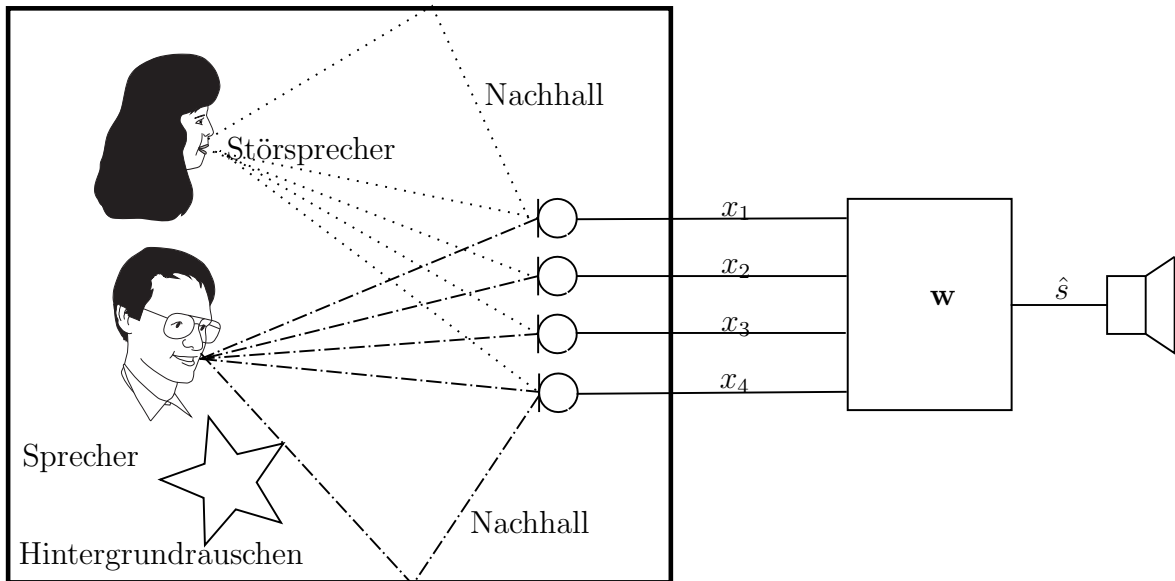


Abbildung 3.1: Anwendungsbeispiel eines Mikrophonrrays

Durch Signalverarbeitung sollen diese ungewünschten Effekte weitestgehend entfernt werden, ohne dass das Signal am Ausgang zu sehr verzerrt wird, das Problem wird hierbei in zwei Teile zerlegt, die im folgenden beschrieben werden sollen.

3.2 Mehrkanaliges Wienerfilter

Im Folgenden werden alle Signale, wenn es nicht anders vermerkt wird als Kurzzeit-Fourier-Transformationen angenommen. Der Übersichtlichkeit wegen werden die Frequenzabhängigkeiten nicht dargestellt. Von einem Sprachsignal s liegt ein Beobachtungsvektor \mathbf{x} vor, der die Eingangssignale der Mikrophone des Arrays enthält (vgl. Abbildung 3.1). Dieser soll durch die Filterung mit einem Vektor \mathbf{w} so verändert werden, dass die Abweichung des Filterausgangs gegenüber dem Nutzsignal minimal wird [7]:

$$e = s - \mathbf{w}^H \mathbf{x} \quad (3.1)$$

Im Fall des mehrkanaligen Wiener-Filters wird hierbei das MMSE-Kriterium (*Minimum Mean Square Error*) angewendet, also ist das Ziel eine Minimierung des quadratischen Fehlers. Dieser entspricht dem Leistungsdichtespektrum ϕ_{ee} des Fehlers, das sich aus Gleichung 3.1 ergibt, wobei zuerst die Kreuzleistungsdichte der verrauschten Signale mit dem Sprachsignal

$$\phi_{xs} = E\{\mathbf{x}s^*\} \quad (3.2)$$

und die Matrix

$$\Phi_{\mathbf{xx}} = E\{\mathbf{xx}^H\}, \quad (3.3)$$

der Autoleistungsdichten der Eingangssignale bestimmt werden müssen. Anschließend ergibt sich [11]:

$$\phi_{ee} = E\{\|e\|_2^2\} = \phi_{ss} - \mathbf{w}^H \phi_{xs} - \mathbf{w} \phi_{xs}^H + \mathbf{w}^H \Phi_{\mathbf{xx}} \mathbf{w}. \quad (3.4)$$

Um den idealen Filtervektor \mathbf{w}_{opt} zu finden, wird der Gradient von ϕ_{ee} gebildet und gleich Null gesetzt:

$$\nabla(\phi_{ee}) = -2\phi_{xs} + 2\Phi_{\mathbf{xx}} \mathbf{w} = 0. \quad (3.5)$$

Nach \mathbf{w} aufgelöst ergibt sich

$$\mathbf{w}_{\text{opt}} = \Phi_{\mathbf{xx}}^{-1} \phi_{xs}. \quad (3.6)$$

Mit dem in Kapitel 2.1 eingeführten Signalmodell und der Annahme der Unkorreliertheit des Sprach- und Rauschsignals ergeben sich die Leistungsdichten aus Gl. 3.2 und Gl. 3.3 zu

$$\phi_{xs} = E\{(s\mathbf{d} + \mathbf{n})s^*\} = \phi_{ss}\mathbf{d} \quad (3.7)$$

und

$$\Phi_{\mathbf{xx}} = E\{(s\mathbf{d} + \mathbf{n})(s^*\mathbf{d}^H + \mathbf{n}^H)\} = \phi_{ss}\mathbf{d}\mathbf{d}^H + \Phi_{\mathbf{nn}}. \quad (3.8)$$

Somit lässt sich der optimale Filtervektor auch folgendermaßen darstellen:

$$\mathbf{w}_{\text{opt}} = \phi_{ss}\mathbf{d}(\phi_{ss}\mathbf{d}\mathbf{d}^H + \Phi_{\mathbf{nn}})^{-1}. \quad (3.9)$$

Eine Anwendung der Sherman-Morrison-Woodbury-Formel, auf die hier nicht weiter eingegangen werden soll (eine komplette mathematische Herleitung findet sich in [2]) führt zu:

$$w_{\text{opt}} = \left[\frac{\phi_{ss}}{\phi_{ss} + \phi_{nn}} \right] \frac{\Phi_{nn}^{-1} \mathbf{d}}{\mathbf{d}^H \Phi_{nn}^{-1} \mathbf{d}}. \quad (3.10)$$

Das mehrkanalige Filter lässt sich nun in einen MVDR-Beamformer (*Minimum Variance Distortionless Response*),

$$\mathbf{w}_{\text{MVDR}} = \frac{\Phi_{nn}^{-1} \mathbf{d}}{\mathbf{d}^H \Phi_{nn}^{-1} \mathbf{d}}, \quad (3.11)$$

auf den im folgenden Kapitel eingegangen werden soll, und ein einkanaliges Wienerfilter

$$h_{\text{Post}} = \frac{\phi_{ss}}{\phi_{ss} + \phi_{nn}}, \quad (3.12)$$

als Postfilter, zerlegen. Da in der Praxis weder Stör- noch Sprachsignale bekannt sind, müssen Modellannahmen getroffen werden, die es ermöglichen, die Signale zu schätzen. Da eine Abwandlung des MVDR-Beamformers in GSC-Struktur in dieser Arbeit gegeben war, gilt es also, einen angemessenen Ersatz für das einkanalige Wiener-Filter zu finden.

3.3 Beamformer

Wenn Audiosignale mit einem Mikrofonarray erfasst werden, ist es möglich, neben den Zeit-und-Frequenzeigenschaften der Signale die Eigenschaften des Schallfeldes zu verwenden. Beamformer benutzen diese zur Unterdrückung von Rauschen, indem jedes Signal mit einer anderen Übertragungsfunktion w_i gefiltert wird und die Ausgänge anschließend aufsummiert werden. Die Übertragungsfunktionen werden üblicherweise im Filtervektor \mathbf{w}_{BF} zusammengefasst.

