

UNIVERSITÄT ERLANGEN-NÜRNBERG

**Lehrstuhl für Multimediakommunikation und
Signalverarbeitung**

PROF. DR.-ING. WALTER KELLERMANN

Bachelorarbeit

**Einkanalige Störreduktionsverfahren zur
Sprachsignalverbesserung in binauralen
Hörgeräten**

von Andreas Heindel

September 2010

Betreuer: Dipl.-Ing. (FH) Klaus Reindl



Bachelorarbeit

für

Herrn cand. ing. Andreas Heindel

Einkanalige Störreduktionsverfahren zur Sprachsignalverbesserung in binauralen Hörgeräten

Für Personen mit Gehörschädigungen ist es insbesondere schwierig, einen Gesprächspartner in akustischen Umgebungen mit hohen Störsignalpegeln zu verstehen. Aus diesem Grund benötigen diese Personen Hörhilfen, die nicht nur die Störsignale in den aufgenommenen Mikrophonsignalen reduzieren, sondern gleichzeitig auch die Sprachverständlichkeit verbessern. Bis vor Kurzem wurden nur monaurale oder bilaterale Algorithmen zur Sprachsignalverbesserung betrachtet, d.h., die Signale wurden unabhängig für jedes Ohr verarbeitet. Diese Vorgehensweise entspricht jedoch nicht dem menschlichen binauralen Hören und deshalb bleiben bei solchen Verarbeitungsstrategien wichtige binaurale Informationen ungenutzt. Deshalb sind aktuell binaurale Verarbeitungsstrategien für die Sprachsignalverbesserung in miteinander kommunizierenden Hörgeräten von großer Bedeutung, die z.B. durch die drahtlose Übertragung ermöglicht werden.

Eine seit langem bekannte Methode zur einkanaligen Sprachsignalverbesserung ist das Wiener-Filter. Neben diesem Ansatz gibt es aber auch zahlreiche andere Verfahren, wie etwa MMSE (minimum mean square error)-, MAP (maximum a posteriori)- und ML (maximum likelihood)-Schätzverfahren, die mit Hilfe der Bayes'schen Schätztheorie abgeleitet werden können.

Die Aufgabe dieser Bachelorarbeit besteht darin, die verschiedenen Schätzmethoden speziell für die Sprachsignalverbesserung in binauralen Hörgeräten zu analysieren. Aus diesem Grund sollen die verschiedenen Methoden hinsichtlich Störunterdrückung, Nutzsignalverzerrung und spektraler Verzerrungen verglichen werden. Der Einfluss von nichtstationären Störsignalen steht dabei im Vordergrund. Hierfür sind eine theoretische sowie eine experimentelle Analyse durch Simulationen mit Matlab erforderlich. Großer Wert wird auf eine gut dokumentierte und gut strukturierte Implementierung gelegt.

Beginn: 19.04.2010
Ende: 19.09.2010

A handwritten signature in black ink, appearing to read 'W. Kellermann'.

(Prof. Dr.-Ing. W. Kellermann)

Erklärung

Ich versichere, dass ich die 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.

Aurachtal, den 17.09.2010

Andreas Heindel

Nelkenstraße 1

91086 Aurachtal

Zusammenfassung

In der vorliegenden Bachelorarbeit wird die Leistungsfähigkeit verschiedener Verfahren zur binauralen Geräuschreduktion in Hörgeräten untersucht. Dazu werden aus der Bayes'schen Schätztheorie verschiedene Schätzer hergeleitet, mit denen sich das Sprachsignal eines gewünschten Sprechers aus den beiden mit Störungen verschiedener Art überlagerten Mikrofonsignalen schätzen lässt. Dazu wird eine gemeinsame Störschätzung und, da die Geräuschreduktion im Frequenzbereich stattfindet, ein statistisches Modell für die DFT-Koeffizienten des Sprachsignals benötigt. Die Evaluierung der verschiedenen entwickelten Schätzverfahren geschieht im Rahmen von Matlabsimulationen, bei denen die Algorithmen in verschiedenen realistischen Szenarien getestet werden. Die Leistungsfähigkeit der entwickelten Schätzverfahren wird schließlich mit dem bisher verwendeten und einem abgeänderten Wiener-Filter Ansatz verglichen und es wird die für Hörgeräteanwendungen am besten geeignete Methode herausgestellt.

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen der Bayes'schen Schätzung	3
2.1	MAP-Schätzung	5
2.2	ML-Schätzung	7
2.3	MMSE-Schätzung	7
3	Bayes'sche Schätzung zur Störreduktion	11
3.1	Zugrundeliegendes Modell	11
3.2	MAP-Schätzung	15
3.3	ML-Schätzung	15
3.4	MMSE-Schätzung	17
4	Evaluierung der verschiedenen Methoden durch Simulation	21
5	Zusammenfassung und Ausblick	41
A	Simulationsergebnisse für Filterlängen 256 und 128	43
A.1	Ergebnisse für Filterlänge 256	43
A.2	Ergebnisse für Filterlänge 128	48
B	Akronyme, Notation und Symbole	53
	Literaturverzeichnis	59

Abbildungsverzeichnis

2.1	Kostenfunktion der MAP-Schätzung	6
2.2	Kostenfunktion der MMSE-Schätzung	8
2.3	Beispielhafte Resultate der Bayes'schen Schätzung eines Skalars für eine symmetrische und eine asymmetrische Wahrscheinlichkeitsdichte	9
3.1	Modell für die additive Überlagerung von Quellensignalen und an- schließender Störreduktion	12
3.2	Verbunddichte des Real- und Imaginärteils von S_μ	14
4.1	Blockschaltbild der binauralen Geräuschreduktion nach [8]	22
4.2	Prinzip der spektralen Analyse durch Verwendung der DFT [7]	24
4.3	Realisierung der Polyphasenfilterbank [4]	24
4.4	Rekursives System zur Schätzung des Leistungsdichtespektrums	26
4.5	Untersuchte Szenarien; der gewünschte Sprecher befindet sich im- mer ungefähr bei 0° vor dem Hörgerätbenutzer	28
4.6	$L = 512$, 1 Störquelle, getrennter Filteransatz	36
4.7	$L = 512$, 1 Störquelle, gemeinsamer Filteransatz	36
4.8	$L = 512$, 2 Störquellen, getrennter Filteransatz	37
4.9	$L = 512$, 2 Störquellen, gemeinsamer Filteransatz	37
4.10	$L = 512$, 3 Störquellen, getrennter Filteransatz	38
4.11	$L = 512$, 3 Störquellen, gemeinsamer Filteransatz	38

4.12	$L = 512$, 3 Störquellen und Hintergrundrauschen, getrennter Filteransatz	39
4.13	$L = 512$, 3 Störquellen und Hintergrundrauschen, gemeinsamer Filteransatz	39
A.1	$L = 256$, 1 Störquelle, getrennter Filteransatz	44
A.2	$L = 256$, 1 Störquelle, gemeinsamer Filteransatz	44
A.3	$L = 256$, 2 Störquellen, getrennter Filteransatz	45
A.4	$L = 256$, 2 Störquellen, gemeinsamer Filteransatz	45
A.5	$L = 256$, 3 Störquellen, getrennter Filteransatz	46
A.6	$L = 256$, 3 Störquellen, gemeinsamer Filteransatz	46
A.7	$L = 256$, 3 Störquellen und Hintergrundrauschen, getrennter Filteransatz	47
A.8	$L = 256$, 3 Störquellen und Hintergrundrauschen, gemeinsamer Filteransatz	47
A.9	$L = 128$, 1 Störquelle, getrennter Filteransatz	48
A.10	$L = 128$, 1 Störquelle, gemeinsamer Filteransatz	49
A.11	$L = 128$, 2 Störquellen, getrennter Filteransatz	49
A.12	$L = 128$, 2 Störquellen, gemeinsamer Filteransatz	50
A.13	$L = 128$, 3 Störquellen, getrennter Filteransatz	50
A.14	$L = 128$, 3 Störquellen, gemeinsamer Filteransatz	51
A.15	$L = 128$, 3 Störquellen und Hintergrundrauschen, getrennter Filteransatz	51
A.16	$L = 128$, 3 Störquellen und Hintergrundrauschen, gemeinsamer Filteransatz	52

Tabellenverzeichnis

3.1	Schätzergebnisse der verschiedenen Schätzer	20
4.1	Berechnungsvorschriften für den getrennten Filteransatz	32
4.2	Berechnungsvorschriften für den gemeinsamen Filteransatz	32
4.3	Verwendete Parameter bei einer Filterlänge von $L = 512$	35
A.1	Verwendete Parameter bei einer Filterlänge von $L = 256$	43
A.2	Verwendete Parameter bei einer Filterlänge von $L = 128$	48

Kapitel 1

Einleitung

Ein großes Problem hörgeschädigter Menschen liegt darin, dass sie im Gespräch mit anderen Personen diese nur schwer verstehen können. In lauten Umgebungen verstärkt sich dieses Problem und das Verstehen anderer Personen wird je nach Lautstärke der Störgeräusche schnell unmöglich. Deshalb sind sie auf Hörgeräte angewiesen, die diese Umgebungsgeräusche reduzieren. Bisher sind zu diesem Zweck nur bilaterale Geräuschreduktionstechniken in Hörgeräten verfügbar. Das bedeutet, dass für jedes Ohr eine getrennte, unabhängige Geräuschreduktion durchgeführt wird. Durch diese unabhängige Verarbeitung entsteht allerdings der große Nachteil, dass es für hörgeschädigte Personen sehr schwierig werden kann, den Ort eines Sprechers oder einer Geräuschquelle zu erkennen, da die hierfür benötigten interauralen Zeit- und Pegelunterschiede bei der Verarbeitung verloren gehen können. Auch die Sprachverständlichkeit leidet unter dieser für jedes Ohr unabhängigen Verarbeitung. Die drahtlose Kommunikationstechnologie ermöglicht nun den Austausch der Mikrofonsignale, so dass für jedes Hörgerät alle Mikrofonsignale zu Verfügung stehen. Dadurch sind nun ganz neue Strategien zur binauralen Geräuschreduktion möglich. Ziel dieses neuen Konzeptes ist es, die Mikrofonsignale in der Form zu verarbeiten, dass die zur räumlichen Lokalisation von Quellen wichtigen interauralen Zeit- und Pegelunterschiede erhalten bleiben und dadurch auch die Sprachverständlichkeit verbessert wird.

In [6] wurde ein möglicher Ansatz zur Störreduktion vorgestellt. Dabei wurden Hörgeräte mit einem Mikrofon je Hörgerät untersucht. Außerdem wurde zur Störreduktion ein Wiener-Filter verwendet. Mit Hilfe der Bayes'schen Schätztheorie lassen sich allerdings noch einige andere Schätzer zur Geräuschunterdrückung herleiten. Ziel dieser Arbeit ist es, verschiedene *maximum likelihood* (ML)-, *maximum a-posteriori* (MAP)- und *minimum mean squared error* (MMSE)-Schätzer mit dem bisher verwendeten Wiener-Filter zu vergleichen. Und dies nicht nur im Sinne der besten Störreduktion, sondern auch in Bezug auf die Sprachverständlichkeit des geräuschreduzierten Mikrofonsignals.

Um die Theorie hinter den zu vergleichenden Schätzern zu erläutern, beschäftigt sich Kapitel 2 dieser Arbeit mit den Grundlagen der Bayes'schen Schätztheorie und geht insbesondere auf die Herleitung der erwähnten ML-, MAP- und MMSE-Schätzer ein. In Kapitel 3 werden zunächst einige Annahmen über das zugrundeliegende Modell bei der Geräuschreduktion bei Sprachsignalen im Frequenzbereich gemacht. Anschließend werden die im Kapitel 2 aus der Bayes'schen Schätztheorie hergeleiteten Schätzer auf das Modell bezogen und die daraus resultierenden Schätzer abgeleitet. Kapitel 4 befasst sich schließlich mit der Evaluierung der verschiedenen Schätzer durch Matlabsimulationen. Dabei wird zunächst auf den Aufbau des Programms zur Simulation eingegangen, bevor die praktische Leistungsfähigkeit der verschiedenen Schätzer näher untersucht wird. Schließlich endet die Arbeit in Kapitel 5 mit einer Zusammenfassung der gewonnenen theoretischen und praktischen Ergebnisse und einem kurzen Ausblick.

Kapitel 2

Grundlagen der Bayes'schen Schätzung

Die Schätztheorie im Allgemeinen ist ein sehr mächtiges Werkzeug. Nach [5] dient sie beispielsweise zur Schätzung von Parametern von Verteilungsdichtefunktionen beziehungsweise statistischen Modellen oder zur Schätzung der Koeffizienten prädiktiver Modelle. In dieser Arbeit wird die Schätztheorie allerdings zur Schätzung von Signalen benutzt, die nicht direkt beobachtbar sind. Nämlich zur Schätzung des Sprachsignals des gewünschten Sprechers aus den mit Störgeräuschen überlagerten Mikrofonsignalen der Hörgeräteträgerin oder des Hörgeräteträgers. Basierend auf [5] und [3] werden in diesem Kapitel einige Grundlagen der Schätztheorie nach Bayes¹ erläutert.

Zuerst sind allerdings noch zwei Konventionen zur Notation notwendig: Im Folgenden werden zufällige Größen mit Großbuchstaben und determinierte Größen mit Kleinbuchstaben bezeichnet. Desweiteren werden vektorielle Größen fettgedruckt dargestellt, skalare Größen jedoch nicht.

In der Schätztheorie unterscheidet man nach [3] zwischen dem *a-priori* Ergebnisraum \mathcal{V} , der alle Werte enthält, die das zu schätzende Signal überhaupt

¹Thomas Bayes: englischer Mathematiker (~1702 - 1761)

annehmen kann, und dem *a-posteriori* Ergebnisraum \mathcal{X} , der nach der Beobachtung aus allen wahrscheinlichen Werten des Signals besteht, die konsistent zu einer Beobachtung und dem a-priori Ergebnisraum sind.

Die Verwendung der Bayes'schen Schätzung bietet gegenüber der *klassischen* Schätzung nach [3] den großen Vorteil auch statistisches Vorwissen über das zu schätzende Signal mit einbringen zu können. Dies geschieht durch die Ausnutzung der sogenannten Bayes'schen Regel, die einen Zusammenhang zwischen der a-priori Wahrscheinlichkeitsdichte $f_{\Phi}(\phi)$ eines Parametervektors ϕ und der a-posteriori Wahrscheinlichkeitsdichte $f_{\Phi|\mathbf{x}}(\phi|\mathbf{x})$ eines Parametervektors ϕ mit gegebenem Beobachtungsvektor \mathbf{x} herstellt:

$$f_{\Phi|\mathbf{x}}(\phi|\mathbf{x}) = \frac{f_{\mathbf{x}|\Phi}(\mathbf{x}|\phi) \cdot f_{\Phi}(\phi)}{f_{\mathbf{x}}(\mathbf{x})} \quad (2.1)$$

Für eine gegebene Beobachtung \mathbf{x} ist die Wahrscheinlichkeitsdichte $f_{\mathbf{x}}(\mathbf{x})$ im Nenner eine einfache Konstante und dient damit nur zur Normierung. Die a-posteriori Wahrscheinlichkeitsdichte ist also proportional zum Produkt aus der sogenannten Likelihood-Funktion $f_{\mathbf{x}|\Phi}(\mathbf{x}|\phi)$ und der a-priori Wahrscheinlichkeitsdichte $f_{\Phi}(\phi)$. Die a-priori Wahrscheinlichkeitsdichte gewichtet also die Likelihood-Funktion mit der ursprünglichen Wahrscheinlichkeitsdichte für das Auftreten des Parameters ϕ und bringt damit nach [5] das Vorwissen über den a-priori Raum \mathcal{V} mit ein.

Ziel der Bayes'schen Schätzung ist die Minimierung einer Risikofunktion $\mathcal{R}(\hat{\Phi})$, die unter anderem laut [3] als Erwartungswert einer Kostenfunktion $\mathcal{C}(\hat{\Phi}, \Phi)$ definiert ist:

$$\begin{aligned} \mathcal{R}(\hat{\Phi}) &= \mathcal{E}\{\mathcal{C}(\hat{\Phi}, \Phi)\} \\ &= \int_{\mathcal{V}} \int_{\mathcal{X}} \mathcal{C}(\hat{\Phi}, \phi) f_{\mathbf{x}, \Phi}(\mathbf{x}, \phi) \, d\mathbf{x} \, d\phi \\ &= \int_{\mathcal{V}} \int_{\mathcal{X}} \mathcal{C}(\hat{\Phi}, \phi) f_{\Phi|\mathbf{x}}(\phi|\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) \, d\mathbf{x} \, d\phi \end{aligned} \quad (2.2)$$

Für eine gegebene Beobachtung \mathbf{x} ist die Wahrscheinlichkeitsdichte $f_{\mathbf{x}}(\mathbf{x})$ wieder eine Konstante. Da es bei der Bayes'schen Schätzung um die Minimierung der Risikofunktion $\mathcal{R}(\hat{\Phi})$ geht, spielt $f_{\mathbf{x}}(\mathbf{x})$ hier keine Rolle und 2.2 kann somit auf

die sogenannte bedingte Risikofunktion reduziert werden:

$$\mathcal{R}(\hat{\phi}|\mathbf{x}) = \int_{\mathcal{V}} \mathcal{C}(\hat{\phi}, \phi) f_{\Phi|\mathbf{X}}(\phi|\mathbf{x}) d\phi \quad (2.3)$$

Die Kostenfunktion $\mathcal{C}(\hat{\phi}, \phi)$ ist dabei ein Maß für die Kosten des Fehlers zwischen dem geschätzten Parametervektor $\hat{\phi}$ und dem tatsächlichen Parametervektor ϕ und ist dementsprechend sinnvoll zu wählen.

Für einen gegebenen Beobachtungsvektor \mathbf{x} ist auch der Parametervektor ϕ determiniert und dessen Bayes'sche Schätzung ist schließlich definiert als derjenige Parametervektor, welcher die Risikofunktion minimiert:

$$\hat{\phi}_{\text{Bayes}} = \underset{\hat{\phi}}{\operatorname{argmin}} \mathcal{R}(\hat{\phi}|\mathbf{x}) = \underset{\hat{\phi}}{\operatorname{argmin}} \left(\int_{\mathcal{V}} \mathcal{C}(\hat{\phi}, \phi) f_{\Phi|\mathbf{X}}(\phi|\mathbf{x}) d\phi \right) \quad (2.4)$$

Unter Verwendung der Bayes'schen Regel lässt sich dies auch umformen zu

$$\hat{\phi}_{\text{Bayes}} = \underset{\hat{\phi}}{\operatorname{argmin}} \left(\int_{\mathcal{V}} \mathcal{C}(\hat{\phi}, \phi) f_{\mathbf{X}|\Phi}(\mathbf{x}|\phi) f_{\Phi}(\phi) d\phi \right). \quad (2.5)$$

Die Kostenfunktion $\mathcal{C}(\hat{\phi}, \phi)$ blieb dabei bisher unbestimmt. Durch die nun folgende Spezifizierung dieser Kostenfunktion wird der allgemeine Ansatz verlassen und es werden konkrete Schätzer nach den Kriterien *maximum a-posteriori*, *maximum likelihood* und *minimum mean squared error* eingeführt.

2.1 MAP-Schätzung

Der erste konkrete Schätzer, der hier betrachtet wird, ist der sogenannte maximum a-posteriori (MAP) Schätzer. MAP-Schätzer liefern denjenigen Schätzvektor $\hat{\phi}$, der die a-posteriori Wahrscheinlichkeitsdichte $f_{\Phi|\mathbf{X}}(\phi|\mathbf{x})$ maximiert. Als Kostenfunktion wird hierfür die sogenannte gleichförmige Kostenfunktion

$$\mathcal{C}_{\text{MAP}}(\hat{\Phi}, \Phi) = 1 - \delta(\|\hat{\Phi} - \Phi\|) \quad (2.6)$$

verwendet. Diese Kostenfunktion ist in Abbildung 2.1 für den Fall der Schätzung eines skalaren Parameters ϕ dargestellt. Die Kosten sind nur dann Null, wenn der

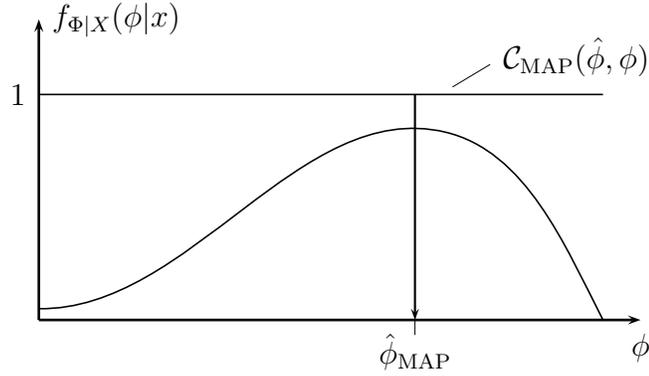


Abbildung 2.1: Kostenfunktion der MAP-Schätzung

geschätzte Parameter $\hat{\phi}$ genau dem tatsächlichen Parameter entspricht. Anderenfalls sind die Kosten immer maximal (eins). Bezogen auf Vektoren bedeutet dies analog, dass die Kosten nur dann Null sind, wenn der geschätzte Parametervektor $\hat{\phi}$ und der tatsächliche Parametervektor ϕ übereinstimmen.

