

Friedrich-Alexander-Universität Erlangen-Nürnberg

**Lehrstuhl für Multimediakommunikation und
Signalverarbeitung**

Prof. Dr.-Ing. Walter Kellermann

Bachelorarbeit

**Evaluation eines auf kopfbezogenen
Übertragungsfunktionen basierenden
robusten Polynombeamformers für
robotisches Hören**

Marcel Möglich

Oktober 2015

Betreuer: Hendrik Barfuss

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	IV
Abkürzungsverzeichnis	V
Formelzeichen	VII
1 Einleitung	1
2 Grundlagen	3
2.1 Signalmodell	3
2.2 Beamforming	5
2.2.1 Charakteristiken von Beamformern	6
2.2.2 Polynombeamforming	9
3 Robuster auf kopfbezogene Übertragungsfunktionen basierender Polynombeamformer	12
3.1 Robustes Superdirektives Beamforming unter Freifeldannahme	13
3.2 Robustes Polynombeamforming unter Freifeldannahme	15
3.3 Robustes auf kopfbezogene Übertragungsfunktionen basierendes Beamforming	17
3.4 Robuster auf kopfbezogene Übertragungsfunktionen basierender Polynombeamformer	18

4	Evaluierung des robusten auf kopfbezogene Übertragungsfunktionen basierenden Polynombeamformers	20
4.1	Aufbau des Experiments	20
4.2	Evaluation der Konfiguration des robusten auf kopfbezogene Übertragungsfunktionen basierenden Polynombeamformers	22
4.3	Signalunabhängige Evaluierung	25
4.4	Signalabhängige Evaluierung	32
5	Zusammenfassung und Ausblick	35
	Abbildungsverzeichnis	36
	Tabellenverzeichnis	38
	Literaturverzeichnis	40

Kurzfassung

Um eine natürliche Interaktion zwischen Mensch und Roboter zu ermöglichen, sollte verbale Kommunikation unterstützt werden. Da sich der Benutzer im Abstand von einigen Metern zu dem Roboter befinden könnte und weitere Sprecher gleichzeitig aktiv sein könnten, ist es die Aufgabe des Roboters das Nutzsignal des Benutzers zu extrahieren. Das Konzept von Beamforming bietet eine Möglichkeit Signale räumlich zu unterscheiden und Signale, aus einer bestimmten Richtung, zu verstärken. Da sich in diesem Kontext die Richtung aus der das Nutzsignal kommt häufig ändern kann, wird in dieser Arbeit ein Polynombeamformer untersucht, welcher in der Lage ist seine Blickrichtung dynamisch anzupassen. Dieses Beamformerkonzept wird für den Roboter NAO [1] evaluiert. Da der Kopf des Roboters das Schallfeld beeinflusst, werden bei dem Entwurf des Beamformers kopfbezogene Übertragungsfunktionen (Head-Related Transfer Function (HRTF)) [2] berücksichtigt. Die Performanz des HRTF-basierten Polynombeamformers wird mit dem HRTF-basierten Beamformer [2] und dem freifeldbasierten Polynombeamformer [3] verglichen. Dabei stellt sich heraus, dass die HRTF-basierten Ansätze deutlich bessere Ergebnisse erzielen als der freifeldbasierte Polynombeamformer. Ausgewertet an den Stützstellen des HRTF-basierten Polynombeamformers unterscheidet sich die Performanz dieses Entwurfs nicht von der des HRTF-basierten Beamformers. Für Blickrichtungen, welche nicht als Stützstellen enthalten sind, werden geringfügig schlechtere Resultate für den HRTF-basierten Polynombeamformer erzielt als für den HRTF-basierten Beamformer.