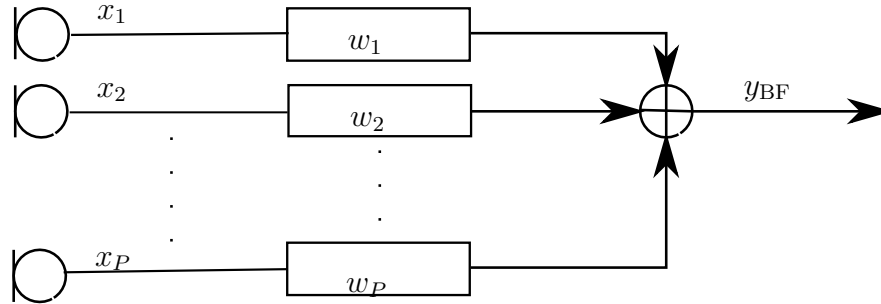


Abbildung 3.2: Blockdiagramm eines Beamformers

3.3.1 LCMV-Beamformer

Der in Gl. 3.11 eingeführte MVDR-Beamformer stellt einen Sonderfall eines LCMV- (*Linearly Constrained Minimum Variance*) Beamformers dar. Der Filtervektor \mathbf{w}_{BF} wird bei dieser Anwendung so gewählt, dass die Varianz am Ausgang, abhängig von $K < P$ Nebenbedingungen, minimal wird [6]:

$$\operatorname{argmin} \mathbf{w}_{BF}^H \Phi_{xx} \mathbf{w}_{BF}. \quad (3.13)$$

Die Nebenbedingungen werden durch die Matrix \mathbf{C}_{BF} der Dimension $P \times K$ und den Vektor \mathbf{c} beschrieben:

$$\mathbf{C}_{BF}^H \mathbf{w}_{BF} = \mathbf{c}. \quad (3.14)$$

Abhängig von den Nebenbedingungen kann eine Optimierung mit Lagrange-Multiplikation durchgeführt werden. Diese findet sich z.B.: in [6] und führt zu dem optimalen Filtervektor:

$$\mathbf{w}_{LCMV} = (\mathbf{C}_{BF}^H \Phi_{xx}^{-1} \mathbf{C}_{BF})^{-1} \Phi_{xx}^{-1} \mathbf{C}_{BF} \mathbf{c}. \quad (3.15)$$

3.3.2 MVDR-Beamformer

Wenn das Nutzsignal nun zusätzlich verzerrungsfrei übertragen werden soll, liegt ein MVDR-Beamformer wie in 3.2 vor. Dieses Kriterium lässt sich erfüllen, indem man

\mathbf{C}_{BF} gleich einem Vektor aus Raumübertragungsfunktionen \mathbf{f} (weitere Informationen zu Raumimpulsantworten/Raumübertragungsfunktionen finden sich z.B.: in [12]), und der Vektor \mathbf{c} gleich 1 gesetzt wird. Somit ist durch

$$\mathbf{f}_{\text{BF}}^H \mathbf{w}_{\text{BF}} \stackrel{!}{=} 1. \quad (3.16)$$

eine verzerrungsfreie Übertragung gewährleistet [3]. Dadurch reicht es aus, die Varianz nur hinsichtlich des Störanteils zu minimieren. Der Vektor der Raumübertragungsfunktionen lässt sich durch den Steuervektor \mathbf{d} approximieren (vgl. [6]). Somit ergibt sich analog zu 3.11:

$$\mathbf{w}_{\text{MVDR}} = \frac{\Phi_{nn}^{-1} \mathbf{d}}{\mathbf{d}^H \Phi_{nn}^{-1} \mathbf{d}}. \quad (3.17)$$

3.3.3 Generalized Sidelobe Canceller (GSC)

Die GSC (*Generalized Sidelobe Canceller*) -Struktur ist eine alternative Formulierung eines LCMV-Beamformers, bei der die Nebenbedingungen durch Orthogonalisierung entfernt werden. Dazu wird zuerst der Vektor des Beamformers in zwei Teile zerlegt [4]:

$$\mathbf{w}_{\text{LCMV}} = \mathbf{w}_{\text{F}} - \mathbf{w}_{\text{B}}, \quad (3.18)$$

wobei \mathbf{w}_{B} im Nullraum von \mathbf{C}_{BF} liegt:

$$\mathbf{C}_{\text{BF}}^H \mathbf{w}_{\text{B}} \stackrel{!}{=} 0, \quad (3.19)$$

und folglich \mathbf{w}_{F} die Nebenbedingungen erfüllen muss:

$$\mathbf{C}_{\text{BF}}^H \mathbf{w}_{\text{F}} = \mathbf{c}. \quad (3.20)$$

Dies gilt zum Beispiel für

$$\mathbf{w}_{\text{F}} = (\mathbf{C}_{\text{BF}}^H \mathbf{C}_{\text{BF}})^{-1} \mathbf{C}_{\text{BF}} \mathbf{c}. \quad (3.21)$$

Der zweite Vektor \mathbf{w}_{B} wird nun als Multiplikation der Blockiermatrix \mathbf{B} , die die Nullraumbedingung aus Gl. 3.19 erfüllt ($\mathbf{B}^H \mathbf{C}_{\text{BF}} \stackrel{!}{=} 0$), und des sogenannten *Interference*

Canceller dargestellt. Dieser kann, ohne Betrachtung der Nebenbedingungen, die schon durch \mathbf{w}_F und \mathbf{B} erfüllt sind, optimiert werden. Wegen der Umformulierung ist es möglich, einfachere Algorithmen zur Optimierung zu verwenden ([11]), die zu

$$\mathbf{w}_{IC} = (\mathbf{B}^H \Phi_{xx} \mathbf{B})^{-1} \mathbf{B}^H \Phi_{xx} \mathbf{w}_F. \quad (3.22)$$

führen. Der tatsächlich verwendete Beamformer basiert auf dem TRINICON-Framework ([1]) basierende Abwandlung eines LCMV-Beamformers, die ebenso zu einer GSC-Struktur führt ([9]). Da dieses Verfahren allerdings den Umfang dieser Arbeit übersteigen würde, wird es hier nicht weiter erklärt und als gegeben angenommen. Ein Blockschaltbild eines GSC ist in Abb. 3.3 dargestellt.

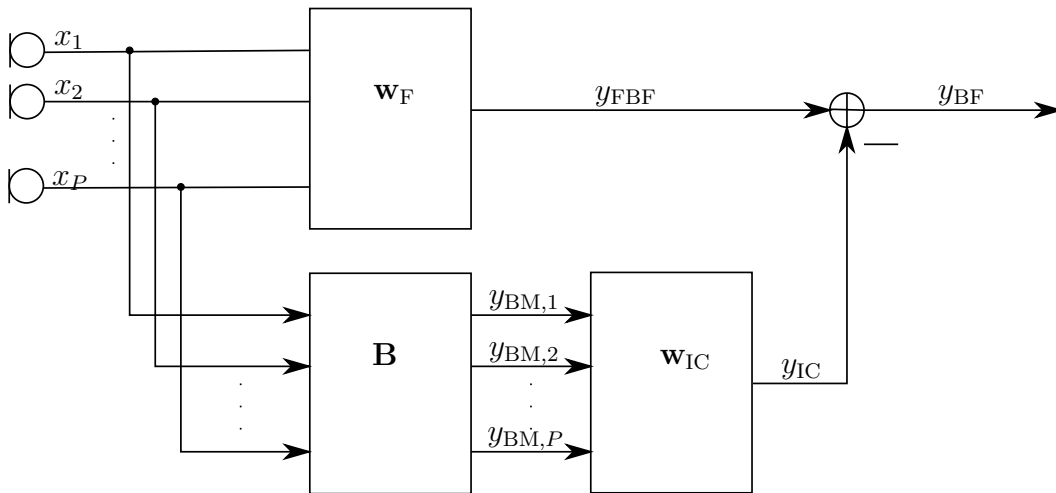


Abbildung 3.3: GSC-Struktur

Kapitel 4

Nachfilter

Dieses Kapitel enthält verschiedene Ansätze zur adaptiven Nachfilterung von Beamformer-Ausgangssignalen. Das Ziel der Filter ist es, die in Gl. 3.12 hergeleitete Übertragungsfunktion mithilfe von bekannten Größen und Modellannahmen zu schätzen. Um das zu erreichen, gehen die meisten Filter von festen Annahmen über das Rauschfeld aus und sind folglich nur in bestimmten Szenarien effektiv.

4.1 Zelinski

Das 1988 von Rainer Zelinski in [13] veröffentlichte Nachfilterverfahren arbeitet unter der Annahme einer schwachen Kohärenz der Rauschsignale an den Mikrofonen und benutzt deshalb das Modell eines inkohärenten Rauschfelds (vgl. Kap. 2.3). Zuerst wird davon ausgegangen, dass der Beamformer die Laufzeitunterschiede der Sprachsignalkomponenten an den Sensoreingängen ausgeglichen hat und somit gilt:

$$\mathbf{x} = s + \mathbf{n}. \quad (4.1)$$

Somit ergeben sich die Autoleistungsdichten der Mikrophonsignale zu:

$$\phi_{x_i x_i} = \phi_{ss} + \phi_{n_i n_i} + 2\Re\{\phi_{sn_i}\}, \quad (4.2)$$

und die Kreuzleistungsdichten zu:

$$\phi_{x_i x_j} = \phi_{ss} + \phi_{n_i n_j} + \phi_{s n_j} + \phi_{n_i s}. \quad (4.3)$$