Setzt man (2.6) in (2.3) ein, ergibt sich damit

$$\begin{aligned} \mathcal{R}_{\text{MAP}}(\hat{\phi}|\mathbf{x}) &= \int_{\mathcal{V}} [1 - \delta(\|\hat{\phi} - \phi\|)] f_{\Phi|\mathbf{X}}(\phi|\mathbf{x}) d\phi \\ &= 1 - f_{\Phi|\mathbf{X}}(\hat{\phi}|\mathbf{x}). \end{aligned} \quad (2.7)$$

Die MAP-Schätzung erhält man nun durch die Minimierung der Risikofunktion (2.7). Diese wird genau dann minimal, wenn die a-posteriori Wahrscheinlichkeitsdichte $f_{\Phi|\mathbf{X}}(\hat{\phi}|\mathbf{x})$ maximal wird. Der Schätzvektor $\hat{\phi}$ lässt sich somit wie folgt berechnen:

$$\begin{aligned} \hat{\phi}_{\text{MAP}} &= \underset{\hat{\phi}}{\operatorname{argmax}} f_{\Phi|\mathbf{X}}(\hat{\phi}|\mathbf{x}) \\ &= \underset{\hat{\phi}}{\operatorname{argmax}} [f_{\mathbf{X}|\Phi}(\mathbf{x}|\hat{\phi}) f_{\Phi}(\hat{\phi})] \end{aligned} \quad (2.8)$$

Da die a-posteriori Wahrscheinlichkeitsdichte $f_{\Phi|\mathbf{X}}(\hat{\phi}|\mathbf{x})$ an der Stelle des geschätzten Vektors $\hat{\phi}$ maximal wird, sind genau an dieser Stelle die Kosten Null, was auch in Abbildung 2.1 dargestellt ist. Die Likelihood-Funktion $f_{\mathbf{X}|\Phi}(\mathbf{x}|\phi)$, welche den Übergang von den Parametern Φ zu den Beobachtungen \mathbf{X} beschreibt, kann hierbei durch Trainingsdaten gemessen werden.

2.2 ML-Schätzung

Die maximum likelihood (ML) Schätzung ist genau genommen nur ein Spezialfall der maximum a-posteriori Schätzung. Die Kostenfunktion ist wie bei der MAP-Schätzung

$$\mathcal{C}_{\text{ML}}(\hat{\Phi}, \Phi) = 1 - \delta(\|\hat{\Phi} - \Phi\|). \quad (2.9)$$

Dementsprechend stimmen auch die Risikofunktion (siehe (2.7)) und der Ausdruck für den geschätzten Parametervektor überein:

$$\begin{aligned} \hat{\phi}_{\text{ML}} &= \operatorname{argmax}_{\hat{\phi}} f_{\Phi|\mathbf{X}}(\hat{\phi}|\mathbf{x}) \\ &= \operatorname{argmax}_{\hat{\phi}} \left[f_{\mathbf{X}|\Phi}(\mathbf{x}|\hat{\phi}) f_{\Phi}(\hat{\phi}) \right] \end{aligned} \quad (2.10)$$

Der einzige Unterschied zur MAP-Schätzung besteht in der Annahme einer anderen, genauer genommen einer speziellen a-priori Wahrscheinlichkeitsdichte $f_{\Phi}(\phi)$. Diese wird bei der ML-Schätzung als konstant angenommen, das heißt alle Elemente des a-priori Ergebnisraums \mathcal{V} sind gleich wahrscheinlich. Dies bedeutet aber auch, dass bei der Schätzung kein Vorwissen mehr über den a-priori Ergebnisraum ausgenutzt wird. (2.10) vereinfacht sich damit zu

$$\hat{\phi}_{\text{ML}} = \operatorname{argmax}_{\hat{\phi}} f_{\mathbf{X}|\Phi}(\mathbf{x}|\hat{\phi}). \quad (2.11)$$

2.3 MMSE-Schätzung

Die Kostenfunktion für die minimum mean squared error (MMSE) Schätzung ist das Quadrat des Fehlers des Schätzers, also das Quadrat der Abweichung des geschätzten Parametervektors vom wahren Parametervektor (siehe auch Abbildung 2.2 zur Veranschaulichung mit einem Skalar),

$$\mathcal{C}_{\text{MSE}}(\hat{\Phi}, \Phi) = \|\hat{\Phi} - \Phi\|^2. \quad (2.12)$$

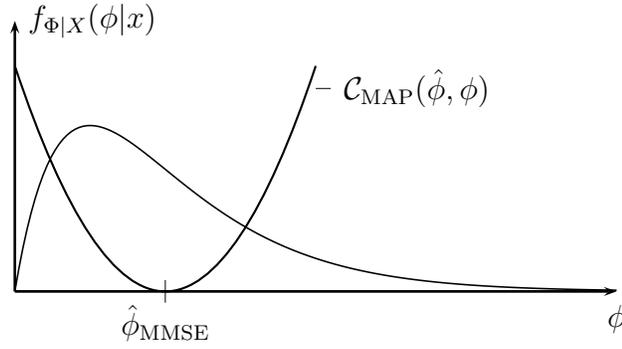


Abbildung 2.2: Kostenfunktion der MMSE-Schätzung

Eingesetzt in (2.3) für die bedingte Risikofunktion ergibt sich für eine gegebene Beobachtung \mathbf{x} nach [3]:

$$\begin{aligned} \mathcal{R}_{\text{MSE}}(\hat{\phi}|\mathbf{x}) &= \int_{\mathcal{V}} \|\hat{\phi} - \phi\|^2 f_{\Phi|\mathbf{X}}(\phi|\mathbf{x}) \, d\phi \\ &= \mathcal{E} \left\{ \|\hat{\phi} - \Phi\|^2 | \mathbf{x} \right\} \end{aligned} \quad (2.13)$$

Unter der Annahme, dass diese bedingte Risikofunktion differenzierbar ist und ein Minimum besitzt, lässt sich dieses dadurch ermitteln, indem man die Ableitung bezüglich $\hat{\phi}$ bildet und diese zu Null setzt:

$$\begin{aligned} \frac{\partial}{\partial \hat{\phi}} \mathcal{E} \left\{ \|\hat{\phi} - \Phi\|^2 | \mathbf{x} \right\} &= \frac{\partial}{\partial \hat{\phi}} \mathcal{E} \left\{ (\hat{\phi} - \Phi)^T (\hat{\phi} - \Phi) | \mathbf{x} \right\} \\ &= \frac{\partial}{\partial \hat{\phi}} \mathcal{E} \left\{ \hat{\phi}^T \hat{\phi} - 2\hat{\phi}^T \Phi + \Phi^T \Phi | \mathbf{x} \right\} \\ &= 2\hat{\phi} - 2\mathcal{E} \{ \Phi | \mathbf{x} \} \stackrel{!}{=} 0 \end{aligned} \quad (2.14)$$

Damit erhält man den bedingten Erwartungswert als Schätzvektor im Sinne des MMSE:

$$\hat{\phi}_{\text{MMSE}} = \mathcal{E} \{ \Phi | \mathbf{x} \} \quad (2.15)$$

Abbildung 2.3 gibt einen Überblick über die Resultate der behandelten Schätzer. Zur Vollständigkeit ist hier auch der Schätzwert bei Verwendung eines *minimum mean absolute value of error* (MAVE)-Schätzers eingetragen. Den MAVE-Schätzwert erhält man nach [5] durch Verwendung der Kostenfunktion

$$\mathcal{C}_{\text{MAVE}}(\hat{\Phi}, \Phi) = \|\hat{\Phi} - \Phi\|. \quad (2.16)$$

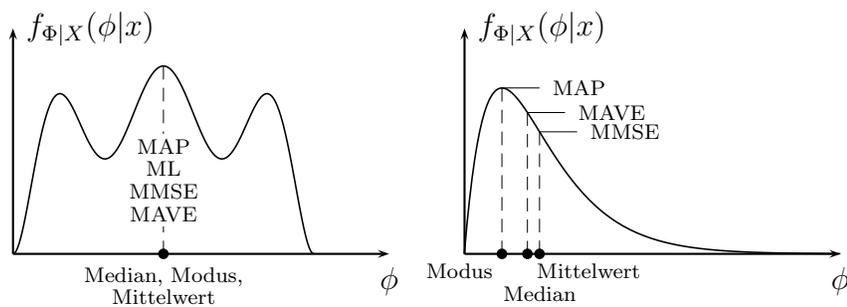


Abbildung 2.3: Beispielhafte Resultate der Bayes'schen Schätzung eines Skalars für eine symmetrische und eine asymmetrische Wahrscheinlichkeitsdichte

Für eine gegebene Beobachtung erhält man nach Einsetzen in (2.3) für die bedingte Risikofunktion und Minimierung laut [5] als Bedingung für den Schätzvektor $\hat{\phi}$

$$\int_{-\infty}^{\hat{\phi}(\mathbf{x})} f_{\Phi|\mathbf{X}}(\phi|\mathbf{x}) d\phi = \int_{\hat{\phi}(\mathbf{x})}^{\infty} f_{\Phi|\mathbf{X}}(\phi|\mathbf{x}) d\phi. \quad (2.17)$$

Nach der MAVE-Schätzung ergibt sich also derjenige Wert ϕ als Schätzwert $\hat{\phi}$, der den Median der a-posteriori Wahrscheinlichkeitsdichte $f_{\Phi|\mathbf{X}}(\phi|\mathbf{x})$ bildet, während man durch den MAP-Schätzer den sogenannten *Modus* und durch den MMSE-Schätzer den Mittelwert erhält.

Kapitel 3

Bayes'sche Schätzung zur Störreduktion

In diesem Kapitel werden die in Kapitel 2 hergeleiteten Schätzer dazu verwendet, das Nutzsignal eines mit Störung überlagerten Mikrofonsignals zu bestimmen. Dazu werden in Abschnitt 3.1 zuerst einige Annahmen getroffen, bevor im restlichen Teil des Kapitels das Nutzsignal mit den verschiedenen in Kapitel 2 vorgestellten Schätzern geschätzt wird.

3.1 Zugrundeliegendes Modell

Wie in der Einleitung beschrieben, sollen die Verfahren der Bayes'schen Schätzung aus Kapitel 2 dieser Arbeit zur Geräuschreduktion in Mikrofonsignalen von Hörgeräten verwendet werden. Abbildung 3.1 zeigt ein einfaches additives Modell für die Überlagerung des Sprachsignals eines Sprechers $s[k]$ mit einem Störsignal $n[k]$, von dem nun erst einmal ausgegangen wird. Das Mikrofonsignal $x[k]$ besteht demnach aus der Summe aus Sprachsignal und Störung ($x[k] = s[k] + n[k]$). Außerdem wird angenommen, dass das Sprachsignal und die Störung gegenseitig unkorreliert sind. Das geräuschreduzierte Mikrofonsignal, also das geschätzte Sprachsignal, wird mit $\hat{s}[k]$ bezeichnet.

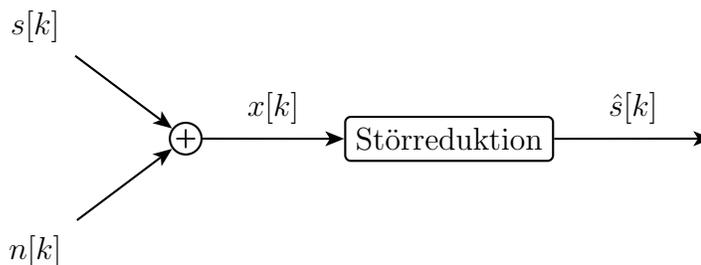


Abbildung 3.1: Modell für die additive Überlagerung von Quellensignalen und anschließender Störreduktion

In dieser Arbeit wird die Störreduktion jedoch nicht im Zeitbereich, sondern im Frequenzbereich betrachtet. Als Transformation wird die diskrete Fouriertransformation (DFT) verwendet. Die Transformationslänge sei hierbei M , so dass man aus M diskreten Abtastwerten im Zeitbereich $x[k]$ ($k \in \{0, 1, \dots, M - 1\}$) einen Satz von M komplexen DFT-Koeffizienten $X_\mu[k]$ für einen bestimmten Zeitpunkt k erhält. Dabei stellt μ den Frequenzindex dar, $\mu \in \{0, 1, \dots, M - 1\}$. μ hängt mit der normierten Mittenfrequenz jedes Frequenzbins über die Beziehung $\Omega_\mu = 2\pi\mu/M$ zusammen. Im weiteren Verlauf dieser Arbeit wird die Abhängigkeit der DFT-Koeffizienten - und deren Betrag und Phase - vom Zeitindex k jedoch bei der Notation unterschlagen, um die Lesbarkeit der Gleichungen zu vereinfachen.

Es sollen also die DFT-Koeffizienten des ungestörten Sprachsignals geschätzt werden. Wie aus Kapitel 2 hervorging, ist für eine MAP-, ML- oder MMSE-Schätzung immer die Wahrscheinlichkeitsdichte des zu schätzenden Signals erforderlich. Um hierüber Aussagen machen zu können, werden im Folgenden einige Annahmen nach [9] getroffen.

Für die weitere Vorgehensweise werden die asymptotischen statistischen Eigenschaften der DFT-Koeffizienten benutzt. Dafür wird angenommen, dass die Transformationslänge gegen unendlich geht ($M \rightarrow \infty$) und dass M viel größer ist als die Zeitspanne der Korrelation von $s[k]$. Für ein ausreichend zufälliges Signal

folgt damit aus dem zentralen Grenzwertsatz, dass für $\mu \notin \{0, M/2\}$ Real- und Imaginärteil von S_μ als voneinander unabhängige, gaußsche Zufallsvariablen mit einem Mittelwert von Null und einer Varianz von $0,5\sigma_{S,\mu}^2 = 0,5\mathcal{E}\{|S_\mu|^2\}$ modelliert werden können. Die Wahrscheinlichkeitsdichten lassen sich damit wie folgt ausdrücken:

$$\begin{aligned} f_{\text{Re}\{S_\mu\}}(\text{Re}\{S_\mu\}) &= \frac{1}{\sqrt{\pi\sigma_{S,\mu}^2}} \exp\left(-\frac{(\text{Re}\{S_\mu\})^2}{\sigma_{S,\mu}^2}\right) \\ f_{\text{Im}\{S_\mu\}}(\text{Im}\{S_\mu\}) &= \frac{1}{\sqrt{\pi\sigma_{S,\mu}^2}} \exp\left(-\frac{(\text{Im}\{S_\mu\})^2}{\sigma_{S,\mu}^2}\right) \end{aligned} \quad (3.1)$$

Für ein reellwertiges Signal $s[k]$ und für $\mu \in \{0, M/2\}$ ist der Imaginärteil von S_μ Null und der Realteil ist ebenso gaußverteilt.

Durch die statistische Unabhängigkeit von Real- und Imaginärteil von S_μ für $\mu \notin \{0, M/2\}$ lässt sich die Verbunddichte durch die Multiplikation der Randdichten aus (3.1) berechnen:

$$f_{\text{Re}\{S_\mu\}, \text{Im}\{S_\mu\}}(\text{Re}\{S_\mu\}, \text{Im}\{S_\mu\}) = \frac{1}{\pi\sigma_{S,\mu}^2} \exp\left(-\frac{(\text{Re}\{S_\mu\})^2 + (\text{Im}\{S_\mu\})^2}{\sigma_{S,\mu}^2}\right) \quad (3.2)$$

Durch das Zusammenfassen von $\text{Re}\{S_\mu\}$ und $\text{Im}\{S_\mu\}$ zu $S_\mu = \text{Re}\{S_\mu\} + j\text{Im}\{S_\mu\}$ lässt sich (3.2) auch schreiben als

$$f_{S_\mu}(S_\mu) = \frac{1}{\pi\sigma_{S,\mu}^2} \exp\left(-\frac{|S_\mu|^2}{\sigma_{S,\mu}^2}\right). \quad (3.3)$$

Zur Veranschaulichung ist diese Verbunddichte in Abbildung 3.2 dargestellt. Durch Umwandlung in die Darstellung mit Polarkoordinaten $S_\mu = A_\mu e^{j\alpha_\mu}$ folgt für den Betrag $|S_\mu| = A_\mu$ eine Rayleigh-Verteilung

$$f_{A_\mu}(A_\mu) = \begin{cases} \frac{2A_\mu}{\sigma_{S,\mu}^2} \exp\left(-\frac{A_\mu^2}{\sigma_{S,\mu}^2}\right) & A_\mu \geq 0 \\ 0 & A_\mu < 0, \end{cases} \quad (3.4)$$

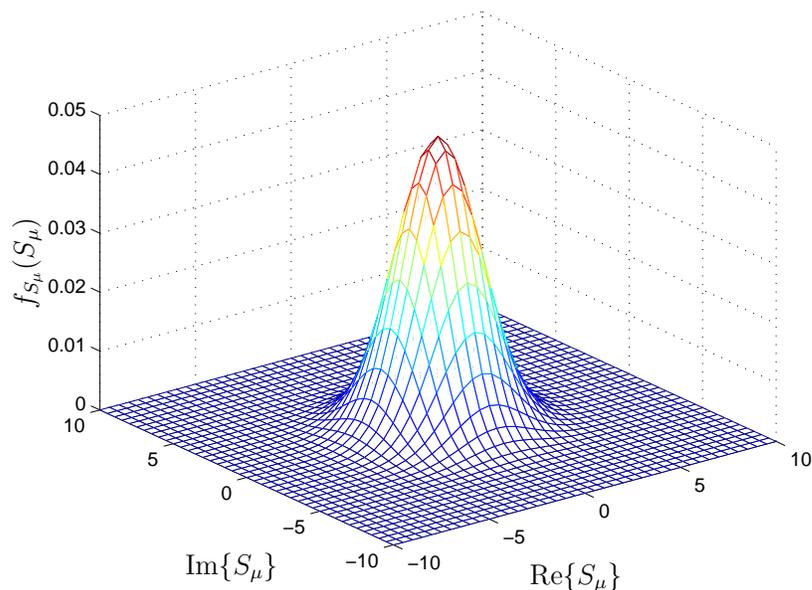


Abbildung 3.2: Verbunddichte des Real- und Imaginärteils von S_μ

und die Phase α_μ ist gleichverteilt für $0 \leq \alpha_\mu \leq 2\pi$

$$f_{\alpha_\mu}(\alpha_\mu) = \begin{cases} \frac{1}{2\pi} & 0 \leq \alpha_\mu \leq 2\pi \\ 0 & \text{sonst.} \end{cases} \quad (3.5)$$

Das bedeutet, dass für die Modellierung von Real- und Imaginärteil als gaußsche Zufallsvariablen der Betrag A_μ und die Phase α_μ von S_μ statistisch unabhängig sind.

Aufbauend auf diesen Annahmen für die Eigenschaften der Fourierkoeffizienten S_μ werden im Folgenden aus den in Kapitel 2 theoretisch hergeleiteten Schätzern konkrete Schätzer zur Störreduktion im Frequenzbereich entwickelt. Die Herleitung dieser konkretisierten Schätzer ist an [9] (Kapitel 11.4) angelehnt.

3.2 MAP-Schätzung

Die Schätzwerte nach der MAP-Schätzung lassen sich nach (2.8) wie folgt berechnen:

$$\hat{\phi}_{\text{MAP}} = \underset{\hat{\phi}}{\operatorname{argmax}} f_{\Phi|\mathbf{x}}(\hat{\phi}|\mathbf{x}) \quad (3.6)$$

Zuerst sollen die komplexwertigen Fourierkoeffizienten des ungestörten Sprachsignals S_μ geschätzt werden. Dafür muss also derjenige Koeffizient \hat{S}_μ gefunden werden, der die bedingte Wahrscheinlichkeitsdichte $f_{S_\mu|X_\mu}(S_\mu|X_\mu)$ maximiert. Durch Anwendung der Bayes'schen Regel ist es aber ebensogut möglich

$$f_{S_\mu|X_\mu}(S_\mu|X_\mu)f_{X_\mu}(X_\mu) = f_{X_\mu|S_\mu}(X_\mu|S_\mu)f_{S_\mu}(S_\mu) \quad (3.7)$$

zu maximieren. Durch die Annahme gaußscher Verteilung der DFT-Koeffizienten aus Abschnitt 3.1 folgt nach [9] daraus das bekannte Wiener-Filter und die geschätzten Koeffizienten des ungestörten Sprachsignals berechnen sich zu

$$\hat{S}_\mu = \left(1 - \frac{\mathcal{E}\{|N_\mu|^2\}}{\mathcal{E}\{|X_\mu|^2\}}\right) X_\mu. \quad (3.8)$$

Außerdem soll auch die Amplitude und die Phase des ungestörten Sprachsignals geschätzt werden. Nach [9] ergibt sich für die Phase α_μ des reinen Sprachsignals die Phase des gestörten Sprachsignals und für die Amplitude A_μ

$$\hat{A}_\mu = \frac{\eta_\mu + \sqrt{\eta_\mu^2 + 2(1 + \eta_\mu)\frac{\eta_\mu}{\gamma_\mu}}}{2(1 + \eta_\mu)} |X_\mu|. \quad (3.9)$$

Dabei entspricht $\eta_\mu = \frac{\mathcal{E}\{|S_\mu|^2\}}{\mathcal{E}\{|N_\mu|^2\}}$ dem sogenannten a-priori SNR (signal-to-noise ratio) und $\gamma_\mu = \frac{|X_\mu|^2}{\mathcal{E}\{|N_\mu|^2\}}$ dem a-posteriori SNR.