Abkürzungsverzeichnis

DOA Direction Of Arrival

HRTF Head-Related Transfer Function

RLSFI Robust Least-Squares Frequency-Invariant

MSE Mean Square Error

WNG White Noise Gain

FSB Filter-and-Sum Beamformer

FIR Finite Impulse Response

DTFT Discrete-Time Fourier Transformation

LTI Linear Time Invariant

FSU Filter-and-Sum Unit

PPF Polynomial Postfilter

FI Frequency Invariant

LS Least-Squares

MATLAB MATrix LABoratory, The MathWorks

RIRs Room Impulse Responses

IDTFT Inverse Discrete-Time Fourier Transformation

fwSegSNR frequency-weighted Segmental Signal-to-Noise Ratio

WER Word Error Rate

Formelzeichen

- * Faltung
- $(\cdot)^T$ Transposition
- $(\cdot)^H$ Hermischer Operator
- \otimes Kronecker-Produkt

Kapitel 1

Einleitung

Ein großes Problem im Bereich der Robotertechnologie ist es, eine zuverlässige Kommunikation zwischen Roboter und Benutzer zu gewährleisten. Diese Bedingung muss für verschiedene Situationen erfüllt werden, zum Beispiel wenn der Benutzer aus unterschiedlichen Entfernungen spricht und mehrere Sprecher gleichzeitig aktiv sind. Bezüglich der Kommunikation besteht die Aufgabe des Roboters darin, das Nutzsignal von Störsignalen zu unterscheiden und zu extrahieren. Wenn dies gelingt, verbessert sich die Performanz in Bezug auf die Spracherkennung deutlich. Diese Arbeit befasst sich mit Beamforming-Verfahren zur Signalextraktion. Durch räumliche Unterscheidung, versucht der Beamformer Signale aus einer bestimmten Richtung zu verstärken und gleichzeitig Signale aus allen anderen Richtungen zu unterdrücken. Ein Beamformer besteht aus einer Anordnung von mehreren Mikrofonen. Da diese Mikrofone in einem bestimmten Abstand zueinander stehen, haben empfangene Schallsignale eine relative Zeitverschiebung, abhängig davon, aus welcher Richtung sie kommen (Direction Of Arrival (DOA)). Diese Eigenschaft wird ausgenutzt, um die Signale räumlich voneinander zu unterscheiden. Die meisten Beamformer-Modelle setzen die Freifeldausbreitung voraus. Diese besagt, dass Signale nicht von der Umgebung beeinflusst werden. Diese Annahme trifft für diese Arbeit nicht zu. Die Mikrofone des Roboters NAO [1], welcher in dieser Arbeit betrachtet wird, befinden sich an dessen Kopf. Um die Verzerrung der Signale durch diesen auszugleichen, werden gemessene, auf dem Kopf

des Roboters basierende Übertragungsfunktionen (HRTF) [2] bei der Anwendung des Beamformers miteinbezogen. Um Signale aus einer bestimmten Richtung zu verstärken, müssen die Parameter des Beamformers für diese Richtung dementsprechend optimiert werden. Diese Berechnung kann für Beamformer relativ viel Zeit in Anspruch nehmen. Da ein Roboter auf schnell wechselnde Situationen reagieren muss, wird eine dynamische Lösung gesucht, die eine akzeptable Performanz liefert. In dieser Arbeit wird ein Polynombeamformer betrachtet, der seine Blickrichtung durch Interpolation von vorher berechneten Einstellungen schnell ändern kann.

Diese Arbeit ist wie folgt strukturiert: In Kapitel 2 werden die benötigten technischen Grundlagen erläutert. Es wird ein Signalmodell anhand eines Filter-and-Sum Beamformer (FSB) [4] vorgestellt, welches die in dieser Arbeit verwendeten technische Notation enthält. Zudem werden die Grundlagen von Beamforming erläutert, zu denen typische Charakteristiken von Beamformern und das Prinzip von Polynombeamformern zählen. Kapitel 3 befasst sich mit der Entwicklung, die zu dem hier betrachteten Experiment führte. Zuerst wird der freifeldbasierte, robuste, nach kleinstem quadratischen Fehler entworfene, frequenzinvariante (Robust Least-Squares Frequency-Invariant (RLSFI)) Beamformer [5] und der freifeldbasierte RLSFI-Polynombeamformer [3] vorgestellt. Die Konzeption des HRTF-basierten RLSFI-Beamformers [2] wird danach präsentiert. Im Anschluss wird die Problembeschreibung des HRTF-basierten RLSFI-Polynombeamformers dargestellt. Die Ergebnisse dieses Versuches werden in Kapitel 4 ausgewertet. Neben dem genauen Versuchsaufbau, wird eine generelle Aussage zu dem Verhalten des HRTF-basierten RLSFI-Polynombeamformers getroffen. Außerdem wird ein direkter Vergleich zwischen diesem und dem freifeldbasierten RLSFI-Polynombeamformer sowie dem HRTF-basierten RLSFI-Beamformer dargestellt. Kapitel 5 schildert eine Zusammenfassung der Ergebnisse und gibt einen Ausblick auf zukünftige Arbeit.