Wenn nun davon ausgegangen wird, dass das Sprachsignal und das Rauschen unkorreliert sind (vgl. Kap. 2.1), die Rauschleistungsdichte an allen Mikrofonen gleich ist ($\phi_{n_i n_i} = \phi_{nn}$) und die Störsignalkomponenten unkorreliert sind ($\phi_{n_i n_j} = 0, \forall i \neq j$), vereinfachen sich die Leistungsdichten zu

$$\phi_{x_i x_i} = \phi_{ss} + \phi_{nn} \quad (4.4)$$

und

$$\phi_{x_i x_j} = \phi_{ss}. \quad (4.5)$$

Nach Einsetzen der Leistungsdichten in die Übertragungsfunktion des Wiener-Filters (vgl. Formel 3.12) ergibt sich, nach Mittelung über die relevanten Sensorkombinationen, für das Postfilter [13]:

$$\hat{h}_{p,Zel} = \frac{\frac{2}{P(P-1)} \sum_{i=1}^{P-1} \sum_{j=1}^P \Re\{\hat{\phi}_{x_i x_j}\}}{\frac{1}{P} \sum_{i=1}^P \hat{\phi}_{x_i x_i}}. \quad (4.6)$$

4.2 Simmer

1992 wurde von Simmer und Wasiljef in [10] ein Postfilter entworfen, das, genau wie das Zelinski-Filter, das Modell des inkohärenten Schallfelds benutzt (vgl. Kap. 2.3.1). Da es bei Anwendung des Zelinski-Filter leicht zu einer Überschätzung der Störsignalkomponenten und damit zu einer Verzerrung des Nutzsignals kommen kann, wurde der Vorschlag veröffentlicht, im Nenner die Leistung des Beamformer-Ausgangs $\phi_{y_{BF} y_{BF}}$ einzusetzen. Eine genaue Herleitung findet sich z.B.: in [2] Es ergibt sich für die Übertragungsfunktion [10] :

$$\hat{h}_{p,Sim} = \frac{\frac{2}{P(P-1)} \sum_{i=1}^{P-1} \sum_{j=1}^P \Re\{\hat{\phi}_{x_i x_j}\}}{\frac{1}{P} \sum_{i=1}^P \phi_{y_{BF} y_{BF}}}. \quad (4.7)$$

Im Falle eines unkorrelierten Schallfeldes wird genau das Wienerfilter aus Gl. 3.12 erreicht [12]. Eine schächere Geräuschreduktion als beim Zelinski-Filter wird durch eine geringere Sprachverzerrung ausgewogen [12].

4.3 McCowan

Während das Zelinski-Nachfilter von der Annahme eines unkorrelierten Rauschfelds ausgeht, verwendet das 2003 von Ian McCowan in [8] veröffentlichte Postfilter ein sphärisch isotropes, lässt allerdings auch die Möglichkeit offen, andere Modelle für das Rauschfeld zu benutzen. Damit stellt das McCowan-Filter eine Verallgemeinerung des Zelinski-Filters für kohärente Rauschfelder dar und ist somit, vor allem bei tiefen Frequenzen und geringen Sensorabständen, optimaler. Das bekannte Modell für die Kohärenzmatrix (vgl. Gl. 2.10) lässt sich mithilfe von Gleichung 2.8 benutzen um für die Schätzung des Sprachsignals folgenden Zusammenhang zu erhalten [8]:

$$\hat{\phi}_{ss}^{ij} = \frac{\Re\{\hat{\phi}_{x_i x_j}\} - \frac{1}{2}\Re\{\Gamma_{n_i n_j}\}(\hat{\phi}_{x_i x_i} + \hat{\phi}_{x_j x_j})}{1 - \Re\{\Gamma_{n_i n_j}\}}. \quad (4.8)$$

Dadurch ändert sich Gleichung 4.6 zu:

$$\hat{h}_p = \frac{\frac{2}{P(P-1)} \sum_{i=1}^{P-1} \sum_{j=1}^P \hat{\phi}_{ss}^{ij}}{\frac{1}{P} \sum_{i=1}^P \hat{\phi}_{x_i x_i}}, \quad (4.9)$$

somit liegt, falls ein inkohärentes Rauschfeld vorliegt ($\Gamma_{nn}=\mathbf{I}$) ein Zelinski-Filter vor, da in diesem Fall $\hat{\phi}_{ss}^{ij} = \hat{\phi}_{x_i x_i}$ gilt. Um eine Division durch Null in Gleichung 4.8 zu vermeiden muss in der Praxis die Kohärenzmatrix auf Werte < 1 zu beschränkt werden [2].

4.4 Wolff

Das von Christian Wolff entwickelte Nachfilter ist für den Einsatz in unterschiedlichen Szenarien ausgelegt. Es verwendet sowohl die Sprachsignal- als auch die Rauschsignal-

kohärenz adaptiv. Alle Formeln in diesem Kapitel wurden aus [12] übernommen. Die Schätzvorschrift für das Nutzsignal ist hierbei:

$$\hat{\phi}_{ss} = \max[\hat{\phi}_{xx} - \hat{G}_{eq} \sum_{q=0}^{P-1} \hat{\phi}_{y_{BM}y_{BM}}, 0], \quad (4.10)$$

wobei $y_{BM}y_{BM}$ die gemittelte Leistungsdichte der Ausgänge der Blockiermatrix darstellt und ϕ_{xx} als identische Autokreuzleistungsdichte für alle Mikrophonsignale angenommen wird. Die spektrale Anpassung G_{eq} lässt sich mithilfe der Blockiermatrix (vgl. Kapitel 3.3.3), der räumlichen Kohärenz und dem Spuroperator $\text{tr}\{\dots\}$ folgendermaßen berechnen:

$$\hat{G}_{eq} = \text{tr}\{\mathbf{B}\Gamma_{nm}\mathbf{B}^H\}^{-1}. \quad (4.11)$$

Eine rekursive Methode zur Schätzung von G_{eq} erfordert eine Detektion von Sprachpausen. Diese wurde bei der Implementierung vereinfacht mit dem gegebenen reinen Sprachsignal durchgeführt, wobei Kurzzeitspektren mit niedrigen Leistungsdichtespektren als Sprachpausen interpretiert wurden und die Werte an diesen Stellen dementsprechend ausgeschlossen wurden. Algorithmen zur Sprachpausenerkennung (VAD, *Voice Activity Detection*) finden sich z.B.: in [11]. Wenn keine Sprachpausen detektiert wurden, bleibt der Anpassungsfaktor gleich, sonst ändert er sich zu:

$$\hat{G}_{eq}(m) = \hat{G}_{eq}(m-1)(1 + \epsilon \text{sign}(\hat{\phi}_{xx} - y_{BM}y_{BM}\hat{G}_{eq}(m-1))). \quad (4.12)$$

Für den ersten Wert von ϵ ist ein Wert $\ll 1$ zu wählen. Das Nachfilter ergibt sich zu (Eine genaue Herleitung findet sich in [12]):

$$h_{P,Wolff} = \frac{\hat{\phi}_{ss}}{\hat{\phi}_{ss} + \beta(\hat{\phi}_{y_{BF}y_{BF}} + \hat{\phi}_{ss})}. \quad (4.13)$$

Der Überschätzfaktor β wird hierbei benutzt, um *musical tones* zu vermeiden. Der Schätzwert des Sprachsignals ergibt sich bei einer inkorrekten Schätzung des Rauschens:

$$\hat{\Phi}_{ss} = \hat{\phi}_{xx} + \phi_{nn} - \hat{\phi}_{nn}. \quad (4.14)$$

Den Fehler $e_n = \phi_{nn} - \hat{\phi}_{nn}$ gilt es hierbei zu minimieren. Hierzu wird ein modifizierter Schätzwert

$$\hat{\phi}_{nn} = \hat{\phi}_{nn} + K \hat{\phi}_{ss} \quad (4.15)$$

benutzt. Anschließend ist zu bestimmen, bei welchem K der mittlere quadratische Fehler gegenüber ϕ_{nn} minimal wird. Es ergibt sich der Optimalwert

$$K_{\text{opt}} = \frac{E\{e_n^2\}}{E\{\phi_{ss}^2\} + E\{e_n^2\}}, \quad (4.16)$$

mit dem sich der Überschätzungskoeffizient

$$\beta = 1 + K_{\text{opt}} \frac{\hat{\phi}_{ss}}{\hat{\phi}_{nn}} \quad (4.17)$$

berechnen lässt.

4.5 Begrenzung der Filterfunktionen

Da es in manchen Fällen, wie bei einer großen Varianz des Nutzsignals oder instationären Störsignalen, zu schwerwiegenden Fehlern bei der Schätzung der Leistungsdichtespektren kommen kann ist es notwendig, die Übertragungsfunktionen der Postfilter betragsmäßig auf einen Wertebereich zwischen 0 und 1 bzw. einem entsprechend kleinen Wert ϵ zu begrenzen [2]:

$$\epsilon \leq h_{\text{post}} \leq 1. \quad (4.18)$$

Kapitel 5

Bewertung der Algorithmen

Es gibt verschiedene Kriterien, die bei der Bewertung von Filteralgorithmen berücksichtigt werden können. So ist neben dem Ausgangs-Signal-Störverhältnis (SNR_{out}) bzw. dem SNR-Gewinn (SNR_{gain}) auch von Bedeutung, wie sehr das Sprachsignal verzerrt wird und wie natürlich das Restgeräusch klingt. Obwohl die zuverlässigste Bestimmung der letzten beiden Kriterien nur durch die subjektive Meinung des Betrachters bestimmbar ist, kann eine grobe Aussage mit mathematischen Modellen getroffen werden. In dieser Arbeit wurden nur objektive mathematische Methoden verwendet, Informationen zur subjektiven Analyse von Sprachsignalen mit Testhörern finden sich in [2] oder [11]. Im folgenden werden alle Signale als mit der Zeitvariable k abgetastet angenommen. In der Praxis können alle Größen nur geschätzt werden, da allerdings bei der Simulation Stör- und Nutzanteil getrennt gefiltert und am Ausgang untersucht werden können konnten die reinen Signale für die Bewertungskriterien benutzt werden.