3.3 ML-Schätzung

Für einen ML-Schätzer ergibt sich der Schätzwert nach (2.11) zu

$$\hat{\phi}_{\text{ML}} = \underset{\hat{\phi}}{\operatorname{argmax}} f_{\mathbf{X}|\Phi}(\mathbf{x}|\hat{\phi}). \quad (3.10)$$

Als erstes soll die Leistung $\sigma_{S,\mu}^2$ des ungestörten Sprachsignals geschätzt werden. Durch die Annahme einer Gaußverteilung für die komplexwertigen Fourierkoeffizienten ergibt sich die Leistung des beobachteten, gestörten Sprachsignals $\sigma_{X,\mu}^2$ als Summe aus den Leistungen der ungestörten Sprache $\sigma_{S,\mu}^2$ und der Störung $\sigma_{N,\mu}^2$. Eingesetzt in (3.3) für die gemeinsame Wahrscheinlichkeitsdichte von Real- und Imaginärteil von X_μ ergibt sich die in (3.10) zu maximierende Wahrscheinlichkeitsdichte laut [9] zu

$$f(\operatorname{Re}\{X_\mu\}, \operatorname{Im}\{X_\mu\} | \sigma_{S,\mu}^2) = \frac{1}{\pi(\sigma_{S,\mu}^2 + \sigma_{N,\mu}^2)} \exp\left(-\frac{|X_\mu|^2}{\sigma_{S,\mu}^2 + \sigma_{N,\mu}^2}\right). \quad (3.11)$$

Das Maximum erhält man durch Bilden der Ableitung nach $\sigma_{S,\mu}^2$ und anschließendem Nullsetzen. Nach [9] erhält man damit den gesuchten Schätzwert $\hat{\sigma}_{S,\mu}^2$ zu

$$\hat{\sigma}_{S,\mu}^2 = |X_\mu|^2 - \sigma_{N,\mu}^2 = R_\mu^2 - \sigma_{N,\mu}^2. \quad (3.12)$$

R_μ ist hier der Betrag des gestörten Sprachsignals X_μ . In einer praktischen Realisierung dieses Schätzers muss darauf geachtet werden, dass die geschätzte Leistung des Sprachsignals nicht negativ ist. Verwendet man nun die Phase θ_μ des gestörten Sprachsignals, so erhält man laut [9] als Schätzer für einen DFT-Koeffizienten des reinen Sprachsignals

$$\hat{S}_\mu = \sqrt{1 - \frac{\sigma_{N,\mu}^2}{|X_\mu|^2}} R_\mu e^{j\theta_\mu} = \sqrt{1 - \frac{\sigma_{N,\mu}^2}{|X_\mu|^2}} X_\mu. \quad (3.13)$$

Für eine praktische Implementierung ist hier dafür zu sorgen, dass der Ausdruck unter der Wurzel nicht negativ wird.

Mit $S_\mu = A_\mu e^{j\alpha_\mu}$ lässt sich, ausgehend von Gleichung 5.119 aus [9], welche die gemeinsame Wahrscheinlichkeitsdichte von Real- und Imaginärteil eines Fourierkoeffizienten des gestörten Sprachsignals X_μ mit gegebenem Koeffizienten des ungestörten Sprachsignals S_μ beschreibt, auch ein ML-Schätzer für die Amplitude A_μ des ungestörten Sprachsignals herleiten. Die Amplitude A_μ ist nun der zu schätzende Parameter, die Phase α_μ wird als gleichverteilt angenommen, mit

$0 \leq \alpha_\mu \leq 2\pi$. Um A_μ aus Gleichung 5.119 aus [9] zu erhalten, ist es notwendig, über die Phase zu mitteln:

$$\begin{aligned} f_{X_\mu|A_\mu}(X_\mu|A_\mu) &= \int_0^{2\pi} f_{X_\mu|A_\mu,\alpha_\mu}(X_\mu|A_\mu, \alpha_\mu) f_{\alpha_\mu}(\alpha_\mu) d\alpha_\mu \\ &= \frac{1}{\pi\sigma_{N,\mu}^2} \exp\left(-\frac{|X_\mu|^2 + A_\mu^2}{\sigma_{N,\mu}^2}\right) \frac{1}{2\pi} \int_0^{2\pi} \exp\left(\frac{2A_\mu \operatorname{Re}\{X_\mu e^{-j\alpha_\mu}\}}{\sigma_{N,\mu}^2}\right) d\alpha_\mu \end{aligned} \quad (3.14)$$

Nach Approximation des Integrals lässt sich (3.14) vereinfachen zu

$$f_{X_\mu|A_\mu}(X_\mu|A_\mu) \approx \frac{1}{2\pi\sqrt{\pi A_\mu |X_\mu| \sigma_{N,\mu}^2}} \exp\left(-\frac{(A_\mu - |X_\mu|)^2}{\sigma_{N,\mu}^2}\right). \quad (3.15)$$

Von dieser Gleichung gilt es das Maximum bezüglich A_μ zu finden, was wieder durch Nullsetzen der Ableitung geschieht. Dadurch erhält man nach [9] als ML-Schätzwert für die Amplitude A_μ

$$\hat{A}_\mu = \frac{|X_\mu|}{2} \left(1 \pm \sqrt{1 - \frac{\sigma_{N,\mu}^2}{|X_\mu|^2}}\right). \quad (3.16)$$

Mit Verwendung der Phase des gestörten Sprachsignals erhält man als Schätzwert für einen Fourierkoeffizienten des ungestörten Sprachsignals

$$\hat{S}_\mu = \left(0,5 + 0,5\sqrt{1 - \frac{\sigma_{N,\mu}^2}{|X_\mu|^2}}\right) X_\mu. \quad (3.17)$$

Auch hier ist bei einer praktischen Realisierung wieder darauf zu achten, dass der Ausdruck unter der Wurzel nicht negativ wird. Außerdem fällt auf, dass die gestörten DFT-Koeffizienten immer mindestens mit 0,5 gewichtet werden, wodurch eine maximale Geräuschreduktion von 6 dB erreicht werden kann. Dies wird bei der Evaluierung in Kapitel 4 im Vergleich zu den anderen Schätzern deutlich bemerkbar werden.

3.4 MMSE-Schätzung

Mithilfe der MMSE-Schätzung werden nun zuerst wieder die Fourierkoeffizienten des reinen Sprachsignals S_μ geschätzt. Nach (2.15) berechnet sich der MMSE-

Schätzwert als der bedingte Erwartungswert

$$\hat{S}_\mu = \mathcal{E}\{S_\mu | X_\mu\}. \quad (3.18)$$

Für diese Schätzung wird wieder auf die Annahmen aus Abschnitt 3.1 Bezug genommen, dass die DFT-Koeffizienten als gaußverteilte Zufallsvariablen modelliert werden können und dass deren Real- und Imaginärteil statistisch unabhängig voneinander sind. Basierend auf diesen Annahmen lässt sich die MMSE-Schätzung in zwei unabhängige Schätzungen für den Real- und den Imaginärteil von S_μ schreiben. (3.18) wird damit zu

$$\begin{aligned} \hat{S}_\mu &= \int_{-\infty}^{\infty} \text{Re}\{S_\mu\} f(\text{Re}\{S_\mu\} | \text{Re}\{X_\mu\}) d\text{Re}\{S_\mu\} \\ &\quad + j \int_{-\infty}^{\infty} \text{Im}\{S_\mu\} f(\text{Im}\{S_\mu\} | \text{Im}\{X_\mu\}) d\text{Im}\{S_\mu\} \\ &= \mathcal{E}\{\text{Re}\{S_\mu\} | \text{Re}\{X_\mu\}\} + j\mathcal{E}\{\text{Im}\{S_\mu\} | \text{Im}\{X_\mu\}\}. \end{aligned} \quad (3.19)$$

Mit den Wahrscheinlichkeitsdichten für die Realteile von S_μ und X_μ nach (3.1) ergibt sich der Realteil des zu schätzenden Koeffizienten nach [9] zu

$$\mathcal{E}\{\text{Re}\{S_\mu\} | \text{Re}\{X_\mu\}\} = \frac{\sigma_{S,\mu}^2}{\sigma_{S,\mu}^2 + \sigma_{N,\mu}^2} \text{Re}\{X_\mu\}. \quad (3.20)$$

Für den Imaginärteil erhält man wieder mit (3.1) analog das Ergebnis

$$\mathcal{E}\{\text{Im}\{S_\mu\} | \text{Im}\{X_\mu\}\} = \frac{\sigma_{S,\mu}^2}{\sigma_{S,\mu}^2 + \sigma_{N,\mu}^2} \text{Im}\{X_\mu\}. \quad (3.21)$$

Zusammengefasst ergibt sich damit für den Schätzwert eines DFT-Koeffizienten S_μ

$$\hat{S}_\mu = \mathcal{E}\{S_\mu | X_\mu\} = \frac{\sigma_{S,\mu}^2}{\sigma_{S,\mu}^2 + \sigma_{N,\mu}^2} X_\mu, \quad (3.22)$$

was allerdings nichts anderes ist als das Wiener-Filter für die DFT-Koeffizienten. Die Phase des gestörten Sprachsignals bleibt hierbei erhalten.

Als letzte Möglichkeit der Anwendung eines Schätzers zur Störreduktion soll nun mit einem MMSE-Schätzer die Amplitude des ungestörten Sprachsignals A_μ

geschätzt werden. Wie in Abschnitt 2.3 gezeigt, ist der gesuchte Schätzwert zunächst einmal gegeben als der bedingte Erwartungswert

$$\hat{A}_\mu = \mathcal{E}\{A_\mu | X_\mu\}. \quad (3.23)$$

Nach seiner Definition berechnet sich der bedingte Erwartungswert zu

$$\begin{aligned} \mathcal{E}\{A_\mu | X_\mu\} &= \int_0^\infty A_\mu f_{A_\mu | X_\mu}(A_\mu | X_\mu) dA_\mu \\ &= \int_0^\infty \int_0^{2\pi} A_\mu f_{A_\mu, \alpha_\mu | X_\mu}(A_\mu, \alpha_\mu | X_\mu) d\alpha_\mu dA_\mu, \end{aligned} \quad (3.24)$$

wobei α_μ die Phase des ungestörten Sprachsignals S_μ ist. Durch Anwendung der Bayes'schen Regel

$$f_{A_\mu, \alpha_\mu | X_\mu}(A_\mu, \alpha_\mu | X_\mu) = \frac{f_{X_\mu | A_\mu, \alpha_\mu}(X_\mu | A_\mu, \alpha_\mu) f_{A_\mu, \alpha_\mu}(A_\mu, \alpha_\mu)}{\int_0^\infty \int_0^{2\pi} f_{X_\mu | A_\mu, \alpha_\mu}(X_\mu | A_\mu, \alpha_\mu) f_{A_\mu, \alpha_\mu}(A_\mu, \alpha_\mu) d\alpha_\mu dA_\mu} \quad (3.25)$$

lässt sich der Schätzwert laut [9] schreiben als

$$\hat{A}_\mu = \frac{\int_0^\infty \int_0^{2\pi} A_\mu f_{X_\mu | A_\mu, \alpha_\mu}(X_\mu | A_\mu, \alpha_\mu) f_{A_\mu, \alpha_\mu}(A_\mu, \alpha_\mu) d\alpha_\mu dA_\mu}{\int_0^\infty \int_0^{2\pi} f_{X_\mu | A_\mu, \alpha_\mu}(X_\mu | A_\mu, \alpha_\mu) f_{A_\mu, \alpha_\mu}(A_\mu, \alpha_\mu) d\alpha_\mu dA_\mu}. \quad (3.26)$$

Mithilfe der Annahmen aus Abschnitt 3.1, dass Real- und Imaginärteil der DFT-Koeffizienten gaußverteilt sind, dass die Phase des reinen Sprachsignals gleichverteilt ist und dass die Amplitude und die Phase des reinen Sprachsignals statistisch unabhängig voneinander sind, wird daraus

$$\hat{A}_\mu = \frac{\int_0^\infty A_\mu f_{A_\mu}(A_\mu) \int_0^{2\pi} \exp\left(-\frac{|X_\mu - A_\mu e^{j\alpha_\mu}|^2}{\sigma_{N,\mu}^2}\right) d\alpha_\mu dA_\mu}{\int_0^\infty f_{A_\mu}(A_\mu) \int_0^{2\pi} \exp\left(-\frac{|X_\mu - A_\mu e^{j\alpha_\mu}|^2}{\sigma_{N,\mu}^2}\right) d\alpha_\mu dA_\mu}. \quad (3.27)$$

Das Integral über die Phase α_μ lässt sich umgehen, indem man es durch einen Ausdruck mit der modifizierten Besselfunktion nullter Ordnung I_0 ersetzt:

$$\frac{1}{2\pi} \int_0^{2\pi} \exp\left(-\frac{|X_\mu - A_\mu e^{j\alpha_\mu}|^2}{\sigma_{N,\mu}^2}\right) d\alpha_\mu = \exp\left(-\frac{|X_\mu|^2 + A_\mu^2}{\sigma_{N,\mu}^2}\right) I_0\left(\frac{2|X_\mu|A_\mu}{\sigma_{N,\mu}^2}\right) \quad (3.28)$$

Eingesetzt in (3.27) erhält man

$$\hat{A}_\mu = \frac{\int_0^\infty A_\mu f_{A_\mu}(A_\mu) \exp\left(-\frac{|X_\mu|^2 + A_\mu^2}{\sigma_{N,\mu}^2}\right) I_0\left(\frac{2|X_\mu|A_\mu}{\sigma_{N,\mu}^2}\right) dA_\mu}{\int_0^\infty f_{A_\mu}(A_\mu) \exp\left(-\frac{|X_\mu|^2 + A_\mu^2}{\sigma_{N,\mu}^2}\right) I_0\left(\frac{2|X_\mu|A_\mu}{\sigma_{N,\mu}^2}\right) dA_\mu}. \quad (3.29)$$

Nach [2] lässt sich der Schätzer nun ableiten zu

$$\hat{A}_\mu = \frac{|X_\mu|}{\gamma_\mu} \sqrt{v_\mu} \Gamma(1, 5) F_1(-0.5, 1, -v_\mu), \quad (3.30)$$

wobei sich v_μ aus dem a-priori SNR η_μ und dem a-posteriori SNR γ_μ zu $v_\mu = \frac{\eta_\mu}{1+\eta_\mu} \gamma_\mu$ berechnet. $\Gamma(\cdot)$ ist hierbei die Gammafunktion mit $\Gamma(1, 5) = \frac{\pi}{2}$ und $F_1(\cdot, \cdot, \cdot)$ die hypergeometrische Funktion nach [2]. Durch die Verwendung der Phase des gestörten Sprachsignals bekommt man als Schätzwert für einen DFT-Koeffizienten des reinen Sprachsignals

$$\hat{S}_\mu = \frac{\sqrt{v_\mu}}{\gamma_\mu} \Gamma(1, 5) F_1(-0.5, 1, -v_\mu) X_\mu. \quad (3.31)$$

Als Abschluss dieses Kapitels sind die verwendeten Schätzer und die dazugehörigen Schätzgrößen noch einmal in Tabelle 3.1 zusammengefasst.

Schätzer	geschätzter Parameter	Schätzergebnis für \hat{S}_μ
MAP	S_μ	$\hat{S}_\mu = \left(1 - \frac{E\{ N_\mu ^2\}}{E\{ X_\mu ^2\}}\right) X_\mu$
	A_μ, α_μ	$\hat{S}_\mu = \frac{\eta_\mu + \sqrt{\eta_\mu^2 + 2(1+\eta_\mu)\frac{\eta_\mu}{\gamma_\mu}}}{2(1+\eta_\mu)} X_\mu$
ML	$\sigma_{S,\mu}^2$	$\hat{S}_\mu = \sqrt{1 - \frac{\sigma_{N,\mu}^2}{ X_\mu ^2}} X_\mu$
	A_μ	$\hat{S}_\mu = \left(0, 5 + 0, 5 \sqrt{1 - \frac{\sigma_{N,\mu}^2}{ X_\mu ^2}}\right) X_\mu$
MMSE	S_μ	$\hat{S}_\mu = \frac{\sigma_{S,\mu}^2}{\sigma_{S,\mu}^2 + \sigma_{N,\mu}^2} X_\mu$
	A_μ	$\hat{S}_\mu = \frac{\sqrt{v_\mu}}{\gamma_\mu} \Gamma(1, 5) F_1(-0.5, 1, -v_\mu) X_\mu$

Tabelle 3.1: Schätzergebnisse der verschiedenen Schätzer

Kapitel 4

Evaluierung der verschiedenen Methoden durch Simulation

Aus den in Kapitel 2 diskutierten Schätzmethoden sind in Kapitel 3 verschiedene Schätzer entwickelt worden, mit denen sich die DFT-Koeffizienten eines Sprachsignals aus einer verrauschten Beobachtung berechnen lassen. Im Folgenden sollen nun die einzelnen Verfahren für realitätsnahe Szenarien miteinander verglichen werden.

Zu diesem Zweck wird eine bereits existierende Implementierung des Wiener-Filters zur Störreduktion in binauralen Hörgeräten mit den in Tabelle 3.1 zusammengefassten Schätzern erweitert. In diesem Programm wird eine binaurale Störgeräuschreduktion nach Abbildung 4.1 im Frequenzbereich durchgeführt. Fettgedruckte Kleinbuchstaben bezeichnen hierbei Spaltenvektoren, die das Spektrum eines Signals oder die spektralen Gewichte eines *Multiple-Input-Single-Output* (MISO)-Systems darstellen. *Multiple-Input-Multiple-Output* (MIMO)-Systeme werden durch Matrizen bezeichnet. Ihre Darstellung erfolgt mit fettgedruckten Großbuchstaben. Die Abhängigkeit der Größen vom Frequenzindex μ und vom Blockindex n wird bei der folgenden Beschreibung aus Gründen der Übersichtlichkeit vernachlässigt.

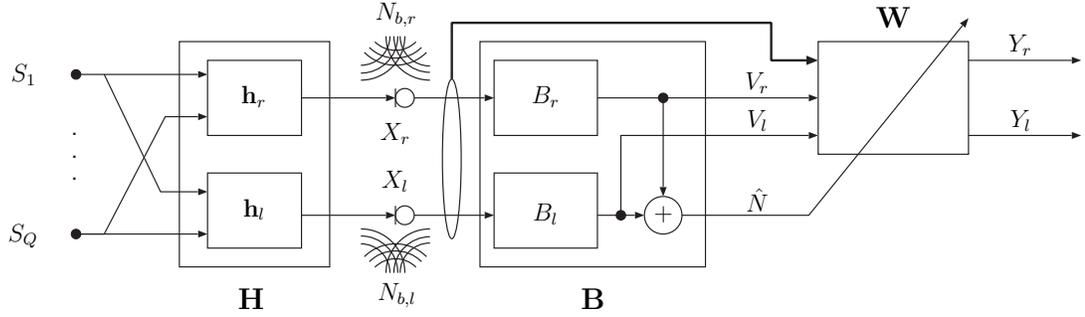


Abbildung 4.1: Blockschaltbild der binauralen Geräuschreduktion nach [8]

Die Spektren der Mikrofon-signale X_r und X_l entstehen hierbei aus Q Quellensignalen S_q ($q \in \{1, \dots, Q\}$) und sind zusätzlich mit dem diffusen Hintergrundrauschen $N_{b,m}$ ($m \in \{r, l\}$) überlagert. Als Nutzsignal wird für die weitere Arbeit S_1 definiert, während die restlichen $Q - 1$ Quellen Störquellen darstellen. Die Reflexionen der verschiedenen Quellensignale im Raum und die Abschattungseffekte am Kopf können durch die Übertragungsfunktionen \mathbf{h}_r und \mathbf{h}_l modelliert werden. Diese wurden in einer wohnzimmerähnlichen Umgebung ($T_{60} \approx 300$) wie in [6] beschrieben gemessen. Jedes Quellensignal erreicht jedes Mikrofon also durch Filterung mit der Übertragungsfunktion H_{qm} ($q \in \{1, \dots, Q\}$ und $m \in \{r, l\}$), wobei H_{qm} die Übertragungsfunktion vom q -ten Quellensignal zum m -ten Mikrofon darstellt. Demnach lassen sich die beiden Mikrofon-signale im Frequenzbereich unter Einbeziehung des Hintergrundrauschens zu

$$X_m = \sum_{q=1}^Q H_{qm} S_q + N_{b,m} \quad m \in \{r, l\} \quad (4.1)$$

berechnen. In Matrix-Schreibweise lässt sich dies auch vereinfacht darstellen als

$$\mathbf{x} = \mathbf{H}^T \mathbf{s} + \mathbf{n}_b, \quad (4.2)$$

wobei $\mathbf{x} = [X_r, X_l]^T$ die beiden Mikrofon-signale, $\mathbf{n}_b = [N_{b,r}, N_{b,l}]^T$ das überlagerte Hintergrundrauschen beider Kanäle und $\mathbf{s} = [S_1, \dots, S_Q]^T$ alle Quellensignale

enthält. Das akustische Mischsystem \mathbf{H} ergibt sich zu

$$\mathbf{H} = [\mathbf{h}_r \quad \mathbf{h}_l] = \begin{bmatrix} H_{1r} & H_{1l} \\ \vdots & \vdots \\ H_{Qr} & H_{Ql} \end{bmatrix}. \quad (4.3)$$

Die Blockiermatrix \mathbf{B} besitzt nun die Aufgabe, eine gemeinsame Störschätzung \hat{N} zu erzeugen, die sowohl die Störung durch die Punktquellen S_q ($q \in \{2, \dots, Q\}$) als auch die Störung durch das Hintergrundrauschen \mathbf{n}_b beinhaltet. Diese wird anschließend für die Berechnung der Filter zur Störreduktion benötigt. Mit den spektralen Gewichten B_m ($m \in \{r, l\}$) und $\mathbf{b} = [B_r, B_l]^T$ ist sie nach [8] definiert als

$$\mathbf{B} = [\mathbf{b} \text{ diag}\{\mathbf{b}\}], \quad (4.4)$$

wobei der Operator $\text{diag}\{\cdot\}$ eine Diagonalmatrix mit den Elementen aus \mathbf{b} auf der Hauptdiagonalen erzeugt. Dies ist notwendig, um die gefilterten Mikrofonsignale V_m ($m \in \{r, l\}$) zu erzeugen. Die Ausgangssignale der Blockiermatrix ergeben sich schließlich zu

$$[\hat{N} \quad V_r \quad V_l]^T = \mathbf{B}^T \mathbf{x}. \quad (4.5)$$

Für die Berechnung der spektralen Gewichte B_m ($m \in \{r, l\}$) wird *directional BSS* nach [10] verwendet. Für diese Blockiermatrix wurden bei allen Simulationen 1024 Koeffizienten verwendet.