Kapitel 2

Grundlagen

In diesem Kapitel werden technische Grundlagen zum besseren Verständnis der Arbeit erläutert. Es wird ein Signalmodell vorgestellt, welches die angewandte, technische Notation einführt. In diesem Teil soll außerdem ein grundlegendes Verständnis von Beamformermodellen und -charakteristiken vermittelt werden, welche in dem durchgeführten Experiment Anwendung finden. Neben den betrachteten Qualitätsmerkmalen, wird der konzeptionelle Aufbau eines Polynombeamformers erläutert.

2.1 Signalmodell

Um ein Grundverständnis für die verwendete technische Notation zu schaffen, wird in diesem Abschnitt ein Signalmodell anhand eines FSB eingeführt. Dieses Modell enthält neben den Bezeichnungen der einzelnen Signale und Filter die Erklärung der mathematischen Operatoren, welche zur Berechnung der Resultate benötigt werden. Abbildung 2.1 zeigt das Modell eines FSB, welches Signale unter relativem Winkel ϑ empfängt. Signale, welche an den M Mikrofonen empfangen werden, werden als $x_m[n]$, $m = 1, \dots, M$ bezeichnet. Der Signalparameter n repräsentiert die diskrete Zeit, da die empfangenen Signale zur Weiterverarbeitung an den Mikrofonen mit einem Analog/Digital Umwandler konvertiert wurden. Die empfangenen Signale werden je mit einem mikrofonspezifischen Filter mit endlicher Impulsantwort (Finite Impulse

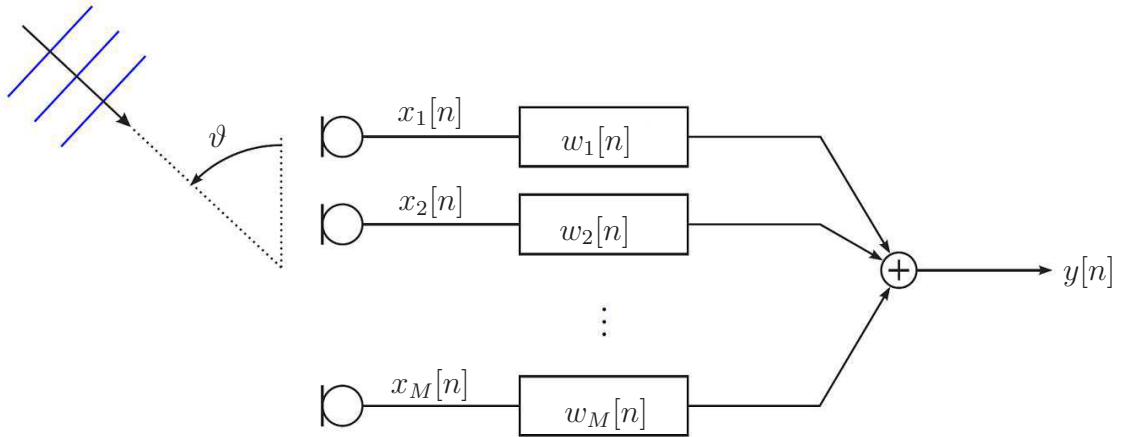


Abbildung 2.1: Model eines FSB [6].