5.1 Signal-Rausch-Verhältnis

Das Signal-Rauschverhältnis (SNR) ist eine der in der Signalverarbeitung am öftesten verwendeten Methoden zur Bewertung der Qualität eines Signals. Es ist der dekadischen

Logarithmus des Verhältnisses der Leistung des Sprach- und Rauschsignals (s und n):

$$\text{SNR} = 10 \log_{10} \frac{\sum_k s^2(k)}{\sum_k n^2(k)} [\text{dB}]. \quad (5.1)$$

Relevant ist auch der SNR-Gewinn (SNR Gain), also die Zunahme des SNR am Ausgang.

$$\text{SNR}_{\text{gain}} = \text{SNR}_{\text{out}} - \text{SNR}_{\text{in}}. \quad (5.2)$$

5.2 Sprachverzerrung

Die Sprachverzerrung SD ist ein Maß für die klangliche Natürlichkeit eines Signals und beschreibt im wesentlichen die betragliche Veränderung des Sprachsignals. Sie wurde für diese Arbeit folgendermaßen definiert:

$$\text{SD} = 10 \log_{10} \frac{\sum_k s_{\text{in}}^2(k)}{\sum_k s_{\text{out}}^2(k)} [\text{dB}]. \quad (5.3)$$

Hohe Werte bedeuten hierbei eine Dämpfung des Signales, negative eine Verstärkung. Ideal sind Werte bei ca. 0.

5.3 Geräuschreduktion

Die Geräuschreduktion NR ist eine Größe, die, analog zur Sprachverzerrung, die betraglichen Veränderungen des Störanteils darstellt.

$$\text{NR} = 10 \log_{10} \frac{\sum_k n_{\text{in}}^2(k)}{\sum_k n_{\text{out}}^2(k)} [\text{dB}]. \quad (5.4)$$

Ein hoher Wert ist hierbei ein Indikator für eine geringe Rauschleistung am Ausgang und somit wünschenswert.

Kapitel 6

Versuchsauswertung

Im folgenden Kapitel wird zuerst der Aufbau der durchgeführten Matlab[®]-Simulation erklärt. Anschließend folgt eine ausführliche Analyse der erzielten Ergebnisse, wobei für weißes und diffuses Rauschen jeweils alle Postfilter nach den Bewertungskriterien in verschiedenen Raumtypen ausgewertet und auch mit einem einkanaligen Wiener-Filter verglichen werden.

6.1 Modellaufbau

Der Modellaufbau für die Simulationen ist in Abb. 6.1 dargestellt. Die gesamte GSC-Struktur (vgl. Kapitel 3.3.3) einschließlich der Simulation des akustischen Szenarios (vgl. Kapitel 3.1) lag als Matlab-Implementierung vor. Diese wurde von Klaus Reindl bei der Arbeit an [9] erstellt, entsprechend finden sich weitere Informationen zur Struktur in dieser Quelle. Das Modell beinhaltete 3 Sprachsignale: Eine männliche (s), eine weibliche (f) und eine Kinderstimme (k), wobei die des Mannes als Nutzsignal und die anderen beiden als Störsprecher verwendet wurden.

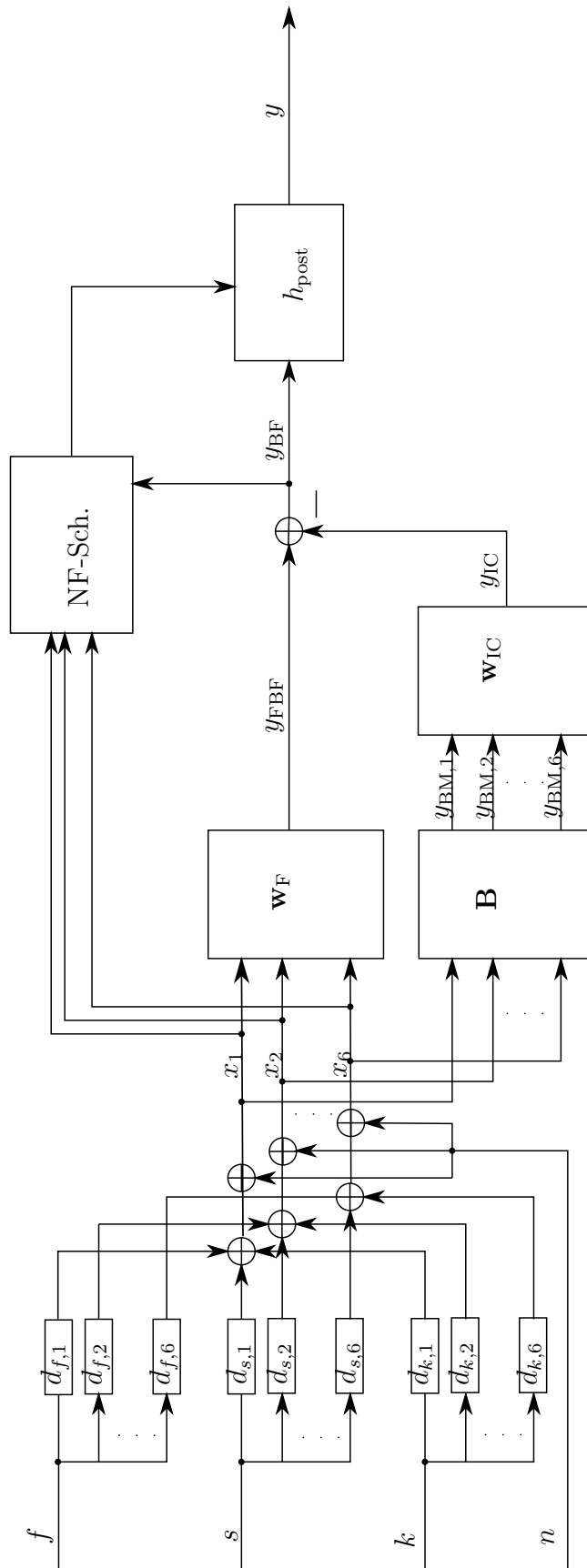


Abbildung 6.1: Modellaufbau

Es konnte zwischen weißem und diffusem Rauschen unterschieden werden. Für die Modellierung des Raumes lagen sowohl synthetische als auch gemessene Raumimpulsantworten d vor, von denen nur die gemessenen verwendet wurden. Hierbei waren vor allem die verschiedenen Nachhallzeiten von Bedeutung. Es wurden 6 Mikrophoneingänge simuliert. Die Abtastfrequenz betrug für alle Signale 16 kHz. Im Frequenzbereich wurden Kurzzeit-Fourier-Spektren (STFTs) mit einer STFT-Länge von 512 benutzt. Der Anteil der Störsprecher an den Mikrophonsignalen x_i wurde durch die Kombination aus Blockiermatrix und *Interference Canceller* geschätzt und dann vom Ausgang des festen Beamformers y_{FBF} abgezogen (vgl. Kap. 3.3). Das resultierende Signal y_{BF} sollte nun durch die Anwendung der in Kapitel 4 beschriebenen Postfilter auch von Hintergrundgeräuschen befreit werden. Für die Nachfilter-Schätzung wurden hierbei sowohl die Mikrophonsignale als auch das Ausgangssignal der Blockiermatrix y_{BM} und der GSC-Struktur y_{BF} verwendet. Nach der Filterung wurden die Signale nach den in Kap. 5 beschriebenen Kriterien bewertet. Neben den in Kap. 4 erklärten Filtern wurde als Vergleichswert auch ein einkanaliges Wienerfilter (vgl. Kap. 3.2) implementiert. Die Implementierung des Wolff-Filters gestaltete sich wegen der vielen zu schätzenden Parameter, von denen für viele vereinfachte Annahmen getroffen wurden, statt sie adaptiv zu berechnen, schwierig. Dieses Filter schnitt deshalb im Vergleich wesentlich schlechter als erwartet ab, weshalb nur ein Teil der Ergebnisse dargestellt wird. In [12] finden sich Versuchsergebnisse für das Filter in verschiedenen Szenarios, die wesentlich vielversprechender sind.

6.2 Anwendung mit weißem Rauschen

Im ersten Versuchsdurchlauf wurde für das Rauschmodell weißes Rauschen verwendet. Das Signal-Rausch-Verhältnis am Eingang des Beamformers wurde zwischen -5 und 20 variiert und es wurden 3 Raumtypen verwendet: Ein hallarmer Raum mit einer Nachhallzeit von 50 ms und zwei gewöhnliche Räume mit Nachhallzeiten von 200 bzw. 400 ms, was z.B.: für gewöhnliche Büroräume realistisch ist.