Die bisherige Verarbeitung der Signale wurde im Frequenzbereich beschrieben und auch die Störgeräuschreduktion soll im Frequenzbereich erfolgen. Die Mikrofonsignale müssen also zuerst einmal in den Frequenzbereich transformiert werden. Dies erfolgt mit der diskreten Fouriertransformation nach Abbildung 4.2 in einer Polyphasenfilterbank. Eine Übersicht über die verwendete Polyphasenfilterbank gibt Abbildung 4.3. Details hierzu können aus [7] entnommen werden. Für die Simulationen werden Prototyp-Tiefpassfilter $h[k]$ verschiedener Längen L benutzt. Angefangen bei einer Länge von 512, werden die Simulationen noch bei einer Länge von 256 und schließlich noch bei einer Länge von 128 durchgeführt.

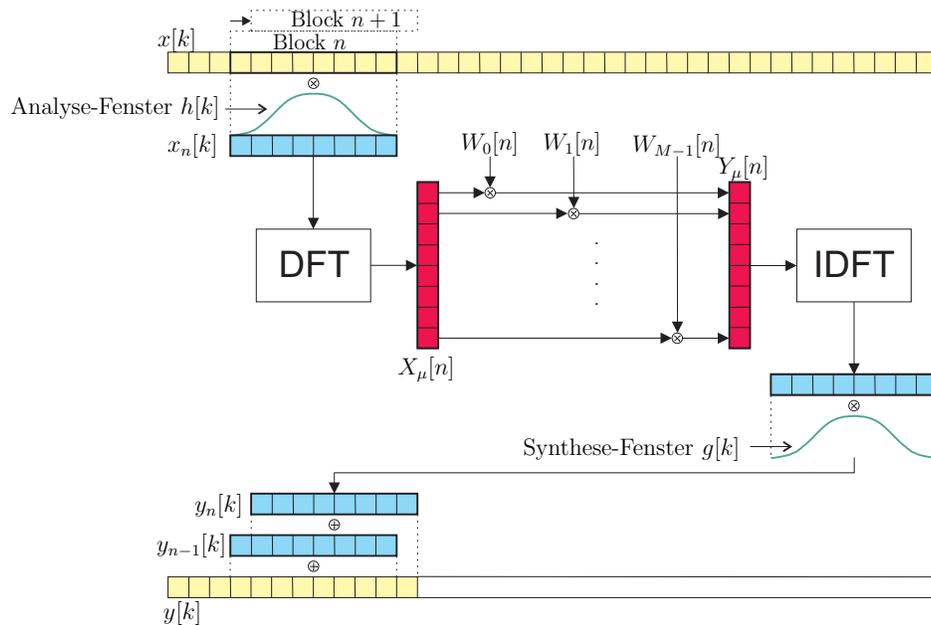


Abbildung 4.2: Prinzip der spektralen Analyse durch Verwendung der DFT [7]

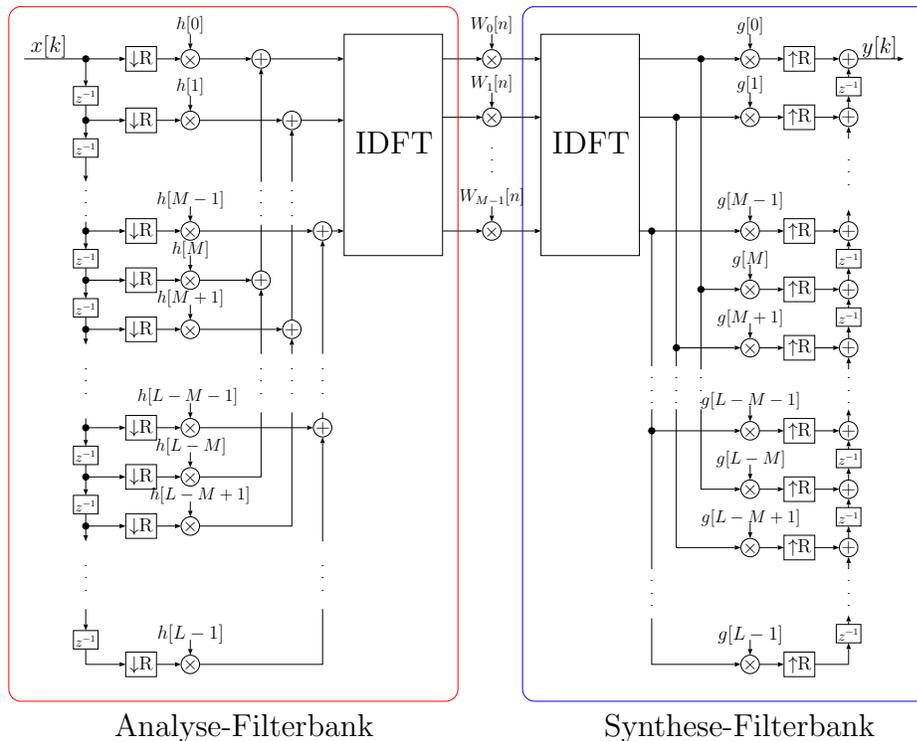


Abbildung 4.3: Realisierung der Polyphasenfilterbank [4]

Der Grund hierfür ist, dass für eine Filterlänge von 512 bessere Ergebnisse erwartet werden, für die Anwendung in Hörgeräten aufgrund der Komplexität und zulässigen Gesamtlaufzeit allerdings nur eine Länge von 128 bei einer Abtastrate von 16 kHz verwendet werden kann. Die Frequenzauflösung M ist immer als die Hälfte der Filterlänge der Prototyp-Tiefpassfilter gegeben und die Abtastratenreduktion R beträgt immer ein Viertel der Frequenzauflösung. Dabei handelt es sich um eine stark überabgetastete Filterbank, dies wurde jedoch so gewählt, damit die Aliasingeffekte möglichst niedrig und speziell in einer Hörgeräteanwendung nicht hörbar sind. Aufgrund der großen Anzahl an Simulationsergebnissen, die durch die Simulationen mit verschiedenen Filterlängen entstehen, werden nur die Ergebnisse mit einer Filterlänge von 512 in dieser Arbeit diskutiert. Die Ergebnisse der Simulationen mit den anderen Filterlängen sind im Anhang A zu finden.

Die eigentliche Geräuschreduktion geschieht nun schließlich durch die Matrix \mathbf{W} (siehe Abbildung 4.1) mit $\mathbf{y} = [Y_r, Y_l]^T$ und

$$\mathbf{y} = \mathbf{W}\mathbf{x}. \quad (4.6)$$

\mathbf{W} besteht hierbei aus den spektralen Gewichten W_m ($m \in \{r, l\}$) mit

$$\mathbf{w} = [W_r \quad W_l]^T \quad (4.7)$$

und

$$\mathbf{W} = \text{diag}\{\mathbf{w}\}. \quad (4.8)$$

Die Berechnung der spektralen Gewichte W_m ($m \in \{r, l\}$) erfolgte bisher durch das Konzept eines Wiener-Filters nach [8]. Dabei werden zwei verschiedene Herangehensweisen unterschieden:

1. Berechnung verschiedener (unabhängiger) spektraler Gewichte für die Filterung am linken und am rechten Ohr
2. Berechnung eines gemeinsamen Filters für die linke und die rechte Seite

Für den Ansatz zweier verschiedener Filter (1. Ansatz) berechnen sich die spektralen Gewichte W_m ($m \in \{r, l\}$) nach [8] zu

$$W_m = \max \left\{ 1 - \beta \frac{\hat{\Phi}_{\hat{N}\hat{N}}}{\hat{\Phi}_{V_m V_m}}, W_{\min} \right\} \quad m \in \{r, l\}. \quad (4.9)$$

$\hat{\Phi}_{\hat{N}\hat{N}}$ und $\hat{\Phi}_{V_m V_m}$ stellen hierbei die geschätzten Leistungsdichtespektren des geschätzten Störsignals und der gefilterten Mikrofon-signale dar. Diese erhält man als rekursive Mittelung zwischen dem Betragsquadrat der DFT-Koeffizienten des aktuellen Datenblocks und dem Leistungsdichtespektrum des vergangenen Datenblocks:

$$\begin{aligned} \hat{\Phi}_{\hat{N}\hat{N}}[k] &= \lambda \cdot \hat{\Phi}_{\hat{N}\hat{N}}[k-1] + (1-\lambda) \cdot |\hat{N}[k]|^2 \\ \hat{\Phi}_{V_m V_m}[k] &= \lambda \cdot \hat{\Phi}_{V_m V_m}[k-1] + (1-\lambda) \cdot |V_m[k]|^2 \end{aligned} \quad (4.10)$$

Durch den Parameter λ lässt sich einstellen, wie stark vergangene Werte gewichtet werden. Er liegt bei den hier durchgeführten Simulationen zwischen 0,85 und 0,92. Abbildung 4.4 zeigt das rekursive System 1. Ordnung zur Schätzung der Leistungsdichtespektren $\hat{\Phi}_{V_m V_m}$. Für $\hat{\Phi}_{\hat{N}\hat{N}}$ erfolgt die Berechnung analog.

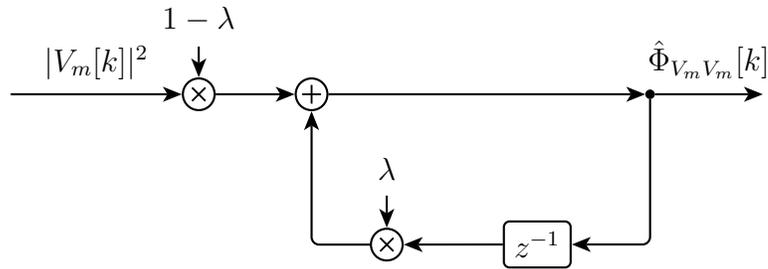


Abbildung 4.4: Rekursives System zur Schätzung des Leistungsdichtespektrums

Für den zweiten Ansatz, dem gemeinsamen Filter, ergeben sich die spektralen Gewichte laut [8] zu

$$W = W_r = W_l = \max \left\{ 1 - \beta \frac{\hat{\Phi}_{\hat{N}\hat{N}}}{\hat{\Phi}_{V_r V_r} + \hat{\Phi}_{V_l V_l}}, W_{\min} \right\}. \quad (4.11)$$

Die Schätzung der Leistungsdichtespektren erfolgt analog zum Ansatz getrennter Filter. Der Parameter β ist in beiden Ansätzen ein Gewichtungsfaktor, welcher einen Kompromiss zwischen Geräuschreduktion und Nutzsignalverzerrung erlaubt. Je größer er gewählt wird, desto größer ist auch die Geräuschreduktion, allerdings steigt damit auch gleichzeitig die Nutzsignalverzerrung. Als guter Kompromiss zwischen Geräuschreduktion und Nutzsignalverzerrung hat sich für den Ansatz zweier getrennter Filter der Wert 0,95 und für den gemeinsamen Filteransatz der Wert 1,1 bewährt. Diese beiden Werte werden auch für die Simulationen in dieser Arbeit verwendet. Der Parameter W_{\min} ($0 \leq W_{\min} \leq 1$) in (4.9) und (4.11) wird als *spectral floor* bezeichnet. Er verhindert, dass die spektralen Gewichte W_m ($m \in \{r, l\}$) zu klein, oder aufgrund von Schätzfehlern (siehe [8]) sogar negativ werden, was zu störenden Effekten im geräuschreduzierten Signal führt. Für den getrennten Filteransatz wird $W_{\min} = 0,2$ und für den Ansatz gemeinsamer Filter $W_{\min} = 0,15$ gewählt.

Als Referenz für alle folgenden Simulationsergebnisse dienen die Ergebnisse aus den Simulationen mit der Berechnung der spektralen Gewichte W_m ($m \in \{r, l\}$) nach (4.9) und (4.11). Zusätzlich zu diesem Ansatz des Wiener-Filters, basierend auf den geschätzten Leistungsdichtespektren, wird als zweites Konzept die Leistungsfähigkeit eines Wiener-Filters basierend auf dem geschätzten a-priori und a-posteriori SNR nach [9] untersucht. Anschließend wird die Geräuschreduktion mithilfe der MAP-, ML- und MMSE-Schätzer aus Kapitel 3 simuliert und die Ergebnisse werden mit denen der beiden Wiener-Filter Ansätze verglichen. Dabei werden allerdings nur diejenigen Schätzer betrachtet, die sich vom Wiener-Filter unterscheiden. Für alle verwendeten Filtertypen wird sowohl der Ansatz zweier getrennter Filter, als auch der gemeinsame Filteransatz simuliert. Außerdem werden, wie bereits erwähnt, jeweils Prototyp-Tiefpassfilter der Längen 512, 256 und 128 verwendet. Als Mikrofonssignale dienen aufgezeichnete Signale, wobei mehrere Simulationen mit verschieden vielen Störquellen durchgeführt werden. Diese sind in Abbildung 4.5 zusammengefasst. Beginnend mit einer Nutz- und einer

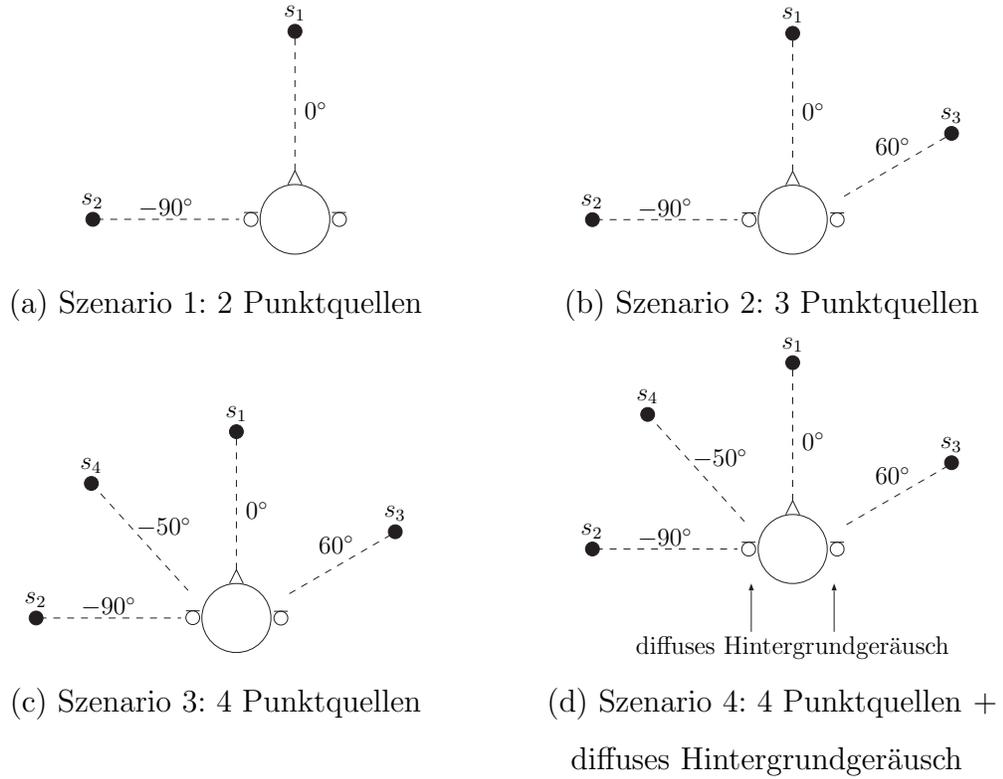


Abbildung 4.5: Untersuchte Szenarien; der gewünschte Sprecher befindet sich immer ungefähr bei 0° vor dem Hörgerätbenutzer

Störquelle, kommen noch weitere Störquellen hinzu bis schließlich auch noch ein diffuses Hintergrundgeräusch überlagert wird. Insgesamt stehen also vier Punktquellen, wovon eine die Nutzsignalquelle ist, und das Hintergrundgeräusch zur Verfügung.

Für den zweiten, neuen Ansatz des Wiener-Filters und für die MAP- und die MMSE-Schätzung der Amplitude A_μ des reinen Sprachsignals ist das a-priori SNR η_μ und das a-posteriori SNR γ_μ erforderlich. Das a-posteriori SNR ist nach [9] definiert als

$$\gamma_\mu[k] = \frac{|V_\mu[k]|^2}{\mathcal{E}\{|\hat{N}_\mu[k]|^2\}} \quad (4.12)$$

und das a-priori SNR lässt sich ebenfalls nach [9] zu

$$\hat{\eta}_\mu[k] = \alpha_\eta \cdot \frac{|\hat{S}_\mu[k-1]|^2}{\mathcal{E}\{|\hat{N}_\mu[k-1]|^2\}} + (1 - \alpha_\eta) \cdot \max\{\gamma_\mu[k] - 1, 0\} \quad (4.13)$$

schätzen, wobei auf die Wahl des Parameters α_η bei der Behandlung der jeweiligen Schätzer genauer eingegangen wird. Für das a-posteriori SNR γ_μ hat sich während der Simulationen jedoch herausgestellt, dass sich bei einer Mittelung von $|V_\mu[k]|^2$ über die Zeit bessere Ergebnisse erzielen lassen. Da hier immer ergodische Prozesse betrachtet werden, kann man $\mathcal{E}\{|\hat{N}_\mu[k]|^2\}$ auch als Mittelung über die Zeit berechnen. Damit lässt sich γ_μ aus den geschätzten Leistungsdichtespektren zu

$$\gamma_\mu[k] = \frac{\hat{\Phi}_{VV}[k]}{\hat{\Phi}_{\hat{N}\hat{N}}[k]} \quad (4.14)$$

berechnen. Für den Ansatz zweier getrennter Filter ergeben sich hierfür auch zwei getrennte Ergebnisse und das Leistungsdichtespektrum $\hat{\Phi}_{VV}[k]$ steht für das Leistungsdichtespektrum eines der beiden gefilterten Mikrofonsignale V_m ($m \in \{r, l\}$). Bei Verwendung eines gemeinsamen Filters ist für $\hat{\Phi}_{VV}[k]$ die Summe aus den Leistungsdichtespektren der gefilterten Mikrofonsignale einzusetzen, also $\hat{\Phi}_{VV}[k] = \hat{\Phi}_{V_r V_r}[k] + \hat{\Phi}_{V_l V_l}[k]$.

Auch für die Bestimmung des a-priori SNRs wird $\mathcal{E}\{|\hat{N}_\mu[k-1]|^2\}$ wieder als Mittelung über die Zeit berechnet und durch das geschätzte Leistungsdichtespektrum des geschätzten Störsignals $\hat{\Phi}_{\hat{N}\hat{N}}[k-1]$ ersetzt:

$$\hat{\eta}_\mu[k] = \alpha_\eta \cdot \frac{|\hat{S}_\mu[k-1]|^2}{\hat{\Phi}_{\hat{N}\hat{N}}[k-1]} + (1 - \alpha_\eta) \cdot \max\{\gamma_\mu[k] - 1, 0\} \quad (4.15)$$

Für den Ansatz eines gemeinsamen Filters wird das Betragsquadrat des geschätzten Sprachsignals $|\hat{S}_\mu[k-1]|^2$ als Summe aus den Betragsquadraten der beiden Kanäle zu $|\hat{S}_\mu[k-1]|^2 = |\hat{S}_{\mu,r}[k-1]|^2 + |\hat{S}_{\mu,l}[k-1]|^2$ berechnet. Je nachdem, ob ein gemeinsames oder zwei getrennte Filter berechnet werden sollen, ergeben sich auch hier wieder ein oder zwei Werte.

Die Berechnungsvorschrift für die spektralen Gewichte W_m ($m \in \{r, l\}$) nach dem Wiener-Filter, basierend auf dem a-priori SNR, erhält man ausgehend von Gleichung 11.32 aus [9]:

$$W_m = \frac{\mathcal{E}\{|S_\mu|^2\}}{\mathcal{E}\{|S_\mu|^2\} + \mathcal{E}\{|N_\mu|^2\}} \quad m \in \{r, l\} \quad (4.16)$$

Durch ausklammern und kürzen von $\mathcal{E}\{|N_\mu|^2\}$ ergibt sich mit der Definition des a-priori SNRs η_μ nach [9]

$$W_m = \frac{\frac{\mathcal{E}\{|S_\mu|^2\}}{\mathcal{E}\{|N_\mu|^2\}}}{\frac{\mathcal{E}\{|S_\mu|^2\}}{\mathcal{E}\{|N_\mu|^2\}} + 1} = \frac{\eta_\mu}{\eta_\mu + 1}. \quad (4.17)$$

Auch hier sollen die spektralen Gewichte nicht beliebig klein werden können, um ungewollte Verzerrungen im gefilterten Signal zu vermeiden. Deshalb wird (4.17) modifiziert zu

$$W_m = \max \left\{ \frac{\eta_\mu}{\eta_\mu + 1}, W_{\min} \right\} \quad m \in \{r, l\}, \quad (4.18)$$

wobei W_{\min} wieder dem sogenannten *spectral floor* entspricht.