Response (FIR)) $w_m[n]$, $m = 1, \dots, M$, der Länge L , gefaltet. Das Ausgangssignal wird wie folgt berechnet:

$$y[n] = \sum_{m=1}^M w_m[n] * x_m[n], \quad (2.1)$$

wobei der Operator $*$ die Faltung zweier Sequenzen bezeichnet. Diese Gleichung lässt sich auch in Matrixnotation formulieren,

$$y[n] = \mathbf{w}^T[n] * \mathbf{x}[n], \quad (2.2)$$

wobei der Vektor $\mathbf{w}[n]$ als

$$\mathbf{w}[n] = [w_1[n], \dots, w_M[n]]^T, \quad (2.3)$$

und der Vektor $\mathbf{x}[n]$ als

$$\mathbf{x}[n] = [x_1[n], \dots, x_M[n]]^T \quad (2.4)$$

definiert sind. Der Operator $(.)^T$ beschreibt die Transposition eines Vektors oder einer Matrix. Diese Berechnung im Zeitbereich lässt sich auch im Frequenzbereich darstellen,

$$Y(\omega) = \sum_{m=1}^M W_m(\omega) X_m(\omega), \quad (2.5)$$

beziehungsweise

$$Y(\omega) = \mathbf{w}_f^T(\omega)\mathbf{X}(\omega), \quad (2.6)$$

wobei $W_m(\omega) = \sum_{l=0}^{L-1} w_{m,l}[n]\exp(-j\omega/f_s l)$ die zeitdiskrete Fouriertransformation (Discrete-Time Fourier Transformation (DTFT)) von $w_m[n]$ beschreibt und $w_{m,l}[n]$ den l -ten Koeffizienten des m -ten Filters darstellt. Der Parameter f_s bezeichnet hier die Frequenz, mit der die Signale an den Mikrofonen abgetastet werden (Abtastrate). $Y(\omega)$ und $X_m(\omega)$ werden analog zu dieser Beschreibung dargestellt. Der Vektor $\mathbf{w}_f(\omega)$ wird wie folgt definiert:

$$\mathbf{w}_f(\omega) = [W_1(\omega), \dots, W_M(\omega)]^T. \quad (2.7)$$

Sämtliche Vektorgrößen werden in dieser Arbeit als fett geschriebene Parameter gekennzeichnet. In dieser Arbeit wird ein Beamformer durch seine Übertragungsfunktion,

$$B(\omega, \vartheta) = \sum_{m=1}^M W_m(\omega)e^{-j\omega\tau_n(\vartheta)}, \quad (2.8)$$

charakterisiert, wobei $\tau_n(\vartheta) = d_n \cos(\vartheta)/c$ die richtungsabhängige Verzögerung darstellt, c die Signalausbreitungsgeschwindigkeit beschreibt und d_n den Abstand des n -ten Mikrofons zu einem Referenzmikrofon, mit $d_{ref} = 0$, angibt. Die im weiteren Verlauf der Arbeit betrachteten FIR Filter haben zusätzlich die Eigenschaft der Linearität und der Zeitinvarianz (Linear Time Invariant (LTI)), wodurch Zeitunabhängigkeit impliziert wird und der diskrete Zeitparameter n bei diesen entfällt. Der nächste Abschnitt führt Beamforming-Charakteristiken und die, in dieser Arbeit verwendeten, Beamformer-Modelle ein.