6.2.1 Zelinski-Filter

Das Zelinski-Filter zeigt in der hallarmen Umgebung erwartungsgemäß ein wesentlich höheres SNR, da dort die getroffenen Annahme eine inkohärenten Rauschfelds wesentlich besser erfüllt ist als bei den höheren Nachhallzeiten.

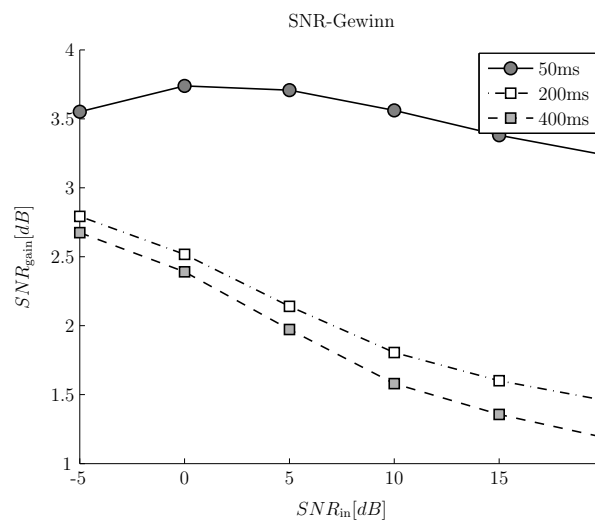


Abbildung 6.2: SNR-Gewinn des Zelinski-Postfilters

Allerdings bewirkt eine Überschätzung der Rauschens bei einer erhöhten Geräuschreduktion im hallarmen Raum eine größere Sprachverzerrung, in den anderen Raumtypen liegt fast eine konstante Sprachverzerrung vor. Bei erhöhtem Eingangs-SNR weist das Filter in den beiden hallstärkeren Räumen nur einen SNR-Gewinns-Abfall von ca. 1,5 dB von auf, während im hallarmen Raum bei mittleren SNRs sogar ein stärkerer Gewinn vorliegt, als bei niedrigen, und bei hohen SNRs kaum ein geringerer erzielt wird als bei niedrigen.

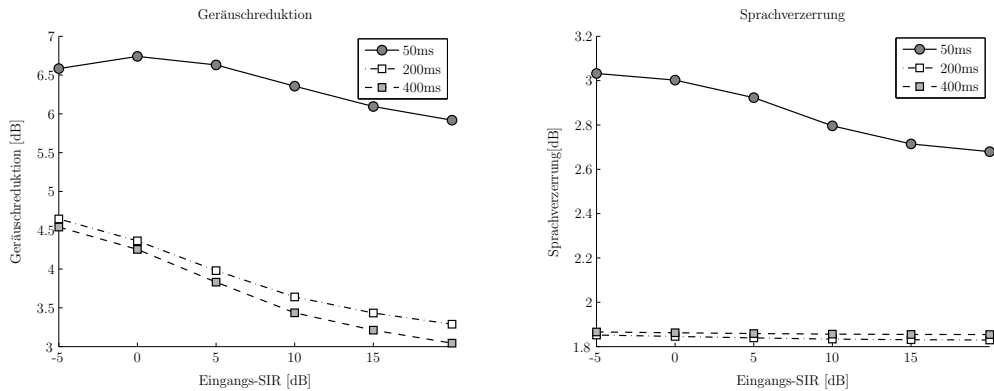


Abbildung 6.3: Geruschreduktion und Sprachverzerrung des Zelinski-Postfilters

6.2.2 Simmer-Filter

Das Simmer-Nachfilter erzeugt einen wesentlich geringeren SNR-Gewinn als die anderen Modelle. Hervorzuheben ist hierbei das fast konstante SNR im hallarmen Raum, dass eine starke ahnlichkeit zum Zelinski-Filter besitzt.

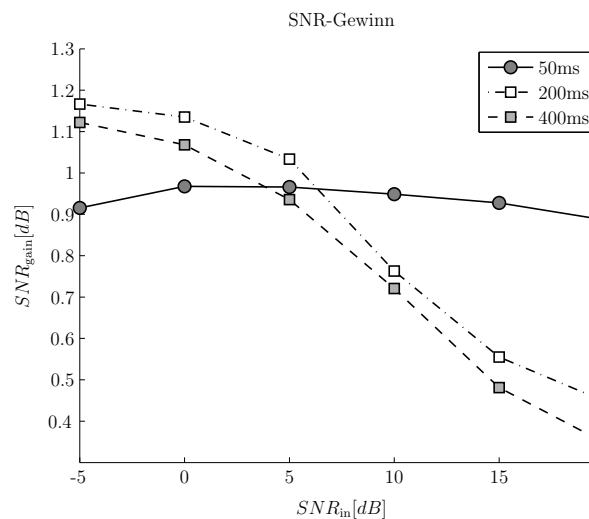


Abbildung 6.4: SNR-Gewinn des Simmer-Postfilters

Die Geruschreduktion ist wesentlich geringer als bei anderen Filtern, dafur wird jedoch auch die Sprache weniger stark verzerrt. Auch hier liegt im hallarmen Raum ein stark

unterschiedlicher SNR-Gewinn-Verlauf vor, der einen fast konstanten Wert mit einem Maximum bei mittleren SNR-Werten aufweist.

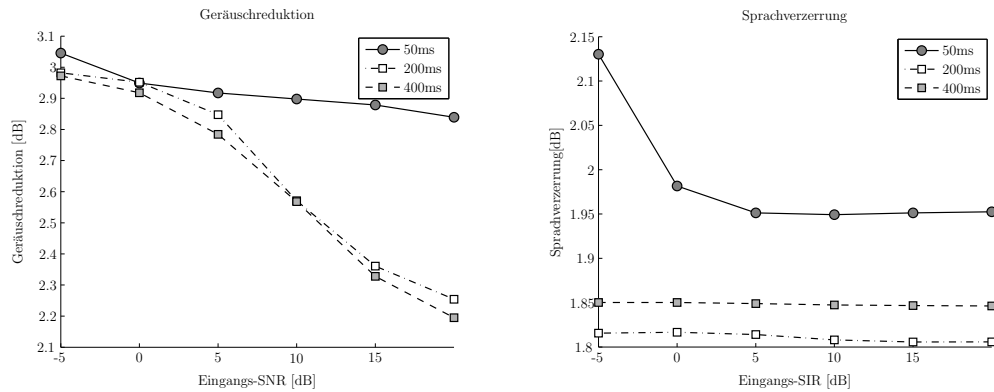


Abbildung 6.5: Geruschreduktion und Sprachverzerrung des Simmer-Postfilters

6.2.3 Mc Cowan-Filter

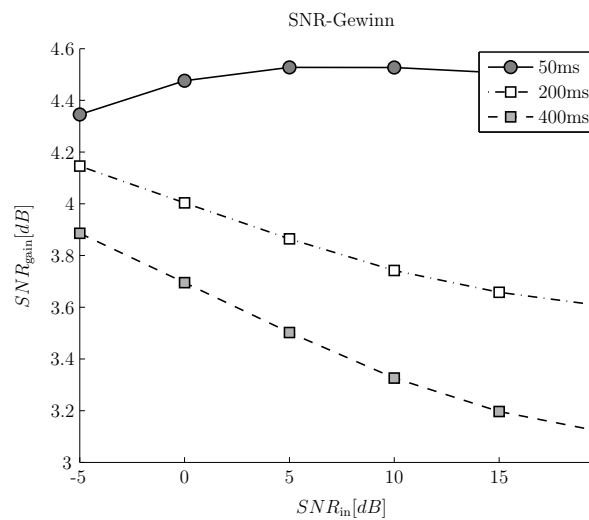


Abbildung 6.6: SNR-Gewinn des McCowan-Postfilters

Das McCowan-Nachfilter erzeugt ein hohes SNR in allen Raumtypen und zeigt eine vom Eingangs-SNR fast unabhangige Sprachverzerrung. Im hallarmen Raum liegt die Geruschreduktion fast konstant bei einem Wert von ca. 7,8. Im allgemeinen zeigt das

Filter, mit durchschnittlich über 3 dB höherem SNR-Gewinn als das Simmer-Filter und ca. 1 dB höherem Gewinn als das Zelinski-Filter, und einer hohen Geräuschreduktion bei nur geringfügig unterschiedlicher Sprachverzerrung, wesentlich bessere Eigenschaften als die anderen Filter.