Diese Vorgehensweise soll nun auch auf die Berechnung der spektralen Gewichte durch die in Kapitel 3 behandelten Schätzer (siehe Zusammenfassung in Tabelle 3.1) angewandt werden. Der MAP-Schätzwert für das reine Sprachsignal berechnet sich nach (3.9) zu

$$\hat{S}_\mu = \frac{\eta_\mu + \sqrt{\eta_\mu^2 + 2(1 + \eta_\mu)\frac{\eta_\mu}{\gamma_\mu}}}{2(1 + \eta_\mu)} X_\mu = W_m \cdot X_\mu. \quad (4.19)$$

Mit der Berücksichtigung des spectral floors ergibt sich für die spektralen Gewichte

$$W_m = \max \left\{ \frac{\eta_\mu + \sqrt{\eta_\mu^2 + 2(1 + \eta_\mu)\frac{\eta_\mu}{\gamma_\mu}}}{2(1 + \eta_\mu)}, W_{\min} \right\} \quad m \in \{r, l\}. \quad (4.20)$$

Für die ML-Schätzung der Sprachleistung $\sigma_{S,\mu}^2$ hat sich in Kapitel 3 die Gleichung

$$\hat{S}_\mu = \sqrt{1 - \frac{\sigma_{N,\mu}^2}{|X_\mu|^2}} X_\mu \quad (4.21)$$

ergeben. Durch die getroffenen Annahmen über die DFT-Koeffizienten (Mittelwertfreiheit, voneinander unabhängiger Real- und Imaginärteil) lässt sich die Varianz $\sigma_{N,\mu}^2$ auch als $\mathcal{E}\{|N_\mu|^2\}$ schreiben. Mit der Berücksichtigung des spectral floors berechnen sich die spektralen Gewichte damit zu

$$W_m = \sqrt{\max \left\{ 1 - \frac{1}{\gamma_\mu}, W_{\min} \right\}} \quad m \in \{r, l\}, \quad (4.22)$$

wobei der $\max\{\cdot, \cdot\}$ -Operator hier bereits unter der Wurzel angewendet wird, um zu verhindern, dass das Argument der Wurzel negativ wird. Für die ML-Schätzung der Amplitude A_μ nach (4.23) ist es nicht notwendig die spektralen Gewichte nach unten hin zu begrenzen, da diese sowieso nicht kleiner als 0,5 werden können. Es ist lediglich dafür zu sorgen, dass das Argument der Wurzel nicht negativ wird, wodurch sich die spektralen Gewichte hierfür zu

$$W_m = \frac{1}{2} + \frac{1}{2} \sqrt{\max\left\{1 - \frac{1}{\gamma_\mu}, 0\right\}} \quad m \in \{r, l\} \quad (4.23)$$

berechnen.

Für die MMSE-Schätzung der Amplitude A_μ nach (3.31) wird ebenso der *spectral floor* eingeführt. Zusätzlich ist es hier jedoch sinnvoll, die annehmbaren Werte der spektralen Gewichte auch nach oben hin zu begrenzen. Durch die Limitierung nach oben hin auf den Wert 1 ergibt sich für deren Berechnung

$$W_m = \min\left\{1, \max\left\{\frac{\sqrt{v_\mu}}{\gamma_\mu} \Gamma(1, 5) F_1(-0.5, 1, -v_\mu), W_{\min}\right\}\right\} \quad m \in \{r, l\}. \quad (4.24)$$

Für alle zu implementierenden Gleichungen zur Berechnung der spektralen Gewichte W_m gilt, dass sich für den Ansatz zweier getrennter Filter die spektralen Gewichte W_r und W_l ergeben. Bei Verwendung eines gemeinsamen Filters ist für die Berechnung des a-priori und des a-posteriori SNRs das gemeinsame Leistungsdichtespektrum der gefilterten Mikrofon-signale $\hat{\Phi}_{V_r V_r} + \hat{\Phi}_{V_l V_l}$ einzusetzen. Dementsprechend ergeben sich auch gemeinsame a-priori und a-posteriori SNRs und daraus gemeinsame spektrale Gewichte W .

(4.9), (4.11), (4.18), (4.20), (4.22), (4.23) und (4.24) wurden schließlich implementiert und das vorhandene Matlab-Programm zur Störreduktion damit erweitert. Zur Übersicht sind diese zu evaluierenden Algorithmen noch einmal in den Tabellen 4.1 und 4.2 zusammengestellt.

Filtertyp	Berechnungsvorschrift
Wiener-Filter (LDS)	$W_m = \max \left\{ 1 - \beta \frac{\hat{\Phi}_{\hat{N}\hat{N}}}{\hat{\Phi}_{V_m V_m}}, W_{\min} \right\}$
Wiener-Filter (SNR)	$W_m = \max \left\{ \frac{\eta_\mu}{\eta_\mu + 1}, W_{\min} \right\}$
MAP-Schätzer	$W_m = \max \left\{ \frac{\eta_\mu + \sqrt{\eta_\mu^2 + 2(1 + \eta_\mu) \frac{\eta_\mu}{\gamma_\mu}}}{2(1 + \eta_\mu)}, W_{\min} \right\}$
ML-Schätzer ($\sigma_{S,\mu}^2$)	$W_m = \sqrt{\max \left\{ 1 - \frac{1}{\gamma_\mu}, W_{\min} \right\}}$
ML-Schätzer (A_μ)	$W_m = \frac{1}{2} + \frac{1}{2} \sqrt{\max \left\{ 1 - \frac{1}{\gamma_\mu}, 0 \right\}}$
MMSE-Schätzer	$W_m = \min \left\{ 1, \max \left\{ \frac{\sqrt{v_\mu}}{\gamma_\mu} \Gamma(1, 5) F_1(-0.5, 1, -v_\mu), W_{\min} \right\} \right\}$
mit $\gamma_\mu[k] = \frac{\hat{\Phi}_{V_m V_m}[k]}{\hat{\Phi}_{\hat{N}\hat{N}}[k]}$, $\hat{\eta}_\mu[k] = \alpha_\eta \cdot \frac{ \hat{S}_\mu[k-1] ^2}{\hat{\Phi}_{\hat{N}\hat{N}}[k-1]} + (1 - \alpha_\eta) \cdot \max\{\gamma_\mu[k] - 1, 0\}$	

Tabelle 4.1: Berechnungsvorschriften für den getrennten Filteransatz

Filtertyp	Berechnungsvorschrift
Wiener-Filter (LDS)	$W = \max \left\{ 1 - \beta \frac{\hat{\Phi}_{\hat{N}\hat{N}}}{\hat{\Phi}_{V_r V_r} + \hat{\Phi}_{V_l V_l}}, W_{\min} \right\}$
Wiener-Filter (SNR)	$W = \max \left\{ \frac{\eta_\mu}{\eta_\mu + 1}, W_{\min} \right\}$
MAP-Schätzer	$W = \max \left\{ \frac{\eta_\mu + \sqrt{\eta_\mu^2 + 2(1 + \eta_\mu) \frac{\eta_\mu}{\gamma_\mu}}}{2(1 + \eta_\mu)}, W_{\min} \right\}$
ML-Schätzer ($\sigma_{S,\mu}^2$)	$W = \sqrt{\max \left\{ 1 - \frac{1}{\gamma_\mu}, W_{\min} \right\}}$
ML-Schätzer (A_μ)	$W = \frac{1}{2} + \frac{1}{2} \sqrt{\max \left\{ 1 - \frac{1}{\gamma_\mu}, 0 \right\}}$
MMSE-Schätzer	$W = \min \left\{ 1, \max \left\{ \frac{\sqrt{v_\mu}}{\gamma_\mu} \Gamma(1, 5) F_1(-0.5, 1, -v_\mu), W_{\min} \right\} \right\}$
mit $\gamma_\mu[k] = \frac{\hat{\Phi}_{V_r V_r}[k] + \hat{\Phi}_{V_l V_l}[k]}{\hat{\Phi}_{\hat{N}\hat{N}}[k]}$, $\hat{\eta}_\mu[k] = \alpha_\eta \cdot \frac{ \hat{S}_\mu[k-1] ^2}{\hat{\Phi}_{\hat{N}\hat{N}}[k-1]} + (1 - \alpha_\eta) \cdot \max\{\gamma_\mu[k] - 1, 0\}$ und $ \hat{S}_\mu[k-1] ^2 = \hat{S}_{\mu,r}[k-1] ^2 + \hat{S}_{\mu,l}[k-1] ^2$	

Tabelle 4.2: Berechnungsvorschriften für den gemeinsamen Filteransatz

Ausgehend von den Leistungsdichtespektren der gefilterten Mikrofonsignale $\hat{\Phi}_{V_m V_m}$ ($m \in \{r, l\}$) und der gemeinsamen Störschätzung $\hat{\Phi}_{\hat{N}\hat{N}}$ listen die beiden Tabellen die Berechnungsvorschriften für die spektralen Gewichte und die dafür notwendigen Gleichungen für die Berechnung des a-priori und a-posteriori SNR auf, jeweils für die Ansätze getrennter und gemeinsamer Filter.

Um die verschiedenen Algorithmen miteinander vergleichen zu können, ist es notwendig einige Indikatoren einzuführen, mit denen sich die Leistungsfähigkeiten der verschiedenen Algorithmen messen lassen. Diese sind im Einzelnen:

- Die Änderung des *signal to interference ratio* (SIR) vom Eingangssignal zum Signal am Ausgang des Systems: Dieser Wert wird als SIR gain bezeichnet und ist über das SIR des Eingangssignals und das SIR des Ausgangssignals definiert als

$$\text{SIR gain} = \text{SIR}|_{\text{out}} - \text{SIR}|_{\text{in}}. \quad (4.25)$$

Die beiden SIRs berechnen sich hierbei zu

$$\begin{aligned} \text{SIR}|_{\text{out}} &= 10 \cdot \log_{10} \frac{\sigma_s^2|_{\text{out}}}{\sigma_n^2|_{\text{out}}} \quad \text{dB} \\ \text{SIR}|_{\text{in}} &= 10 \cdot \log_{10} \frac{\sigma_s^2|_{\text{in}}}{\sigma_n^2|_{\text{in}}} \quad \text{dB}, \end{aligned} \quad (4.26)$$

wobei σ_s^2 und σ_n^2 die (Langzeit-) Leistungen des Nutzsignals beziehungsweise der Stör- und Interferenzsignale darstellen. Durch die Angaben „in“ und „out“ wird festgelegt, ob es sich um das Signal am Eingang oder am Ausgang des Systems handelt.

- Das nach Sprachverständlichkeit (*speech intelligibility*) gewichtete SIR (*SIR gain SI weighted*): Auch der *SIR gain SI weighted* ist über die entsprechenden Werte am Ein- und am Ausgang des Systems definiert:

$$\text{SIR gain SI weighted} = \text{SIR SI weighted}|_{\text{out}} - \text{SIR SI weighted}|_{\text{in}} \quad (4.27)$$

Hierbei berechnen sich die beiden gewichteten SIRs zu

$$\begin{aligned} \text{SIR SI weighted}|_{\text{out}} &= \int_{f_L}^{f_R} w(f) \left[10 \cdot \log_{10} \frac{\Phi_{SS}(f)|_{\text{out}}}{\Phi_{NN}(f)|_{\text{out}}} \right]_{-15}^{+15} df \quad \text{dB} \\ \text{SIR SI weighted}|_{\text{in}} &= \int_{f_L}^{f_R} w(f) \left[10 \cdot \log_{10} \frac{\Phi_{SS}(f)|_{\text{in}}}{\Phi_{NN}(f)|_{\text{in}}} \right]_{-15}^{+15} df \quad \text{dB}. \end{aligned} \quad (4.28)$$

Die Notation $[\cdot]_{-15}^{+15}$ steht dabei für eine Begrenzung auf Werte zwischen -15 und $+15$ und $w(f)$ ist ein frequenzabhängiger Gewichtungsfaktor. Die Bestimmung der Integralgrenzen f_R und f_L und der Gewichtsfunktion $w(f)$ ist in [1] nachzulesen. $\Phi_{SS}(f)$ und $\Phi_{NN}(f)$ bezeichnen die Langzeit-Leistungsdichtespektren des Nutzsignals beziehungsweise der Stör- und Interferenzsignale.

- Die Geräuschreduktion (*noise reduction*, NR): Sie ist definiert als das logarithmierte Verhältnis zwischen der Störleistung am Eingang und der Störleistung am Ausgang des Systems:

$$\text{NR} = 10 \cdot \log_{10} \frac{\sigma_n^2|_{\text{in}}}{\sigma_n^2|_{\text{out}}} \quad \text{dB} \quad (4.29)$$

- Die Sprachverzerrung (*speech distortion*, SD): Der Wert der Sprachverzerrung berechnet sich als das logarithmierte Verhältnis zwischen der Leistung der Differenz des Nutzsignals am Eingang und am Ausgang zur Leistung des Nutzsignals am Eingang zu

$$\text{SD} = 10 \cdot \log_{10} \frac{\mathcal{E} \{ (s[k - \kappa]|_{\text{in}} - s[k]|_{\text{out}})^2 \}}{\sigma_s^2|_{\text{in}}} \quad \text{dB}. \quad (4.30)$$

Dabei entspricht κ der Gruppenlaufzeit des Systems.

In der Auswertung der Ergebnisse in dieser Arbeit wird von den hier eingeführten Größen jeweils der arithmetische Mittelwert zwischen dem rechten und dem linken Kanal betrachtet, um die Diagramme nicht zu unübersichtlich werden zu lassen.

Die Ergebnisse der verschiedenen Simulationen mit einer Filterlänge des Prototyp-Tiefpassfilters von $L = 512$ sind in den Abbildungen 4.6 bis 4.13 dargestellt.

	getrennte Filter				gem. Filter	
	λ	α_η	β	\mathbf{W}_{\min}	β	\mathbf{W}_{\min}
Wiener-Filter (LDS)	0,85	-	0,95	0,2	1,1	0,15
Wiener-Filter (SNR)	0,85	0,85	-	0,25	-	0,24
MAP-Schätzer	0,85	0	-	0,25	-	0,15
ML-Schätzer ($\sigma_{\mathbf{S},\mu}^2$)	0,85	0,85	-	0,12	-	0,03
ML-Schätzer (\mathbf{A}_μ)	0,85	-	-	-	-	-
MMSE-Schätzer	0,85	0	-	0,25	-	0,12

Tabelle 4.3: Verwendete Parameter bei einer Filterlänge von $L = 512$

Dabei wird für alle Simulationen der Wert $\lambda = 0,85$ gewählt, um die Leistungsdichtespektren zu schätzen. Eine Übersicht über alle für diese Simulation gewählten Parameter zeigt Tabelle 4.3. Die Anzahl der Quellen wird, beginnend bei zwei, schrittweise bis zu vier gesteigert, und schließlich wird bei einem 4-Quellen-Szenario zusätzlich diffuses Hintergrundrauschen addiert. Die gewählten Parameter nach Tabelle 4.3 sind für die verschiedenen Geräuschsituationen jeweils konstant.

Beim Vergleich der Ergebnisse schneidet der bisher verwendete Wiener-Filter Ansatz, basierend auf den Leistungsdichtespektren, bezogen auf die Geräuschreduktion, fast am besten von allen untersuchten Algorithmen ab. Lediglich das Wiener-Filter basierend auf dem a-priori und dem a-posteriori SNR liefert hier noch bessere Werte. Allerdings ist auch zu sehen, dass die Sprachverzerrung bei diesen beiden Ansätzen etwas höher ist als bei den anderen Schätzern. Einzig und allein der ML-Schätzer der Amplitude des Sprachsignals sticht negativ ins Auge. Mit einer maximalen Geräuschreduktion von nur 4,21 dB liegt er deutlich unter allen anderen Ansätzen. Der Grund hierfür lässt sich aber recht leicht in der Berechnungsvorschrift für die DFT-Koeffizienten des Nutzsymbols nach (4.23) erkennen. Wie bereits in Kapitel 3 erwähnt, werden die Koeffizienten des Mikro-

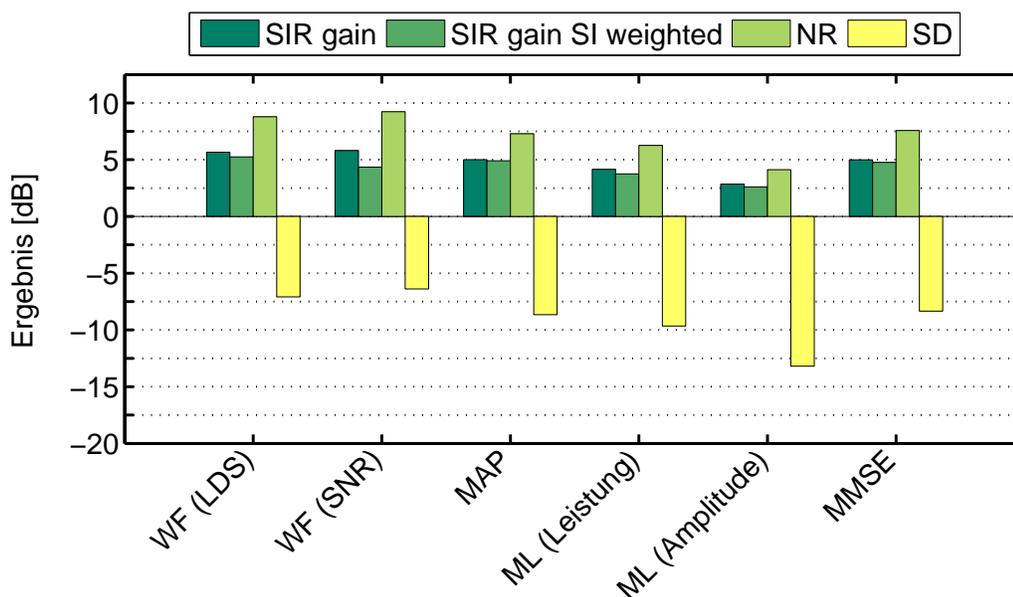


Abbildung 4.6: $L = 512$, 1 Störquelle, getrennter Filteransatz

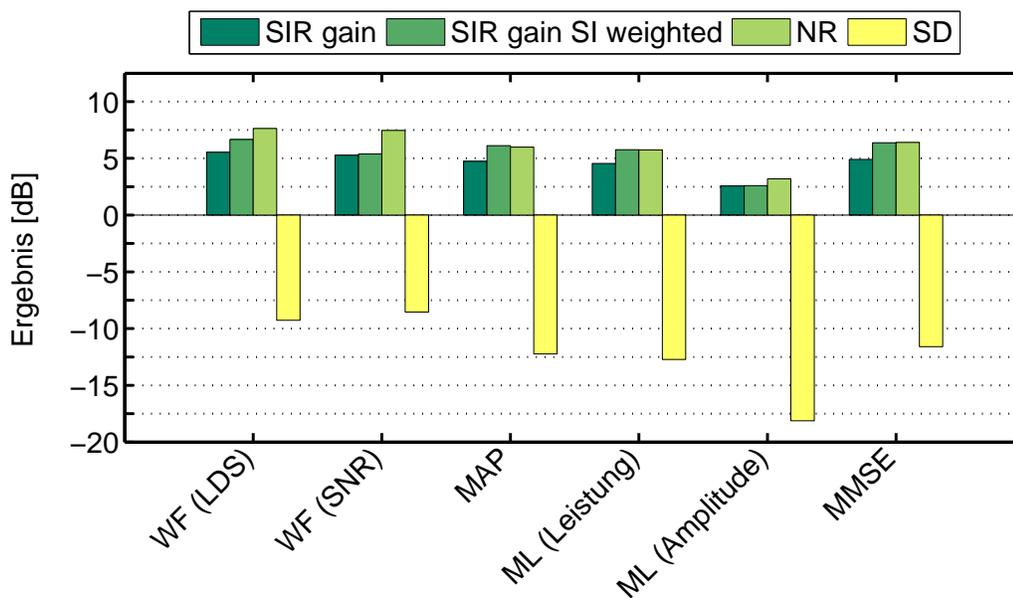


Abbildung 4.7: $L = 512$, 1 Störquelle, gemeinsamer Filteransatz

fonsignals immer mindestens mit 0,5 gewichtet. Selbst für den Fall, dass das Mikrofonsignal kein Sprachsignal der Nutzquelle enthält, könnte das Signal also nicht vollständig unterdrückt werden. Mit diesem Schätzer ist eine maximale Ge-