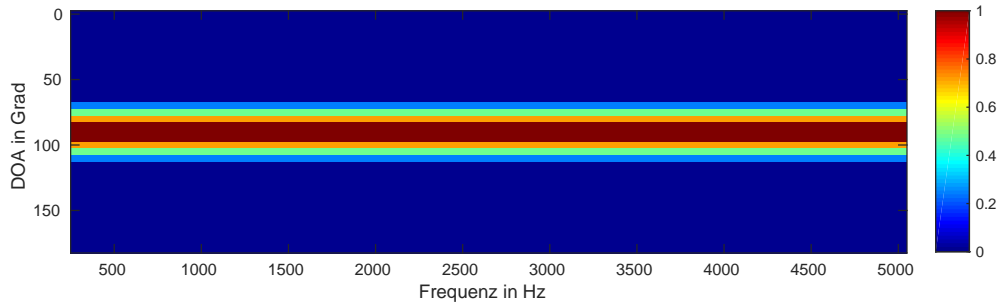
2.2 Beamforming

Eine Methode zur Signalextraktion stellt das Beamforming-Prinzip dar. Ein Beamformer besteht aus einer Reihe Mikrofone, welche in einer bestimmten Struktur angeordnet sind. Diese werden dazu genutzt, Signale räumlich zu unterscheiden. Diese Einteilung

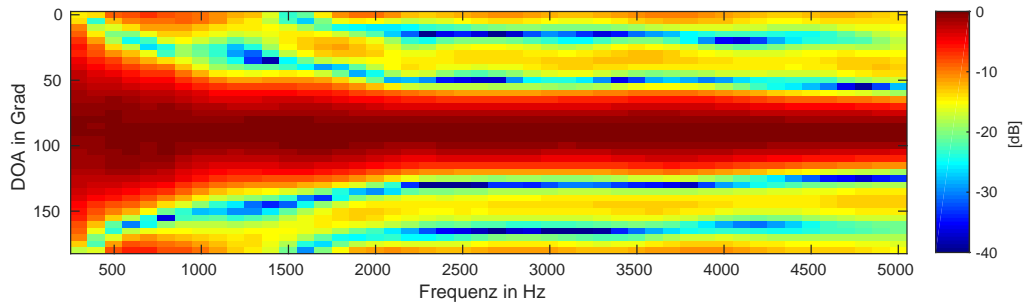
erfolgt durch die Messung der relativen Zeitunterschiede der Signale zueinander, welche je nach DOA verschieden sind. Um ein Signal aus einer bestimmten Richtung zu verstärken, müssen die an den Mikrofonen gemessenen Signale bezüglich der aus dieser Richtung resultierenden Zeitverschiebung synchronisiert und aufsummiert werden. Dies hat den Effekt, dass Signale aus genau dieser Richtung konstruktiv überlagert und Signale aus anderen Richtungen geschwächt werden. Wenn eine Schallwelle von einer punktförmigen Quelle erzeugt wird, dehnt diese sich kugelförmig, beziehungsweise kreisförmig in der Ebene aus. In dieser Arbeit werden Fernfeldbedingungen angenommen. In diesem Zusammenhang bedeutet Fernfeld, dass der Abstand zwischen Signalursprung und Mikrofonen, relativ zum Abstand der Mikrofone untereinander, so groß ist, dass die an den Mikrofonen eintreffenden Wellen als ebene Wellen angesehen werden können. Der folgende Abschnitt erläutert einige Charakteristiken, welche bestimmte Qualitätsmerkmale eines Beamformers untersuchen. Anschließend wird das Prinzip von Polynombeamforming vorgestellt.

2.2.1 Charakteristiken von Beamformern

In diesem Abschnitt werden Methoden eingeführt, welche sich zur Untersuchung gewisser Qualitätsmerkmale von Beamformern eignen. Als erste Untersuchungsmethode wird das Beampattern vorgestellt. Es beschreibt das Betragsquadrat der Übertragungsfunktion des Beamformers als generelles Verhalten auf Signale, abhängig aus welcher Richtung sie kommen (DOA) und welche Frequenzen enthalten sind. Abbildung 2.2 zeigt zwei Beispiele für Beampattern. Beide wurden für die Blickrichtung $\vartheta_{des} = 90^\circ$ entworfen. In Abbildung 2.2 (a) wird ein optimales, beziehungsweise angestrebtes Beampattern unter Berücksichtigung eines Durchlassbereichs gezeigt, welches in der Realität nur angenähert werden kann. Als Vergleich dazu, veranschaulicht Abbildung 2.2 (b) ein Beispiel eines realen Beampatterns. Beide besitzen ihre Maxima von ungefähr 0dB entlang der Frequenzachse bei 90° . Während das optimale Beampattern ein frequenzunabhängiges Verhalten aufweist, besitzt das reale Beampattern Nebenmaxima