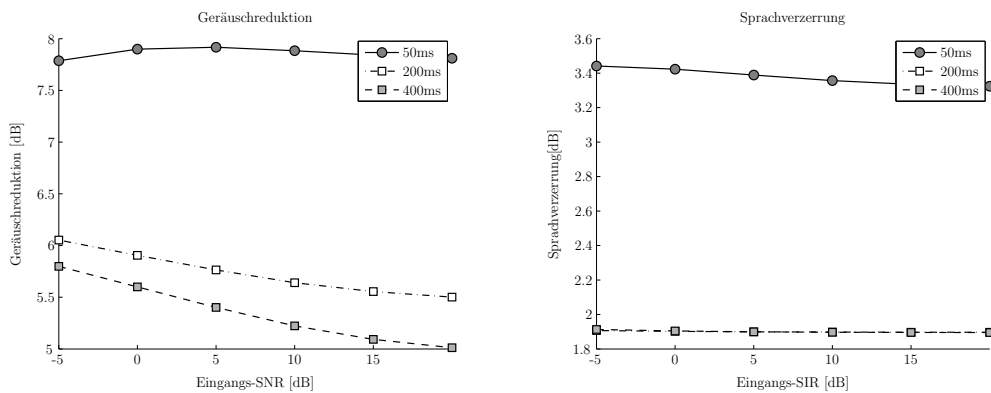


Abbildung 6.7: Geräuschreduktion und Sprachverzerrung des McCowan-Postfilters

6.2.4 Wolff-Filter

Beim Wolff-Filter wurde ein konstanter Wert von $\beta = 8$ für den Überschätzungsfaktor verwendet, um die Anzahl der Ergebnisse in einem angemessenen Rahmen zu halten. Möglichkeiten für die iterative Berechnung von β und Test mit anderen Werten finden sich in [12].

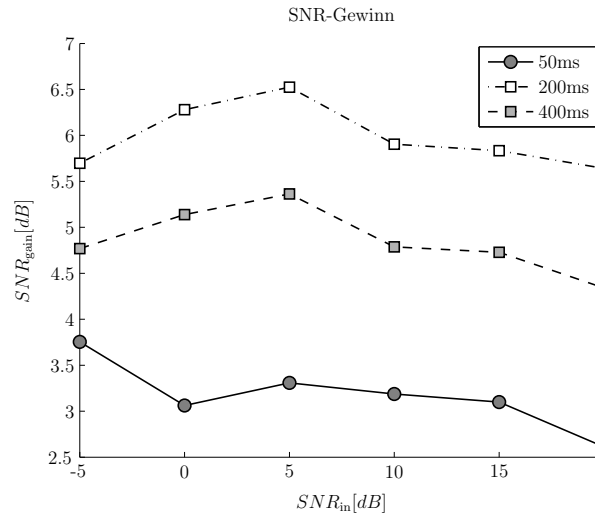


Abbildung 6.8: SNR-Gewinn des Wolff-Postfilters

Das Wolff-Filter zeigte eine unerwartete Unbeständigkeit im Bezug auf SNR-Gewinne. Allerdings liegt vor allem in den Räumen mit mehr Hall ein sehr hohes SNR vor.

6.2.5 Vergleich

Bei weißem Rauschen zeigt somit das McCowan-Filter den stärksten SNR-Gewinn bei einer Sprachverzerrung, die nur etwas stärker als die des Zelinski-Filters ist. Das Simmer-Filter besitzt eine Sonderrolle: Es erzielt nur geringe SNR-Gewinne, allerdings bleibt dafür die Natürlichkeit des Sprachsignals im hallarmen Raum weiter erhalten als bei den andern Ansätzen.

Tabelle 6.1: Werte im Hallarmen Raum bei 0 dB Eingangs-SNR in dB

| | SNR _{gain} | SD | NR |
|----------|---------------------|-----|------|
| Zelinski | 3.7 | 3.0 | 6.7 |
| Simmer | 0.95 | 1.9 | 2.95 |
| McCowan | 4.5 | 3.4 | 7.8 |

Im 200ms und 400ms-Raum weisen alle Filter einen sehr ähnlichen Verlauf auf. Das

McCowan-Filter erzeugt mit Abstand das höchste Ausgangs-SNR bei einer nur etwas höheren Sprachverzerrung.

Tabelle 6.2: Werte im 200ms-Raum bei 0 dB Eingangs-SNR in dB

| | SNR _{gain} | SD | NR |
|----------|---------------------|------|------|
| Zelinski | 2.6 | 1.85 | 4.4 |
| Simmer | 1.1 | 1.8 | 2.95 |
| McCowan | 4.1 | 1.9 | 5.9 |

Wenn man die Ergebnisse mit einem einkanaligen Wienerfilter vergleicht, fällt auf, dass der SNR-Gewinn bei niedrigem Eingangsrauschen wesentlich geringer ist, das McCowan-und Zelinski-Filter jedoch vor allem bei hohen und mittleren SNRs und im hallarmen Raum an die Leistung des Wienerfilters herankommen.

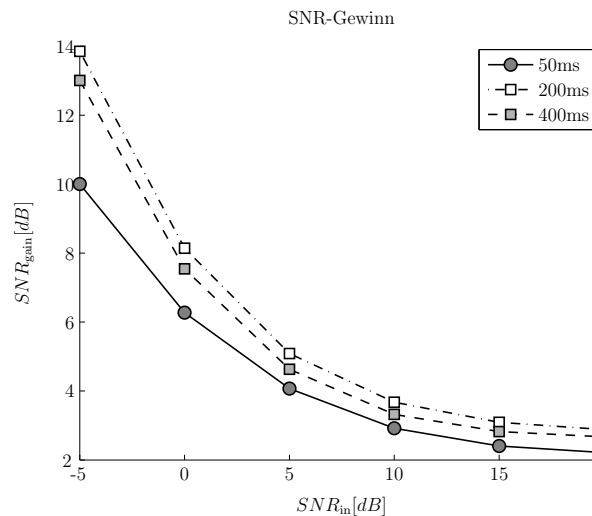


Abbildung 6.9: SNR-Gewinn des Wienerfilters

6.3 Anwendung mit diffusem Rauschen

Im zweiten Versuchsdurchgang wurde als Rauschmodell sechs-kanaliges, diffuses Rauschen verwendet. Somit besaß auch das Hintergrundrauschen räumliche Eigenschaften.

Der weitere Aufbau war identisch zu Kap. 6.2.

6.3.1 Zelinski-Filter

Das Zelinski-Filter erzeugt bei diffusem Rauschen einen wesentlich höheren SNR-Gewinn, während die Sprachverzerrung sich in einem ähnlichen Wertebereich befindet.

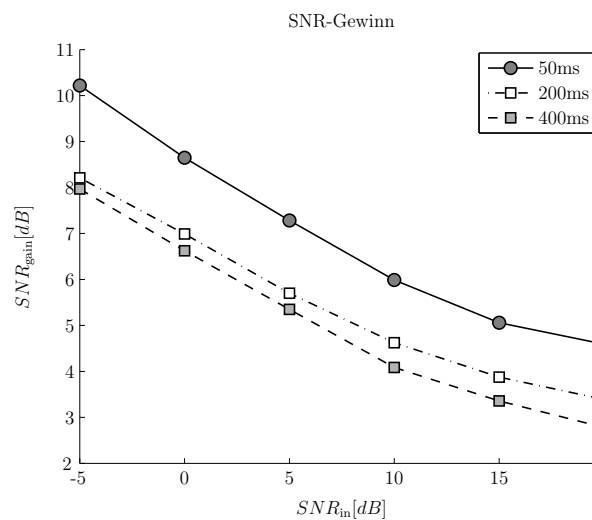


Abbildung 6.10: SNR-Gewinn des Zelinski-Postfilters

Hervorzuheben ist auch die mit wesentlich höhere Geräuschreduktion mit bis zu 7,5 dB Zunahme.

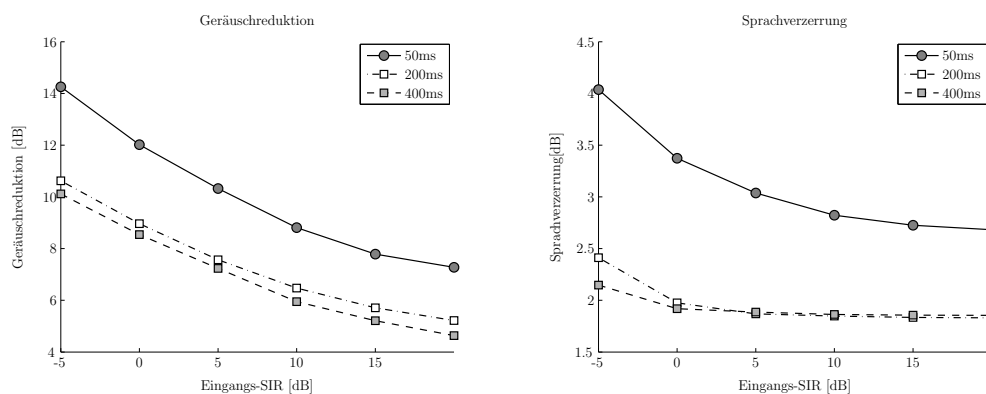


Abbildung 6.11: Geräuschreduktion und Sprachverzerrung des Zelinski-Postfilters

6.3.2 Simmer-Filter

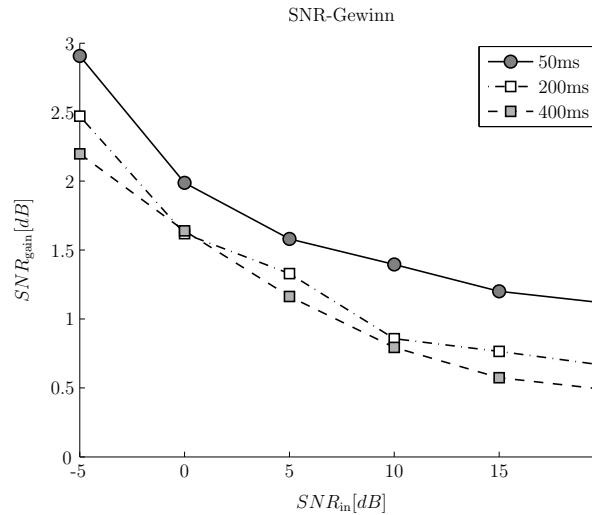


Abbildung 6.12: SNR-Gewinn des Simmer-Postfilters

Bei diffusem Rauschen zeigt das Simmer-Filter einen sehr kleinen SNR-Gewinn von 1,5-3 dB, allerdings wurde auch im hallarmen Raum eine sehr geringe Sprachverzerrung von ca. 2 dB erzielt, während bei den anderen Filtern die Sprachverzerrung um ca. 1,5 dB höher ist.