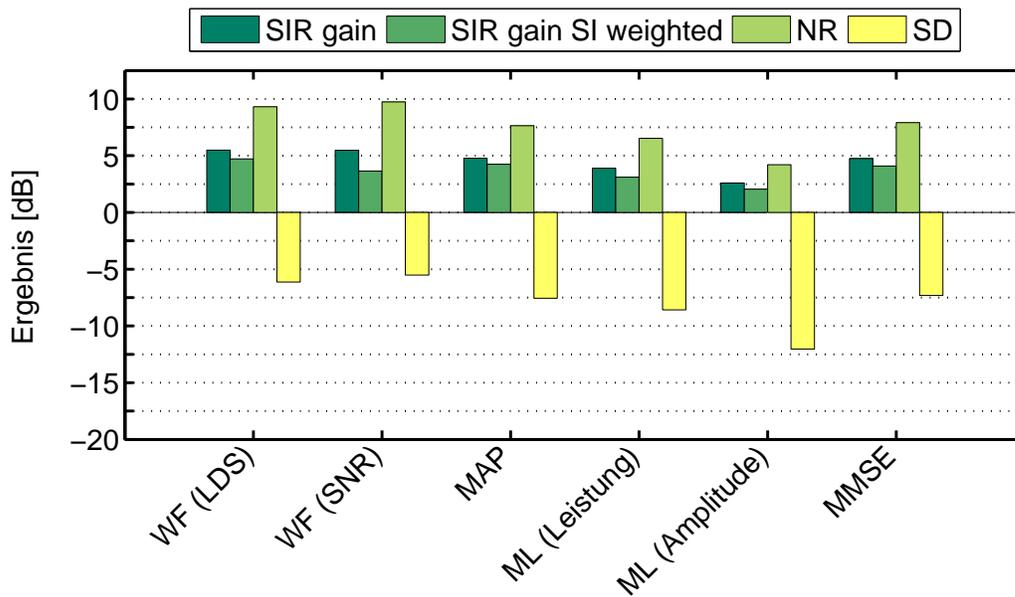


Abbildung 4.8: $L = 512$, 2 Störquellen, getrennter Filteransatz

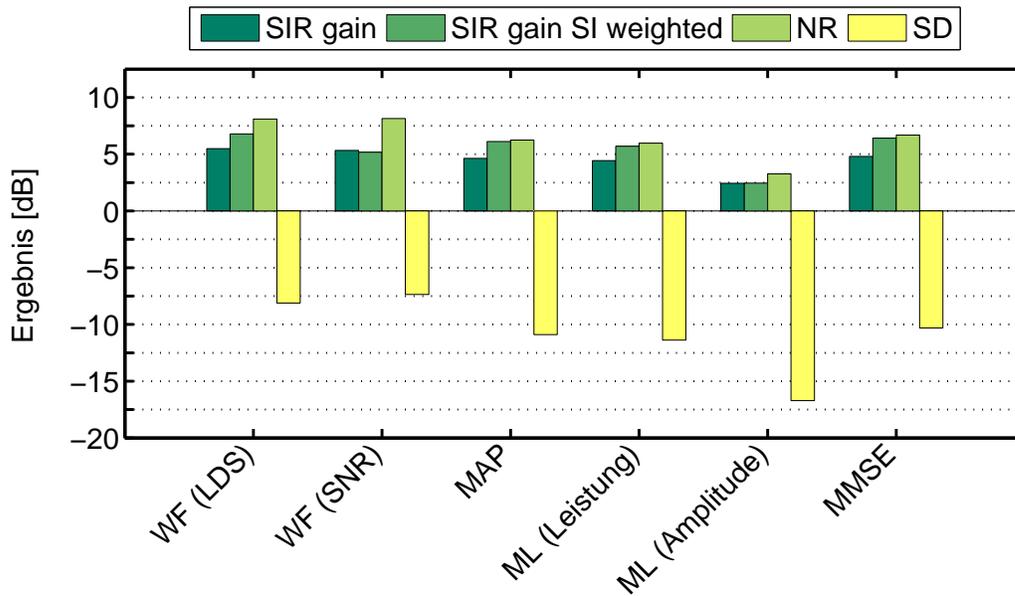


Abbildung 4.9: $L = 512$, 2 Störquellen, gemeinsamer Filteransatz

räuschreduktion von 6 dB erreichbar. Allein aus den erzielten Zahlenwerten den besten Algorithmus auszuwählen ist also nicht sehr sinnvoll.

Durch informelle Hörtests lassen sich die Unterschiede besser erkennen. Hier-

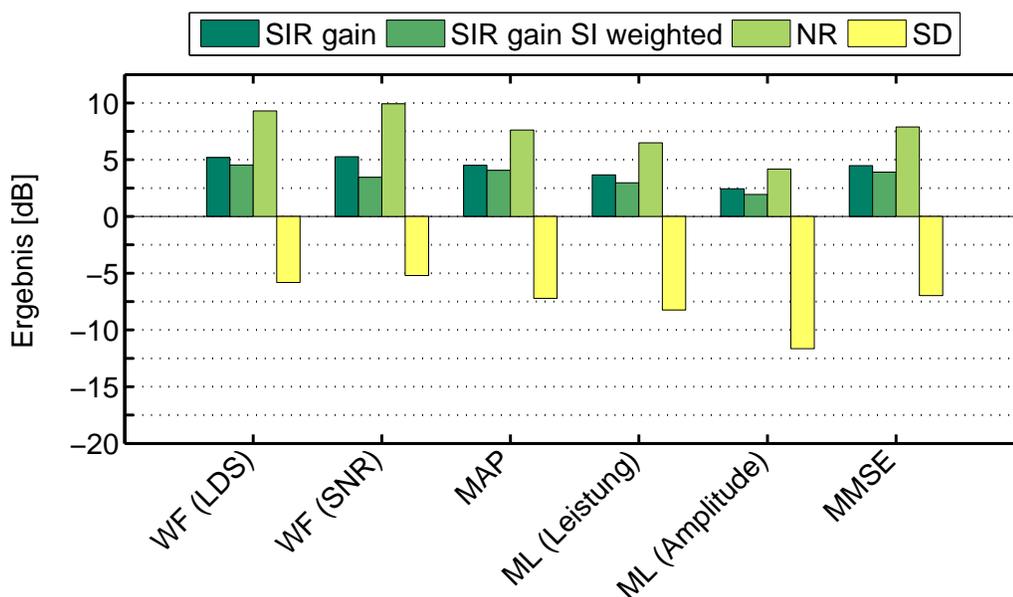


Abbildung 4.10: $L = 512$, 3 Störquellen, getrennter Filteransatz

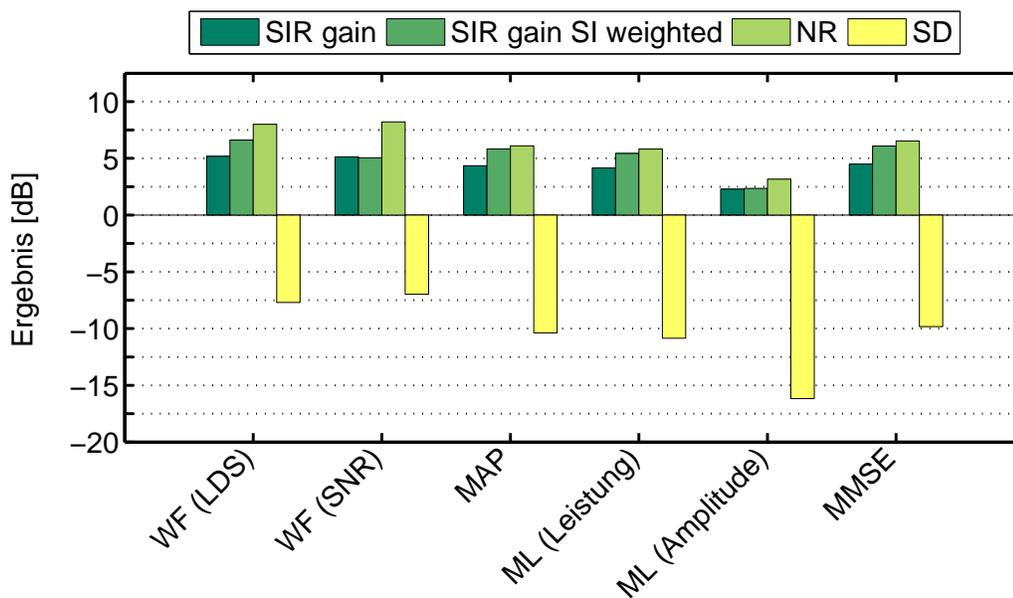


Abbildung 4.11: $L = 512$, 3 Störquellen, gemeinsamer Filteransatz

bei stellt sich heraus, dass die beiden Wiener-Filter Ansätze einen leichten Tiefpasscharakter besitzen. Ihre Ausgangssignale klingen etwas dumpf und unnatürlich. Bei den in dieser Arbeit vorgestellten Schätzmethoden ist dies nicht der

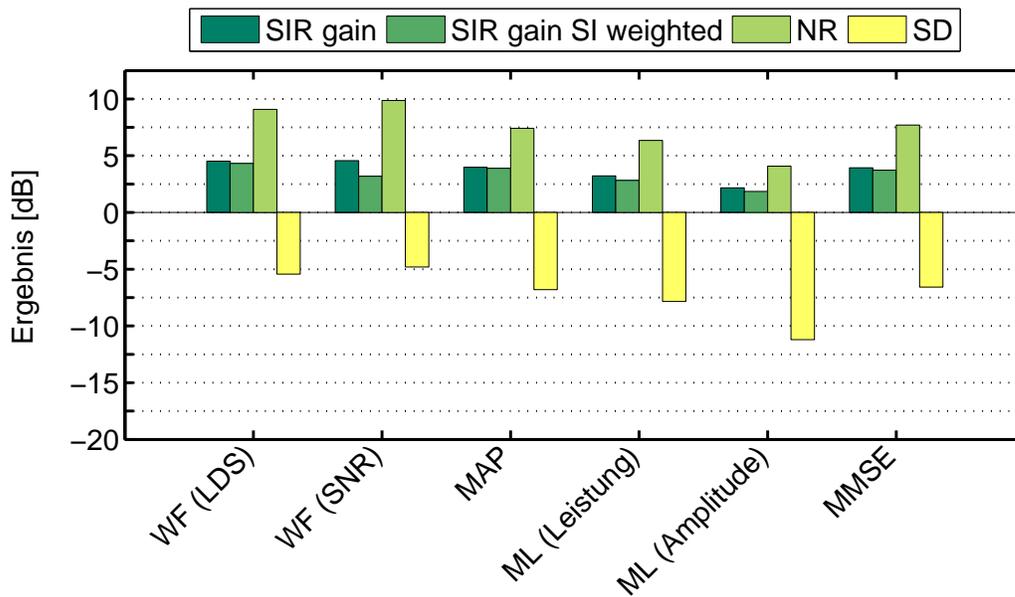


Abbildung 4.12: $L = 512$, 3 Störquellen und Hintergrundrauschen, getrennter Filteransatz

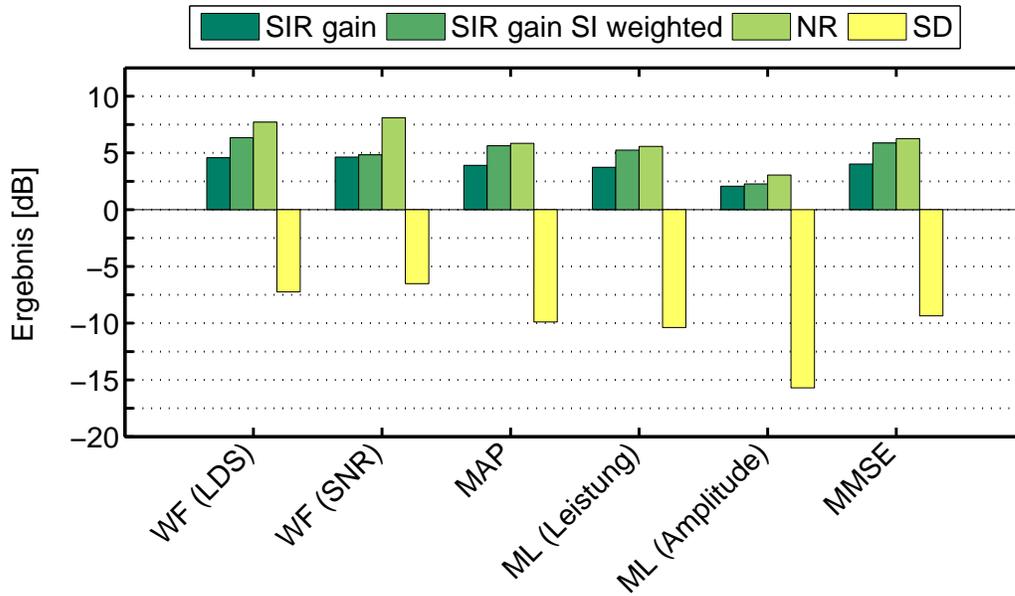


Abbildung 4.13: $L = 512$, 3 Störquellen und Hintergrundrauschen, gemeinsamer Filteransatz

Fall. Die niedrige Geräuschunterdrückung des ML-Schätzers der Amplitude ist zwar im Ausgangssignal deutlich hörbar, weshalb die Sprachverständlichkeit bei diesem Ansatz nicht sehr gut ist, die drei anderen Schätzer (MAP-Schätzer, ML-Schätzer der Leistung und MMSE-Schätzer) liefern aber eine deutlich höhere Geräuschunterdrückung und gleichzeitig ein natürlicher klingendes Ausgangssignal, als es die beiden Wiener-Filter tun. Zwischen dem MAP-Schätzer, dem ML-Schätzer der Sprachleistung und dem MMSE-Schätzer lässt sich kaum ein Unterschied feststellen. Diese drei Methoden liefern eine gute Störreduktion bei sehr wenig Nutzsignalverzerrung.

Schließlich ist auch noch ein Unterschied zwischen der Verwendung eines gemeinsamen Filters für beide Kanäle und der Verwendung zweier getrennter Filter festzustellen. Für alle Simulationen fällt auf, dass die Sprachverzerrung bei Verwendung eines gemeinsamen Filters deutlich niedriger ist als bei Verwendung getrennter Filter. Gleichzeitig ist zwar auch die Störreduktion etwas niedriger, insgesamt allerdings lässt sich durch die Verwendung eines gemeinsamen Filters der Wert des *SIR gain SI weighted* beträchtlich erhöhen, sodass im Allgemeinen das Konzept des gemeinsamen Filters zu bevorzugen ist.

Darüber hinaus soll noch erwähnt werden, dass alle in dieser Arbeit betrachteten Verfahren linearphasige Systeme sind. Wie bereits in [8] diskutiert, werden diese Verfahren in einer binauralen Hörgeräteanwendung nie die interauralen Zeitunterschiede beeinflussen (solange beide Filter die gleiche Gruppenlaufzeit haben). Außerdem hat ein gemeinsamer Filteransatz den weiteren Vorteil, dass interaurale Pegelunterschiede ebenfalls erhalten bleiben (siehe [8]), weshalb in einer binauralen Hörgeräteanwendung ein gemeinsamer Ansatz bevorzugt verwendet werden sollte. Schließlich wird aufgrund der einfachen Berechnungsvorschrift der ML-Schätzer der Nutzsignalleistung (auch bekannt als spektrale Subtraktion) Vorteile in der Implementierung bieten, weshalb sich dieses Konzept als das beste für Hörgeräteanwendungen herausgestellt hat.

Kapitel 5

Zusammenfassung und Ausblick

Im Verlauf dieser Arbeit wurden - aufbauend auf der Bayes'schen Schätztheorie - verschiedene Schätzer zur Störgeräuschunterdrückung in Hörgeräteanwendungen hergeleitet und deren Leistungsfähigkeit schließlich in Matlabsimulationen untersucht. Um konkrete Signale schätzen zu können, wurden hierfür idealisierte Annahmen über die Wahrscheinlichkeitsdichten der beteiligten Signale getroffen. Auf der Grundlage dieser angenommenen Dichten war es schließlich möglich, Gleichungen für die Berechnung der Schätzwerte der verschiedenen untersuchten Schätzer aufzustellen. Die Erweiterung des bereits bestehenden Simulationsprogramms mit diesen Schätzern und die durchgeführten Simulationen führten schließlich zu Erkenntnissen über die Leistungsfähigkeit der gewonnenen Schätzer.

Zusammenfassend lässt sich feststellen, dass fast alle der in den Simulationen eingesetzten Schätzer gute Ergebnisse liefern. Nur der ML-Schätzer zur Schätzung der Amplitude schneidet sehr schlecht ab. Der MAP-Schätzer, der ML-Schätzer der Sprachleistung und der MMSE-Schätzer bieten sich aber durchaus als Alternative zum bisher verwendeten Wiener-Filter an. Wenn auch die Störreduktion etwas niedriger ist, so hört sich das Ausgangssignal dieser drei Schätzer doch besser an, weil es natürlicher klingt als das mit dem Wiener-Filter gefilterte Signal. Beide Wiener-Filter Ansätze besitzen einen leichten Tiefpasscharakter, sodass sich die Ausgangssignale etwas dumpf anhören. Unter den drei alternativen Schätzern

ist schließlich der ML-Schätzer der Nutzsignalleistung aufgrund seiner einfachen Berechnungsvorschrift bei der Implementierung für eine Hörgeräteanwendung zu bevorzugen.

Alle untersuchten Schätzer basieren auf dem gaußschen Modell für die DFT-Koeffizienten des Sprachsignals. Für weitere Untersuchungen wäre es von Interesse auch Schätzer basierend auf einer supergaußförmigen Wahrscheinlichkeitsdichte, die eigentlich besser zu Sprachsignalen passt, herzuleiten und deren Leistungsfähigkeit in Simulationen zu überprüfen. Ein weiterer Schritt wäre schließlich die Untersuchung von Mehrkanalverfahren zur Geräuschunterdrückung.

Anhang A

Simulationsergebnisse für Filterlängen 256 und 128

Im Folgenden sind die Ergebnisse der Simulationen zu finden, bei denen Prototyp-Tiefpassfilter der Längen 256 und 128 verwendet wurden.

A.1 Ergebnisse für Filterlänge 256

Für die Simulationen mit einer Prototyp-Tiefpassfilterlänge von $L = 256$ wurden die Parameter nach Tabelle A.1 verwendet.

	getrennte Filter				gem. Filter	
	λ	α_η	β	\mathbf{W}_{\min}	β	\mathbf{W}_{\min}
Wiener-Filter (LDS)	0,88	-	0,95	0,2	1,1	0,15
Wiener-Filter (SNR)	0,88	0,85	-	0,25	-	0,24
MAP-Schätzer	0,88	0	-	0,25	-	0,15
ML-Schätzer ($\sigma_{\mathcal{S},\mu}^2$)	0,88	0,85	-	0,12	-	0,03
ML-Schätzer (\mathbf{A}_μ)	0,88	-	-	-	-	-
MMSE-Schätzer	0,88	0	-	0,25	-	0,12