(a) Optimales Beampattern



(b) Beispiel eines realen Beampattern

Abbildung 2.2: Optimales und reales Beampattern, entworfen für die Blickrichtung von 90° .

bei mehreren Frequenzen. Diese Nebenmaxima wirken sich negativ auf die Qualität des Beamformers aus, da Signalanteile mit diesen Frequenzen aus unerwünschten Richtungen nicht die maximale Unterdrückung erfahren. Um einen direkten Vergleich zwischen realen Beampattern zu schaffen, wird der mittlere quadratische Fehler (Mean Square Error (MSE)) zwischen angestrebtem und realem Beampattern ermittelt. Dieser Fehler wird als skalare Zahl berechnet, welche es ermöglicht, reale Beampattern zu vergleichen, unabhängig von der Beschaffenheit der einzelnen Nebenmaxima. Der mittlere quadratische Fehler wird wie folgt berechnet [7]:

$$\text{MSE}(\vartheta_{des}) = \frac{1}{N_f N_a} \sum_{q=0}^{N_f-1} \sum_{n=0}^{N_a-1} (|B_{act}(\omega_q, \vartheta_n)| - |B_{des}(\vartheta_n, \vartheta_{des})|)^2. \quad (2.9)$$

Dabei beschreibt N_f die Anzahl der berücksichtigten Frequenzen, N_a die Anzahl der berücksichtigten Winkel, $|B_{act}|$ den Betrag der berechneten und $|B_{des}|$ den Betrag der

angestrebte Übertragungsfunktion.

Ein weiteres Qualitätsmerkmal von Beamformern ist die Eigenschaft der Verzerrungsfreiheit. In diesem Fall bedeutet Verzerrungsfreiheit, dass Signale aus der gewünschten Richtung über alle Frequenzen unverzerrt empfangen werden, $|B_{act}(\omega_q, \vartheta_{des})| = 1$ ($\equiv 0\text{dB}$), $q = 0, \dots, N_f - 1$. Um Verzerrungsfreiheit bei einem Beamformer zu garantieren, muss diese Voraussetzung als Nebenbedingung bei der Berechnung der Beamformerfilter berücksichtigt werden. Abbildung 2.3 zeigt die Übertragungsfunktion, ausgewertet in der gewünschten Blickrichtung ϑ_{des} .

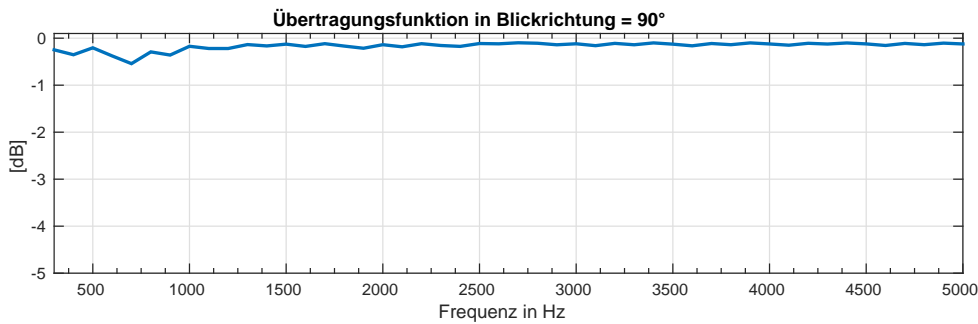


Abbildung 2.3: $B_{act}(\omega, \vartheta_{des} = 90^\circ)$ von Abbildung 2.2 (b).