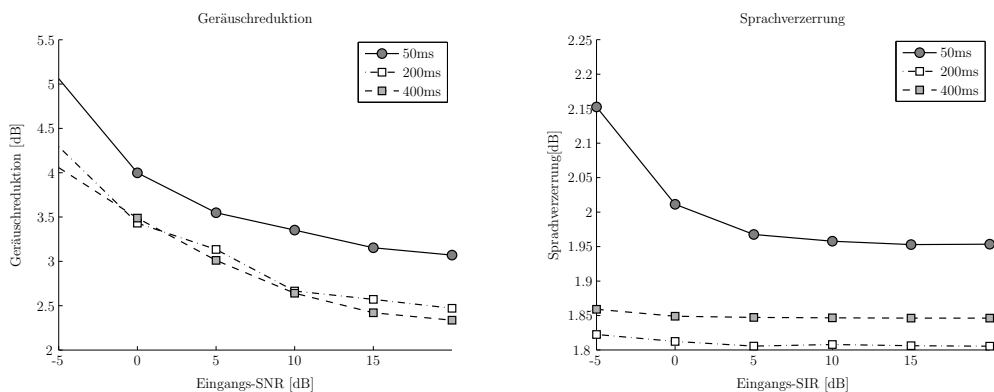


Abbildung 6.13: Geräuschreduktion und Sprachverzerrung des Simmer-Postfilters

6.3.3 Mc Cowan-Filter

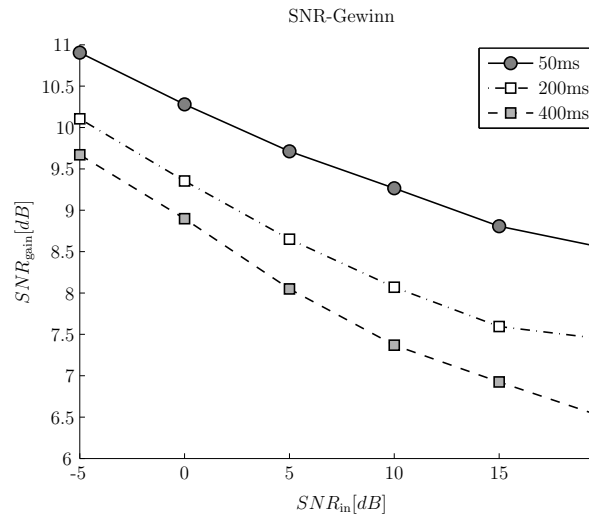


Abbildung 6.14: SNR-Gewinn des McCowan-Postfilters

Da im Fall des diffusen Rauschens die bei der Implementierung des Filters getroffene Annahme eines sphärisch isotropen Rauschfeld (vgl. Kap 2.3.2) für die Eingangssignale erfüllt ist, zeigt das McCowan Filter in diesem Fall eine sehr hohe Geräuschreduktion und einen sehr hohen SNR-Gewinn. Trotzdem liegen vor allem in den Räumen mit längeren Nachhallzeiten eine geringe und fast konstante Sprachverzerrung vor.

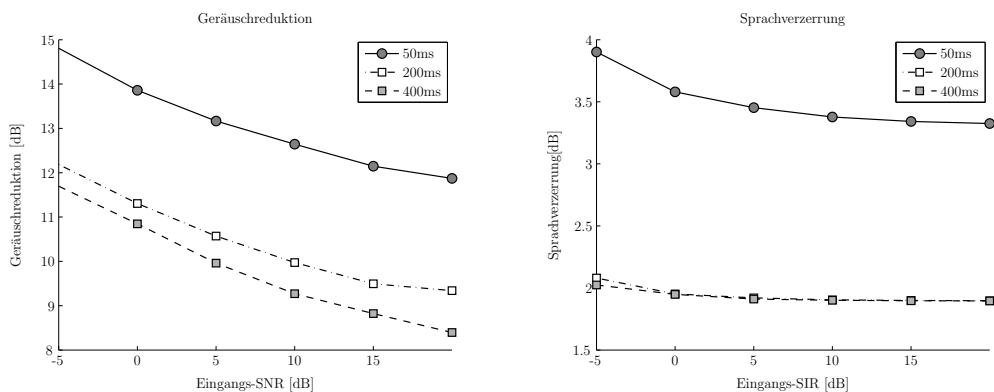


Abbildung 6.15: Geräuschreduktion und Sprachverzerrung des McCowan-Postfilters

6.3.4 Wolff-Filter

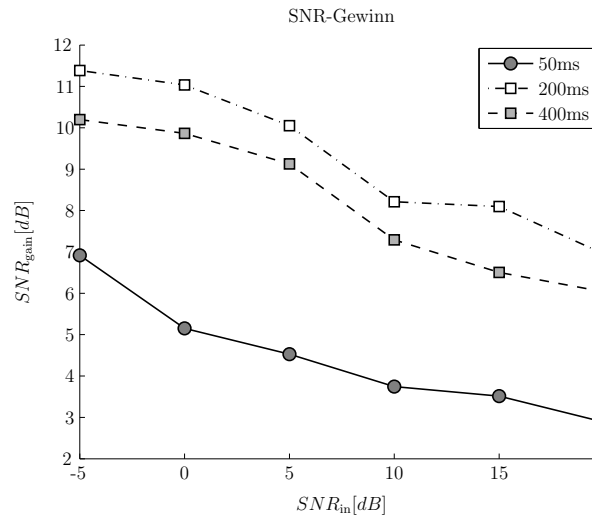


Abbildung 6.16: SNR-Gewinn des Wolff-Postfilters

Das Wolff-Filter zeigt bei diffusem Rauschen genau wie beim weißen Rauschen sehr gute Ergebnisse bei längeren Nachhallzeiten, während unerwarteterweise im hallarmen Raum ein wesentlich kleinerer SNR-Gewinn erzielt wurde.

6.3.5 Vergleich

Auch bei diffusem Rauschen zeigt das McCowan-Filter die besten Ergebnisse mit hohen SNR-Gewinnen und geringer Sprachverzerrung, was auf die passende Annahme für die Rauschkorrelation zurückzuführen ist. Das Simmer-Filter erzeugt im hallarmen Raum eine um ca. 1.5 geringere Sprachverzerrung als die anderen Filter.

Tabelle 6.3: Werte im Hallarmen Raum bei 0 dB Eingangs-SNR in dB

| | SNR_{gain} | SD | NR |
|----------|--------------|-----|------|
| Zelinski | 8.9 | 3.4 | 12.2 |
| Simmer | 2.0 | 2.0 | 4.0 |
| McCowan | 10.4 | 3.6 | 13.2 |

In den anderen Räumen liefert das Simmer-Filter nur eine etwas bessere Sprachverzerrung. Auch hier ist das McCowan-Filter mit nur geringfügig höherer Sprachverzerrung bei der mit über 1,5 dB Abstand höchsten Geräuschreduktion und dem bei 0 dB um 1,8 dB höheren SNR-Gewinn als das Zelinski-Filter

Tabelle 6.4: Werte im 200ms-Raum bei 0 dB Eingangs-SNR in dB

| | SNR _{gain} | SD | NR |
|----------|---------------------|-----|------|
| Zelinski | 7.0 | 2.0 | 8.9 |
| Simmer | 1.7 | 1.8 | 3.5 |
| McCowan | 8.8 | 2.0 | 11.5 |

6.4 Abschließende Analyse

Abschließend gilt zu erwähnen, dass sich vor allem das McCowan-Postfilter mit hohen SNR-Gewinnen und Geräuschreduktionen bei geringer Sprachverzerrung hervorhebt. Vor allem bei bekannten räumlichen Eigenschaften des Rauschens können diese ausgenutzt und zur Geräuschreduktion verwendet werden. Das gilt nicht nur für diffuses Rauschen sondern ist auch für andere Modelle möglich (vgl. 4.8). Das Simmer-Nachfilter stellt einen Minimalanatz dar: Geringe Geräuschreduktion und SNR-Gewinne, dafür jedoch kaum Sprachverzerrung. Das Zelinski-Filter erzeugt eine etwas geringere Sprachverzerrung als McCowans Ansatz, der allerdings wesentlich höhere SNR-Gewinne erzielt. Im hallarmen Raum konnte mit Simmer-Filter eine wesentlich kleinere Sprachverzerrung erzielt werden.