Tabelle A.1: Verwendete Parameter bei einer Filterlänge von $L = 256$

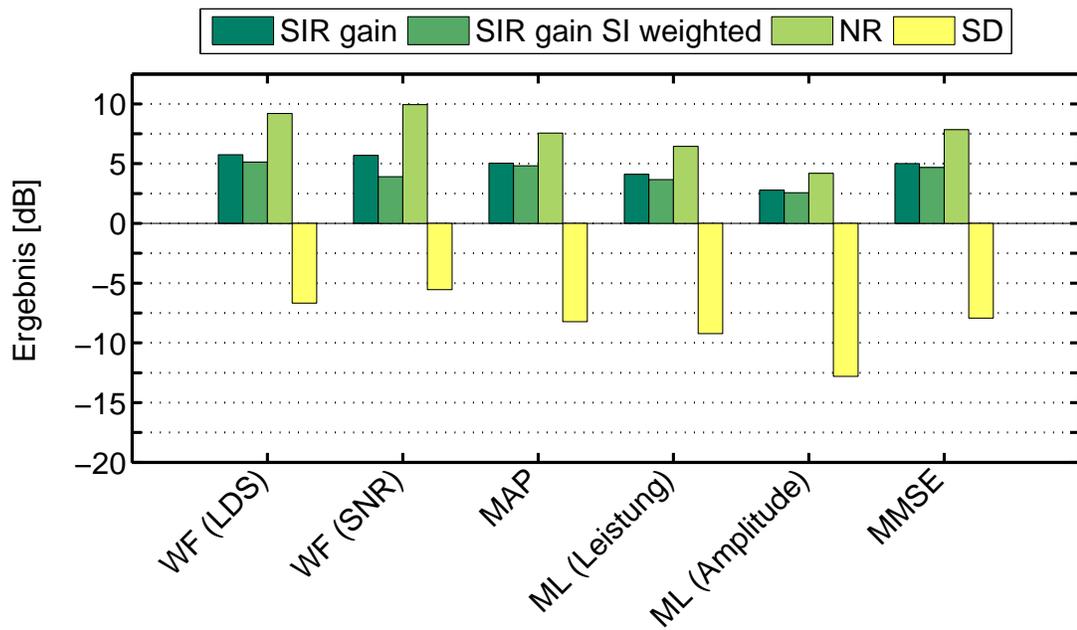


Abbildung A.1: $L = 256$, 1 Störquelle, getrennter Filteransatz

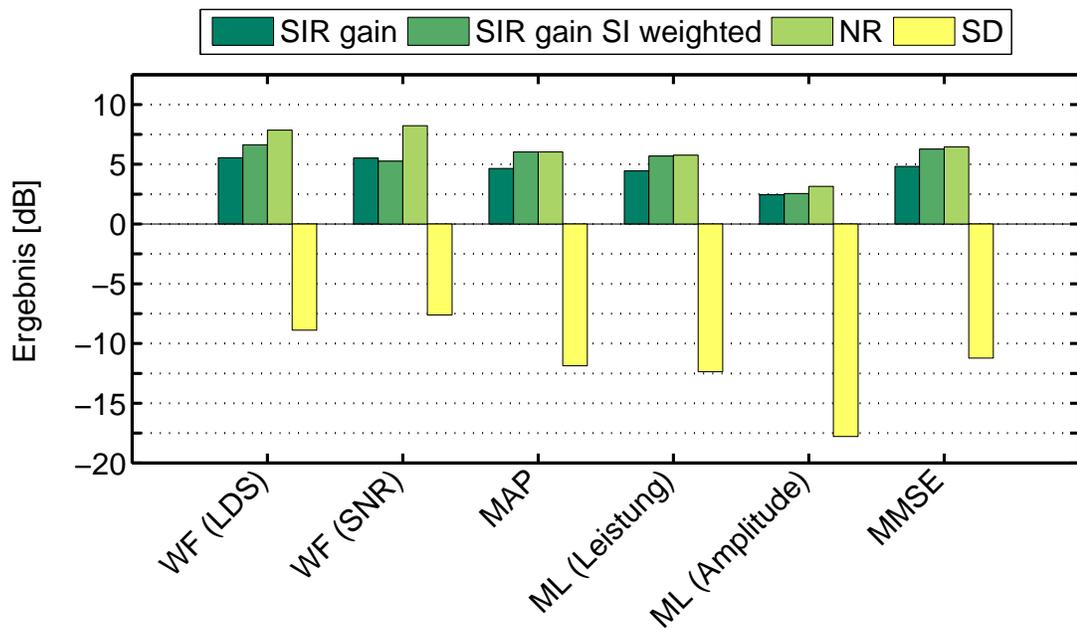


Abbildung A.2: $L = 256$, 1 Störquelle, gemeinsamer Filteransatz

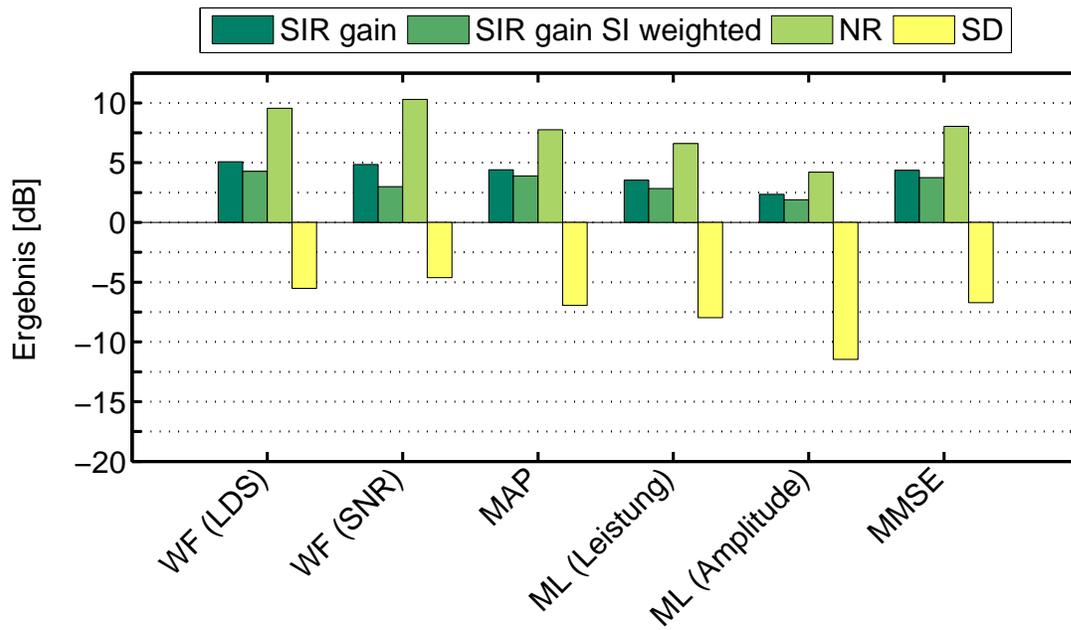


Abbildung A.3: $L = 256$, 2 Störquellen, getrennter Filteransatz

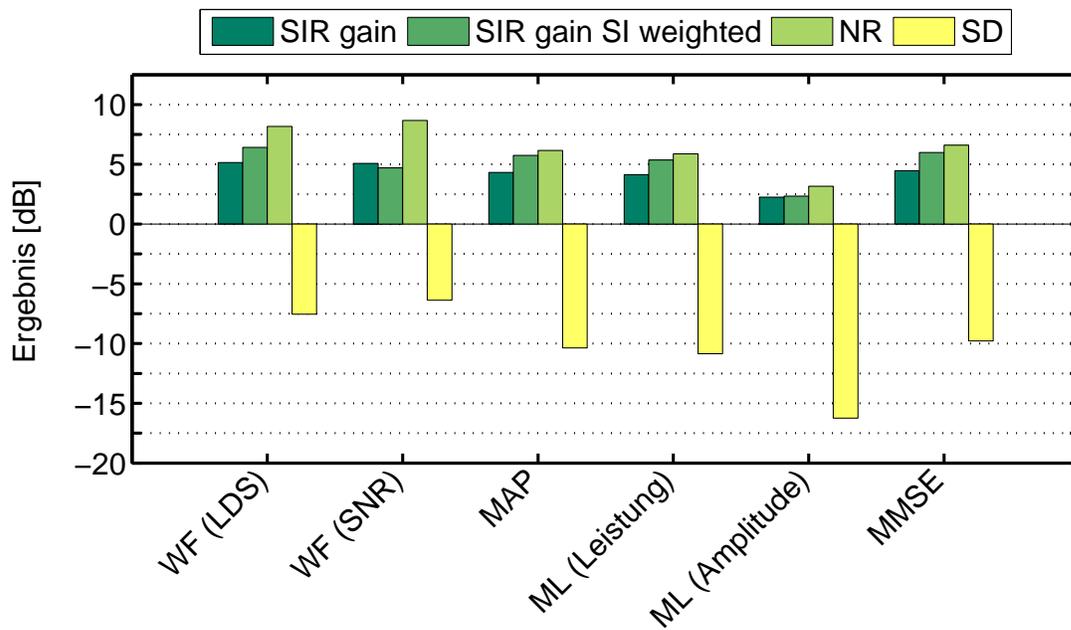


Abbildung A.4: $L = 256$, 2 Störquellen, gemeinsamer Filteransatz

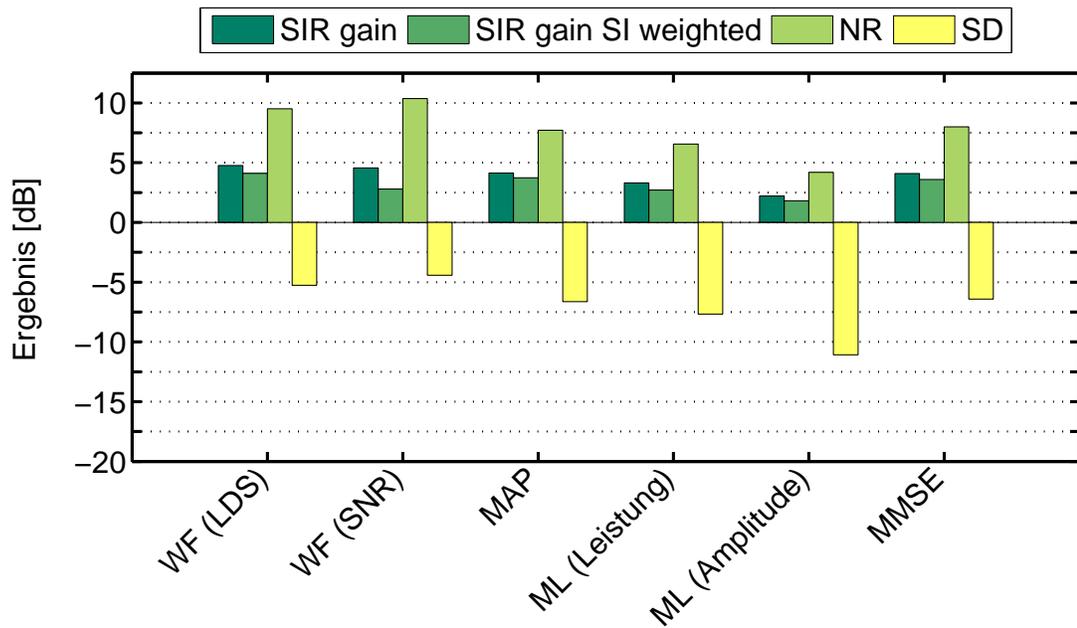


Abbildung A.5: $L = 256$, 3 Störquellen, getrennter Filteransatz

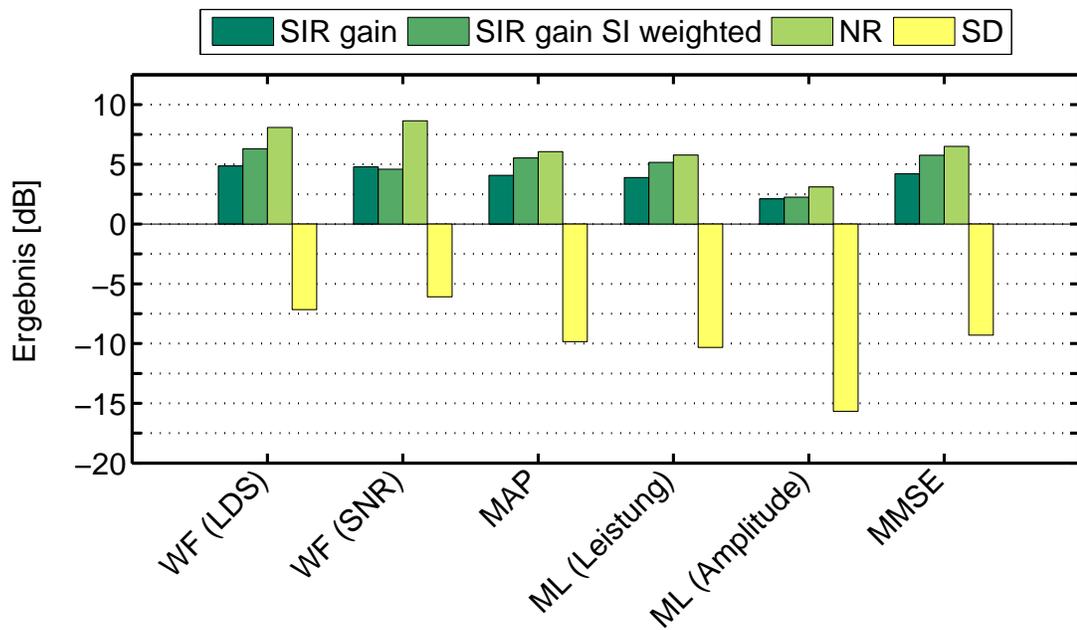


Abbildung A.6: $L = 256$, 3 Störquellen, gemeinsamer Filteransatz

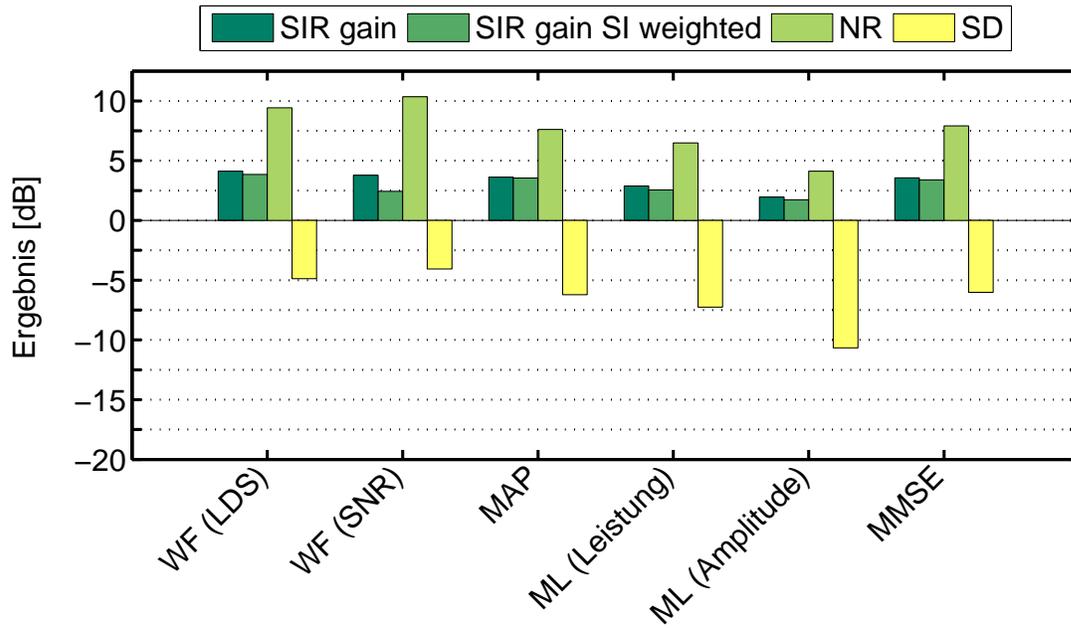


Abbildung A.7: $L = 256$, 3 Störquellen und Hintergrundrauschen, getrennter Filteransatz

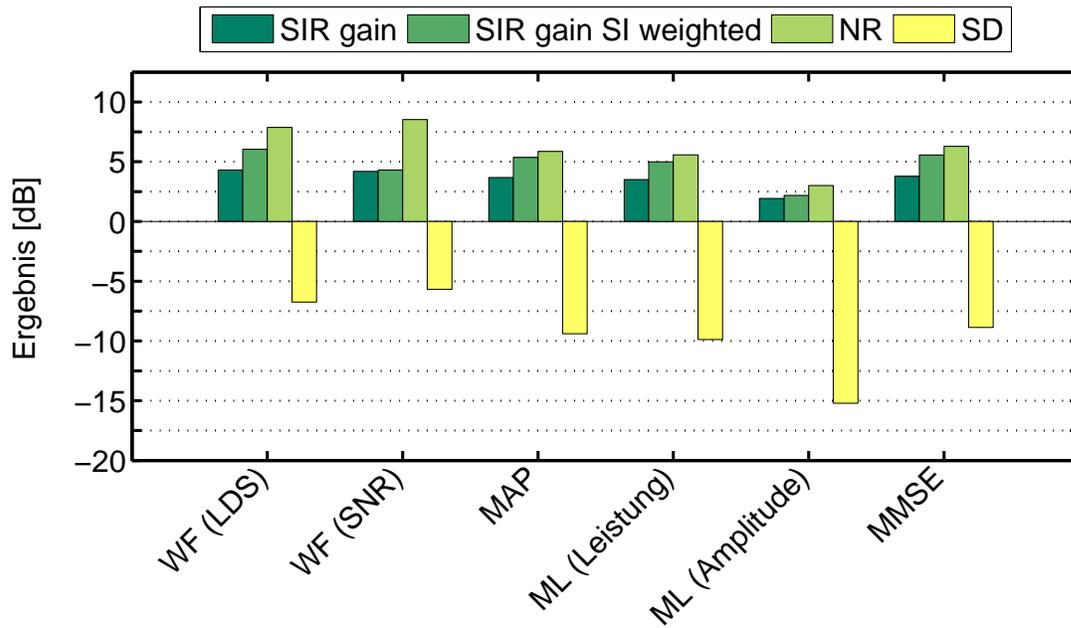


Abbildung A.8: $L = 256$, 3 Störquellen und Hintergrundrauschen, gemeinsamer Filteransatz

A.2 Ergebnisse für Filterlänge 128

Die für die Simulationen mit einer Filterlänge von $L = 128$ verwendeten Parameter sind in Tabelle A.2 zusammengefasst.

	getrennte Filter				gem. Filter	
	λ	α_η	β	W_{\min}	β	W_{\min}
Wiener-Filter (LDS)	0,92	-	0,95	0,2	1,1	0,15
Wiener-Filter (SNR)	0,92	0,85	-	0,25	-	0,24
MAP-Schätzer	0,92	0	-	0,25	-	0,15
ML-Schätzer ($\sigma_{S,\mu}^2$)	0,92	0,85	-	0,12	-	0,03
ML-Schätzer (A_μ)	0,92	-	-	-	-	-
MMSE-Schätzer	0,92	0	-	0,25	-	0,12

Tabelle A.2: Verwendete Parameter bei einer Filterlänge von $L = 128$

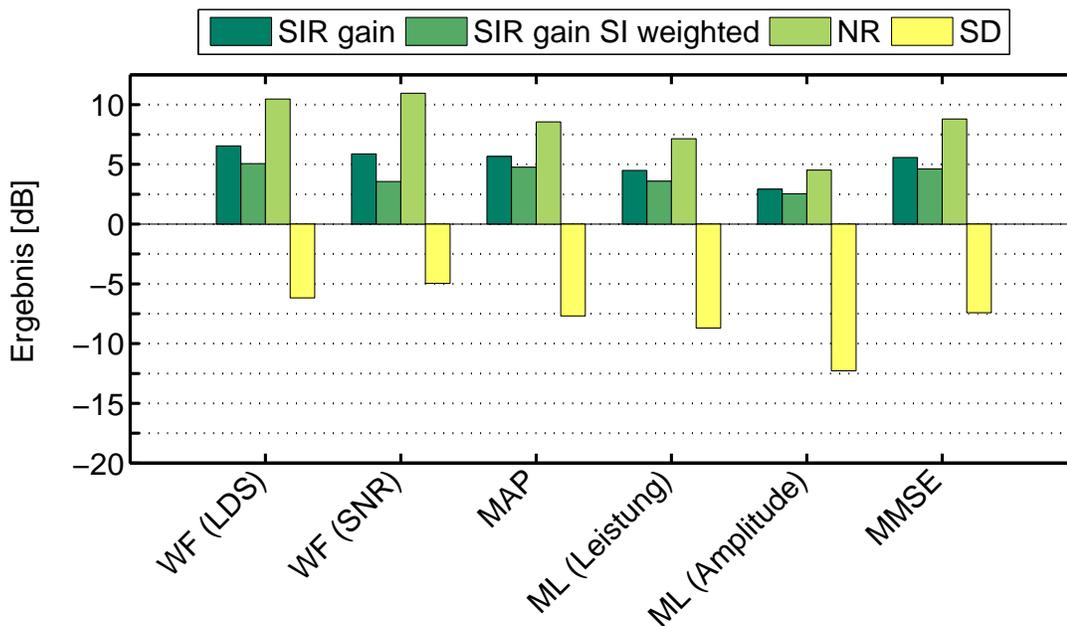


Abbildung A.9: $L = 128$, 1 Störquelle, getrennter Filteransatz

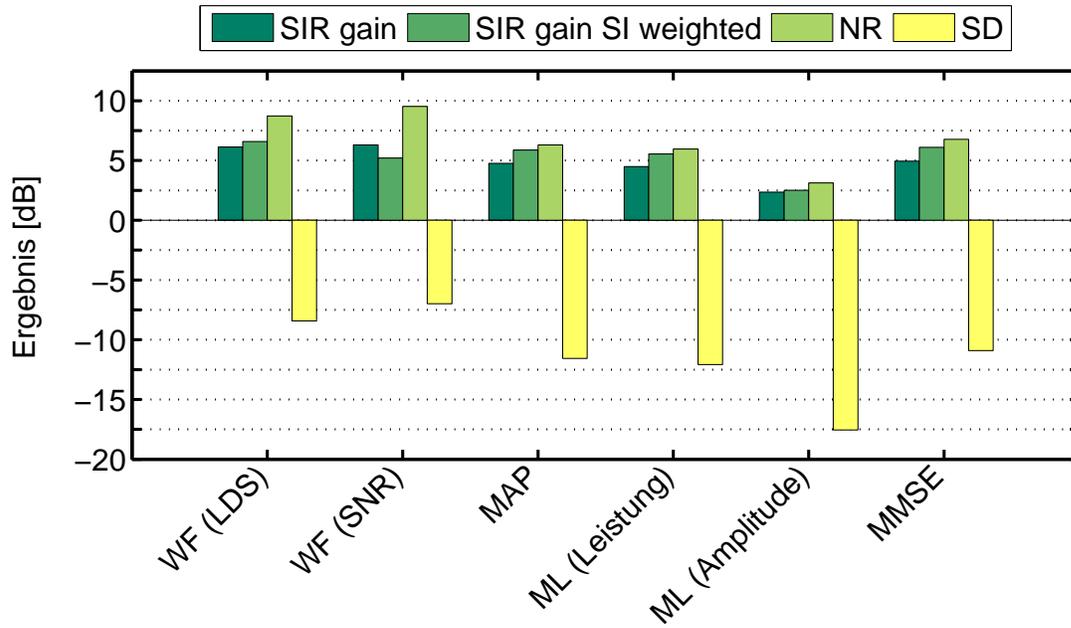


Abbildung A.10: $L = 128$, 1 Störquelle, gemeinsamer Filteransatz

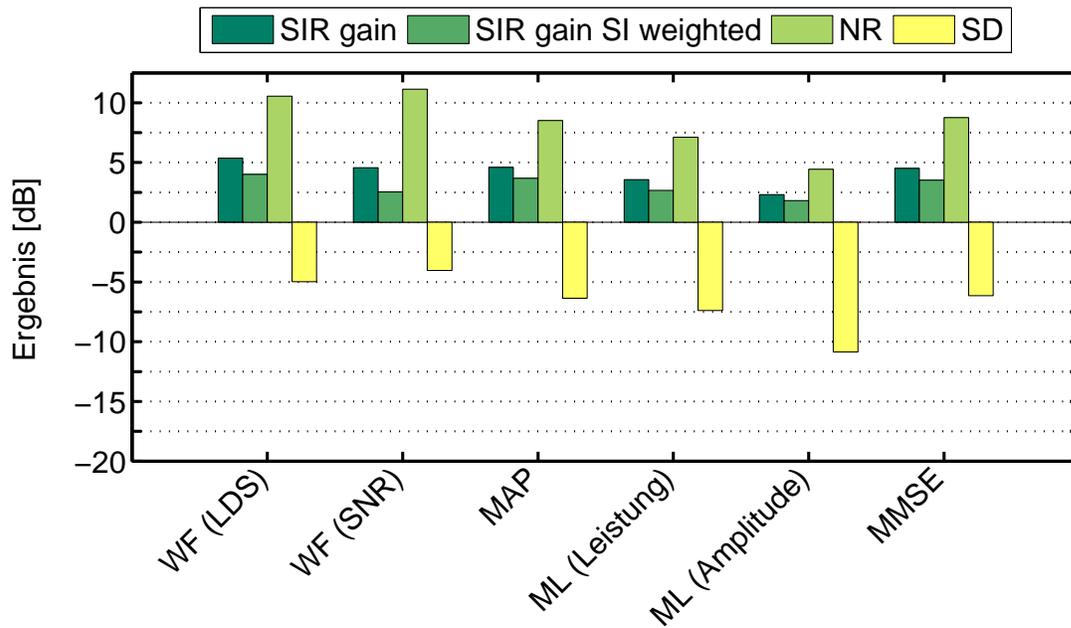


Abbildung A.11: $L = 128$, 2 Störquellen, getrennter Filteransatz

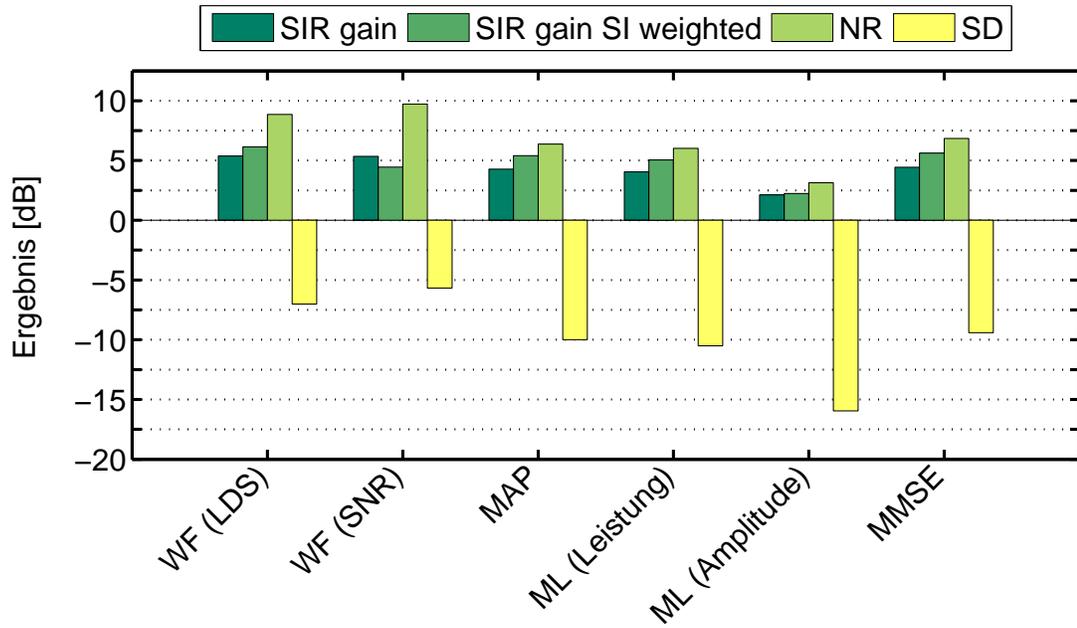


Abbildung A.12: $L = 128$, 2 Störquellen, gemeinsamer Filteransatz

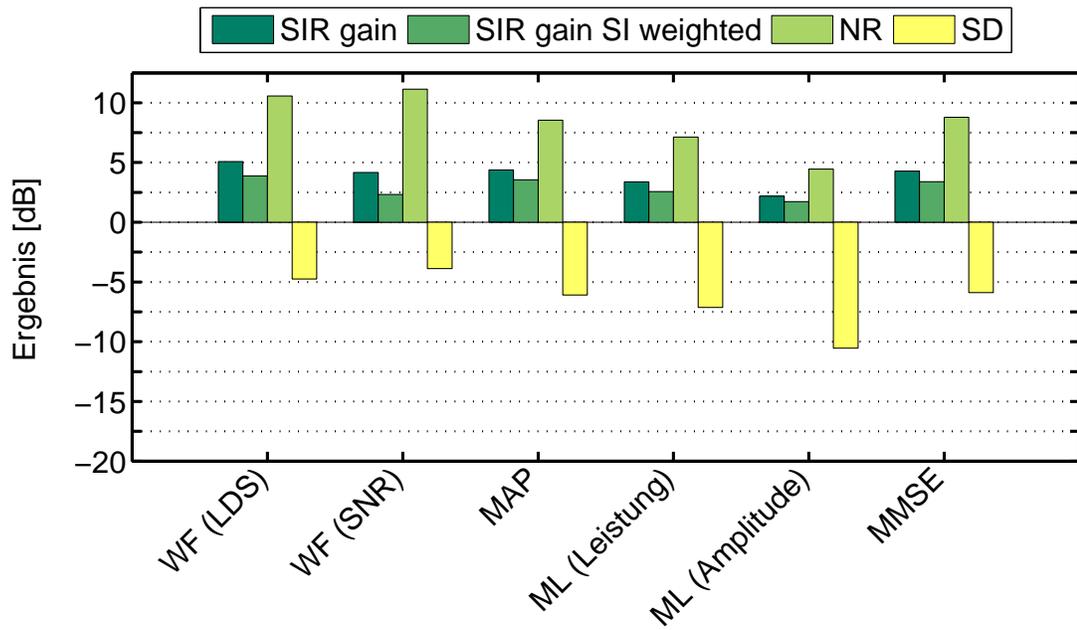


Abbildung A.13: $L = 128$, 3 Störquellen, getrennter Filteransatz

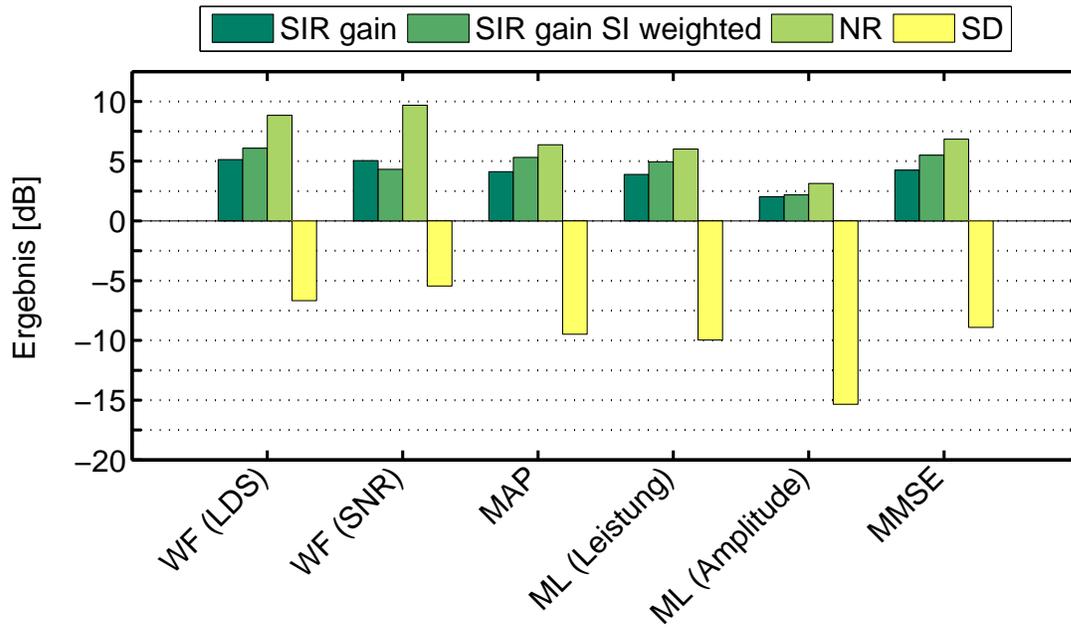


Abbildung A.14: $L = 128$, 3 Störquellen, gemeinsamer Filteransatz

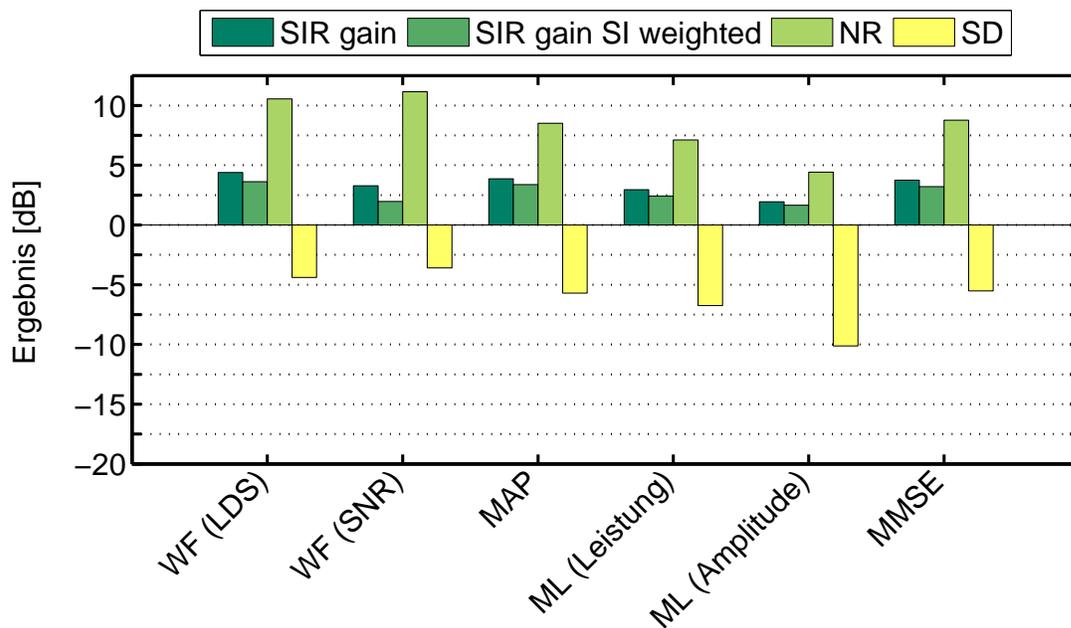


Abbildung A.15: $L = 128$, 3 Störquellen und Hintergrundrauschen, getrennter Filteransatz

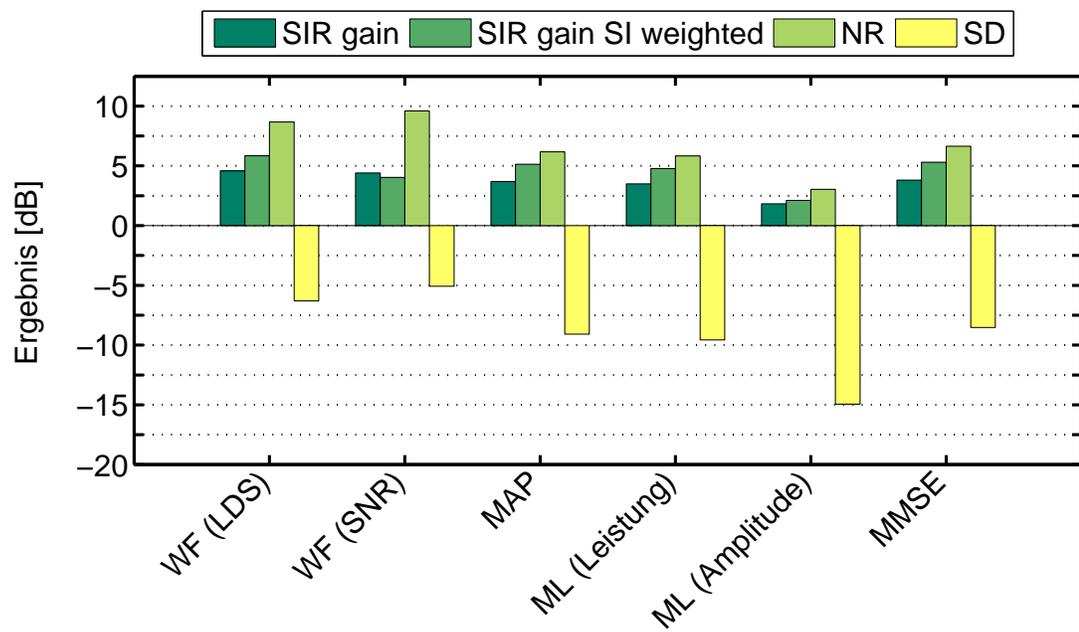


Abbildung A.16: $L = 128$, 3 Störquellen und Hintergrundrauschen, gemeinsamer Filteransatz

Anhang B

Akronyme, Notation und Symbole

Akronyme

DFT	diskrete Fouriertransformation
IDFT	inverse diskrete Fouriertransformation
LDS	Leistungsdichtespektrum
MAP	maximum a-posteriori
MAVE	minimum mean absolute value of error
ML	maximum likelihood
MMSE	minimum mean squared error
NR	noise reduction
SD	speech distortion
SI	speech intelligibility
SIR	signal to interference ratio
SNR	signal to noise ratio
WF	Wiener-Filter

Notation und Symbole

Allgemeines

$\hat{\cdot}$	Schätzwert einer Größe
\cdot^T	transponierte Matrix
argmax_x	liefert das Argument x , für das eine Funktion maximal wird
argmin_x	liefert das Argument x , für das eine Funktion minimal wird
$\mathcal{E}\{\cdot\}$	Erwartungswert-Operator
$F_1(\cdot, \cdot, \cdot)$	hypergeometrische Funktion [2]
$\Gamma(\cdot)$	Gammafunktion: $\Gamma(y) = \int_0^\infty x^{y-1} e^{-x} dx$

Grundlagen der Bayes'schen Schätzung

$\ \cdot\ $	euklidische Norm eines Vektors: $\ \mathbf{x}\ = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$
$\mathcal{C}(\cdot, \cdot)$	Bayes'sche Kostenfunktion
$f_{\Phi}(\phi)$	a-priori Wahrscheinlichkeitsdichte
$f_{\Phi \mathbf{X}}(\phi \mathbf{x})$	a-posteriori Wahrscheinlichkeitsdichte
$f_{\mathbf{X} \Phi}(\mathbf{x} \phi)$	Likelihood-Funktion
$\mathcal{R}(\cdot)$	Bayes'sche Risikofunktion
\mathcal{V}	a-priori Ergebnisraum
\mathbf{x}	determinierter Beobachtungsvektor
\mathbf{X}	zufälliger Beobachtungsvektor
\mathcal{X}	a-posteriori Ergebnisraum
$\delta(\cdot)$	Einheitsimpuls
ϕ	determinierter skalarer Parameter
ϕ	determinierter Parametervektor
Φ	zufälliger Parametervektor

Bayes'sche Schätzung zur Störreduktion

$ \cdot $	Betrags-Operator: $ z = \sqrt{\text{Re}\{z\} + \text{Im}\{z\}}$
A_μ	Betrag von S_μ
$\exp(x) / e^x$	Exponentialfunktion
$f_X(x)$	Wahrscheinlichkeitsdichte
$f_{X,Y}(x, y)$	Verbunddichte
$f_{X Y}(x y)$	bedingte Wahrscheinlichkeitsdichte
$I_0(\cdot)$	modifizierte Besselfunktion nullter Ordnung
$\text{Im}\{\cdot\}$	Imaginärteil-Operator
j	imaginäre Einheit: $j = \sqrt{-1}$
k	diskreter Zeitindex
M	Transformationslänge der DFT
$n[k]$	Störsignal im Zeitbereich
N_μ	μ -te Komponente des Spektrums von $n[k]$
R_μ	Betrag von X_μ
$\text{Re}\{\cdot\}$	Realteil-Operator
$s[k]$	Nutzsignal im Zeitbereich
S_μ	μ -te Komponente des Spektrums von $s[k]$
$x[k]$	Mikrofonsignal im Zeitbereich
$X_\mu / X_\mu[k]$	μ -te Komponente des Spektrums von $x[k]$
α_μ	Phase von S_μ
γ_μ	a-posteriori SNR
η_μ	a-priori SNR
θ_μ	Phase von X_μ
μ	diskreter Frequenzindex
$\sigma_{N,\mu}^2$	Varianz von N_μ

$\sigma_{S,\mu}^2$	Varianz von S_μ
$\sigma_{X_\mu}^2$	Varianz von X_μ
Ω_μ	normierte Mittenfrequenz eines Frequenzbins

Evaluierung

B	Blockiermatrix
$\text{diag}\{\cdot\}$	erzeugt eine Diagonalmatrix mit den Elementen des Argumentvektors auf der Hauptdiagonalen
$h[k]$	Prototyp-Tiefpassfilter
$\mathbf{h}_r/\mathbf{h}_l$	sogenannte <i>Head-Related Impulse Responses</i> (HRIRs) [6]
H	Matrix des akustischen Mischsystems
H_{qr}/H_{ql}	Übertragungsfunktion von der q -ten Punktquelle zum rechten/linken Mikrofon
$\log_{10} \cdot$	Logarithmus zur Basis 10
L	Länge des Prototyp-Tiefpassfilters
$\max\{\cdot, \cdot\}$	Maximum der beiden Argumente
$\min\{\cdot, \cdot\}$	Minimum der beiden Argumente
n	Blockindex
\mathbf{n}_b	Vektor mit den Spektren des Hintergrundrauschens
\hat{N}	Spektrum der gemeinsamen Störschätzung
$N_{b,r}/N_{b,l}$	Spektrum des Hintergrundrauschens am rechten/linken Mikrofon
Y_r/Y_l	Spektrum des geräuschreduzierten rechten/linken Mikrofonsignals
s	Vektor mit den Spektren der Quellensignale
S_q	Spektrum einer Punktquelle
W	Matrix zur Geräuschreduktion

W_{\min}	untere Grenze für W_r/W_l
W_r/W_l	spektrale Gewichte zur Geräuschreduktion
V_r/V_l	Spektrum des gefilterten rechten/linken Mikrofonsignals
\mathbf{x}	Vektor mit den Spektren der Mikrofonsignale
X_r/X_l	Spektrum des rechten/linken Mikrofonsignals
\mathbf{y}	Vektor mit den Spektren der geräuschreduzierten Signale
α_η	Gewichtungsfaktor für Schätzung des a-priori SNR
β	Gewichtungsfaktor für Berechnung von W_r/W_l
κ	Gruppenlaufzeit des Systems
λ	Gewichtungsfaktor für Schätzung der Leistungsdichtespektren
σ_n^2	(Langzeit-) Leistung der Stör- und Interferenzsignale
σ_s^2	(Langzeit-) Leistung des Nutzsignals
$\Phi_{NN}(f)$	Langzeit-Leistungsdichtespektrum der Stör- und Interferenzsignale
$\Phi_{SS}(f)$	Langzeit-Leistungsdichtespektrum des Nutzsignals
$\Phi_{XX}/\Phi_{XX}[k]$	Leistungsdichtespektrum

Literaturverzeichnis

- [1] ANSI-S3.5, *American national standard methods for the calculation of the speech intelligibility index*, American National Standards Institute, New York, 1997.
- [2] Yariv Ephraim, David Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” in *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, No. 6, Dec. 1984, pp. 1109-1121.
- [3] Walter Kellermann, *Stochastische Prozesse*, Skript zur Vorlesung, Lehrstuhl für Multimediakommunikation und Signalverarbeitung, FAU Erlangen-Nürnberg, Sommersemester 2009.
- [4] Walter Kellermann, *Digital Signal Processing*, Skript zur Vorlesung, Lehrstuhl für Multimediakommunikation und Signalverarbeitung, FAU Erlangen-Nürnberg, Wintersemester 2009/2010.
- [5] Bert-Uwe Köhler, *Konzepte der statistischen Signalverarbeitung*, Springer-Verlag Berlin Heidelberg, 2005.
- [6] Klaus Reindl, Yuanhang Zheng, Walter Kellermann, “Speech Enhancement for Binaural Hearing Aids based on Blind Source Separation,” in Proc. *4th International Symposium on Communications, Control, and Signal Processing (ISCCSP)*, Limassol, Cyprus, Mar. 2010.

- [7] Klaus Reindl, Yuanhang Zheng, Walter Kellermann, “Blind Binaural Signal Extraction for Hearing Aids“, Report on Project Year 1, Lehrstuhl für Multimediakommunikation und Signalverarbeitung, FAU Erlangen-Nürnberg, 2010.
- [8] Klaus Reindl, Yuanhang Zheng, Walter Kellermann, “Analysis of two generic Wiener filtering concepts for binaural speech enhancement in hearing aids“ in Proc. *European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010.
- [9] Peter Vary, Rainer Martin, *Digital Speech Transmission - Enhancement, Coding, and Error Concealment*, John Wiley & Sons, 2006.
- [10] Yuanhang Zheng, Klaus Reindl, Walter Kellermann, “BSS for Improved Interference Estimation for Blind Speech Signal Extraction with Two Microphones“ in Proc. *3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Aruba, Dutch Antilles, Dec. 2009.