Tabelle 6.5: Durchschnittswerte im hallarmen Raum bei diffusem Eingangsrauschen in dB

| | SNR _{gain} | SD | NR |
|----------|---------------------|------|------|
| Zelinski | 6.95 | 3.13 | 10.4 |
| Simmer | 1.8 | 2.01 | 3.78 |
| McCowan | 9.6 | 3.53 | 13.2 |

Bei allen Filtern weisen der 200ms- und 400ms-Raum sehr ähnliche Werte auf, wobei im 400ms-Raum etwas geringere SNR-Gewinne und Geräuschreduktionen erzeugt wurden. In diesen Umgebungen unterscheiden sich die Sprachverzerrungen kaum, wodurch das Simmer-Filter mit der wesentlich geringeren Geräuschreduktion schlechtere Eigenschaften zeigt als die beiden anderen Ansätze.

Tabelle 6.6: Durchschnittswerte im 200ms-Raum bei diffusem Eingangsruschen in dB

| | SNR _{gain} | SD | NR |
|----------|---------------------|------|-------|
| Zelinski | 5.62 | 2.0 | 7.65 |
| Simmer | 1.32 | 1.81 | 3.20 |
| McCowan | 8.63 | 2.0 | 10.63 |

Da McCowans Ansatz eine verallgemeinerte Version des Zelinski-Filters darstellt, verwundert es nicht, dass das McCowan-Filter im allgemeinen bessere Ergebnisse als das Zelinski-Filter erzielt, vor allem in Umgebungen mit mehr Hall und wenn das Kohärenzmodell zutrifft. Das Wolff-Filter zeigt unerwartete Ergebnisse, was auf Probleme bei der adaptiven Berechnung der Parameter zurückzuführen ist. Wesentlich aussagekräftigere Testergebnisse können jedoch in [12] eingesehen werden.

Kapitel 7

Zusammenfassung und Ausblick

In dieser Arbeit wurde, ausgehend von einer Beschreibung der benötigten Grundlagen der akustischen Signalverarbeitung und Modellbildung, eine zweiteilige Lösung, bestehend aus einem Beamformer und einem Nachfilter, für die Geräuschreduktion aufbauend auf einem mehrkanaligen Wienerfilter hergeleitet. Anschließend wurden verschiedene Möglichkeiten, Postfilter zu schätzen hergeleitet und erklärt. Schließlich wurde noch auf mögliche Bewertungskriterien eingegangen. Eine Implementierung der Filter in Matlab[®] führte zu dem Ergebniss, dass das McCowan-Nachfilter, vor allem im Fall eines bekannten Raummodells für das Rauschen die besten Ergebnisse zeigt und dem Zelinski-Filter vor allem in diesen Fall vorzuziehen ist. Allerdings ist die Schätzung des Zelinski-Filters weniger aufwendig, was es evtl. für Anwendungen mit Ressourcenknappheit attraktiver macht. Das Simmer-Filter stellt einen minimalistischen Ansatz dar und zeigt nur geringe Gewinne allerdings auch nur kleine Verluste an Sprachqualität. Die für das Wolff-Filter getroffenen Annahmen machen die Ergebnisse zu wenig repräsentativ, um eine Aussage zu treffen, allerdings lassen die in [12] erzielten Ergebnisse auf eine hohe Flexibilität im Bezug auf das akustische Szenario schließen. Ein weiterer Schritt wäre der ausführliche Test verschiedener Nachfilter-Beamformer-Kombinationen in einer Vielzahl realistischer akustischer Umgebungen, mit dem Ziel eine Methode zur adaptiven Schätzungen und damit des Filtertyps abhängig vom akustischen Szenarios. Eine erster Ansatz hierzu findet sich in [12]. Tests mit einer größe-

ren Menge von Testhörern zu Bestimmung der subjektiven Qualität wären ein weiterer Schritt, da diese für manche Anwendungen (Entertainment-Systeme, Mobilkommunikation usw.) relevanter als die rein objektive Bewertung ist.

Abbildungsverzeichnis

| | | |
|------|--|----|
| 2.1 | Blockdiagramm für ein Mikrophonsignal | 3 |
| 2.2 | Laufzeitunterschiede zwischen den Mikrofonen am Beispiel eine vierkanaligen Mikrofonarrays | 4 |
| 2.3 | Kohärenzfunktion eines Büroraums im Vergleich zum Modell [8]. | 7 |
| 3.1 | Andwendungsbeispiel eines Mikrophonrrays | 9 |
| 3.2 | Blockdiagramm eines Beamformers | 12 |
| 3.3 | GSC-Struktur | 14 |
| 6.1 | Modellaufbau | 23 |
| 6.2 | SNR-Gewinn des Zelinski-Postfilters | 25 |
| 6.3 | Geräuschreduktion und Sprachverzerrung des Zelinski-Postfilters | 26 |
| 6.4 | SNR-Gewinn des Simmer-Postfilters | 26 |
| 6.5 | Geräuschreduktion und Sprachverzerrung des Simmer-Postfilters | 27 |
| 6.6 | SNR-Gewinn des McCowan-Postfilters | 27 |
| 6.7 | Geräuschreduktion und Sprachverzerrung des McCowan-Postfilters | 28 |
| 6.8 | SNR-Gewinn des Wolff-Postfilters | 29 |
| 6.9 | SNR-Gewinn des Wienerfilters | 30 |
| 6.10 | SNR-Gewinn des Zelinski-Postfilters | 31 |
| 6.11 | Geräuschreduktion und Sprachverzerrung des Zelinski-Postfilters | 31 |
| 6.12 | SNR-Gewinn des Simmer-Postfilters | 32 |
| 6.13 | Geräuschreduktion und Sprachverzerrung des Simmer-Postfilters | 32 |

| | |
|---|----|
| 6.14 SNR-Gewinn des McCowan-Postfilters | 33 |
| 6.15 Geräuschreduktion und Sprachverzerrung des McCowan-Postfilters . . . | 33 |
| 6.16 SNR-Gewinn des Wolff-Postfilters | 34 |

Tabellenverzeichnis

| | | |
|-----|---|----|
| 6.1 | Werte im Hallarmen Raum bei 0 dB Eingangs-SNR in dB | 29 |
| 6.2 | Werte im 200ms-Raum bei 0 dB Eingangs-SNR in dB | 30 |
| 6.3 | Werte im Hallarmen Raum bei 0 dB Eingangs-SNR in dB | 34 |
| 6.4 | Werte im 200ms-Raum bei 0 dB Eingangs-SNR in dB | 35 |
| 6.5 | Durchschnittswerte im hallarmen Raum bei diffusem Eingangsrauschen in dB | 35 |
| 6.6 | Durchschnittswerte im 200ms-Raum bei diffusem Eingangsrauschen in dB | 36 |

Literaturverzeichnis

- [1] H. Buchner. *Breitbandige adaptive MIMO-Filterung: Eine vereinheitlichte Behandlung und Anwendungen auf akustische Mensch-Maschine-Schnittstellen*. June 2010.
- [2] P. Fertl. *Mikrofonarray mit adaptivem Postfilter zur Sprachsignalentstörung*. 2011.
- [3] O. L. Frost. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE, vol. 60, NO. 8*, 1972.
- [4] L. J. Griffiths and C. W. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. AP-30, NO. 1*, 1982.
- [5] W. Kellermann. *Digital Signal Processing, Lecture Notes Summer 2013*. 2013.
- [6] W. Kellermann. *Statistical Signal Processing, Lecture Notes Summer 2013*. 2013.
- [7] B.-U. Köhler. *Konzepte der statistischen Signalverarbeitung*. 2005.
- [8] I. A. McCowan and H. Bourslard. Microphone array postfilter based on noise field coherence. *IEEE Transactions on Speech and Audio Processing, Vol. 11, No.6*, 2003.
- [9] K. Reindl, S. Meier, H. Barfuss, and W. Kellermann. Minimum mutual information-based linearly constrained broadband signal extraction. *unreleased?*, 2014.

-
- [10] K. Simmer and A. Wasiljeff. Adaptive microphone arrays for noise suppression in the frequency domain. *Second Cost 229 Workshop on Adaptive Algorithms in Communications*, 1992.
- [11] P. Vary and R. Martin. *Digital Speech Transmission*. 2006.
- [12] C. Wolff. *Entwurf und Analyse von Beamformer-Nachfilter-Systemen*. 2011.
- [13] R. Zelinski. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. *Proc. ICASSP-88, vol. 5*, 1988.