Ein weiteres Qualitätsmerkmal stellt die Eigenschaft der Rauschkompensation (White Noise Gain (WNG)) dar. Das WNG beschreibt die Fähigkeit, weißes Rauschen über alle betrachteten Frequenzen relativ zur Übertragung, zu unterdrücken. Es wird sich nicht nur auf weißes Rauschen aus der Umgebung bezogen, sondern auch auf signalunabhängige Fehler und Ungenauigkeiten in der Mikrofonstruktur, welche als weißes Rauschen modelliert werden können. Dies bedeutet, je höher das WNG ist, desto robuster ist der Beamformer. Diese Eigenschaft lässt sich ebenfalls bei der Berechnung eines Beamformers berücksichtigen. Als Bedingung formuliert, würde eine untere Schranke γ gewählt werden, welche vom WNG bei keiner Frequenz unterschritten werden darf. Hat ein Beamformer eine entsprechend restriktive WNG-Schranke, wird dieser robuster gegenüber Rauschen und Fehlern in der Anordnung. Durch die Möglichkeit, diese Schranke frei wählen zu können, bekommt der Anwender direkte Kontrolle über die Robustheit des Beamformers. In Abbildung 2.4 wird das zu Abbildung 2.2 (b) gehörige

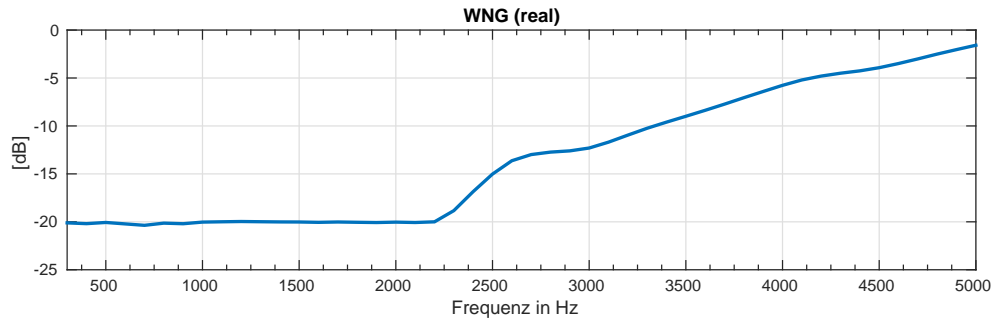


Abbildung 2.4: WNG von Abbildung 2.2 (b).

WNG veranschaulicht. In diesem Fall erfährt niederfrequentes Rauschen bis 2250Hz eine Unterdrückung von -20dB und höherfrequenteres Rauschen eine stärkere Unterdrückung von bis zu über -5dB . Wie eine solche Bedingung formuliert und berücksichtigt werden kann, wird in Kapitel 3 erklärt.

2.2.2 Polynombeamforming

In diesem Abschnitt wird das Konzept von Polynombeamforming vorgestellt. Bei der Anwendung von Beamformern, werden hohe Performanz der Beamformer und ein schnelles Ausrichten der Blickrichtung voraus gesetzt. Da die Berechnung einer neuen Konfiguration mit veränderter Blickrichtung für Beamformer relativ zeitaufwendig sein kann, wird eine dynamische Lösung benötigt. In dieser Arbeit wird der Polynombeamformer als dynamischer Lösungsansatz herangezogen und untersucht. Ein Polynombeamformer besteht aus $P + 1$ Beamformerkonfigurationen (Filter-and-Sum Unit (FSU)) und einem polynomiellen Nachfilter (Polynomial Postfilter (PPF)). Diese FSUs stehen jeweils für eine Stützstelle des Polynombeamformers in eine unterschiedliche Blickrichtung. Im PPF werden die Ausgänge der verschiedenen FSUs zum Ausgang des Polynombeamformers kombiniert. Abbildung 2.5 zeigt den schematischen Aufbau eines Polynombeamformers. Der Vorteil dieses Verfahrens besteht darin, dass sich die Blickrichtung dieses Beamformers, durch eine einzige Variable D kontrollieren lässt, während die Filter der FSUs nur einmal im Vorraus berechnet werden müssen. Der Ausgang des Polynombeamformers lässt sich analog zu Gleichung (2.1) im Zeitbereich

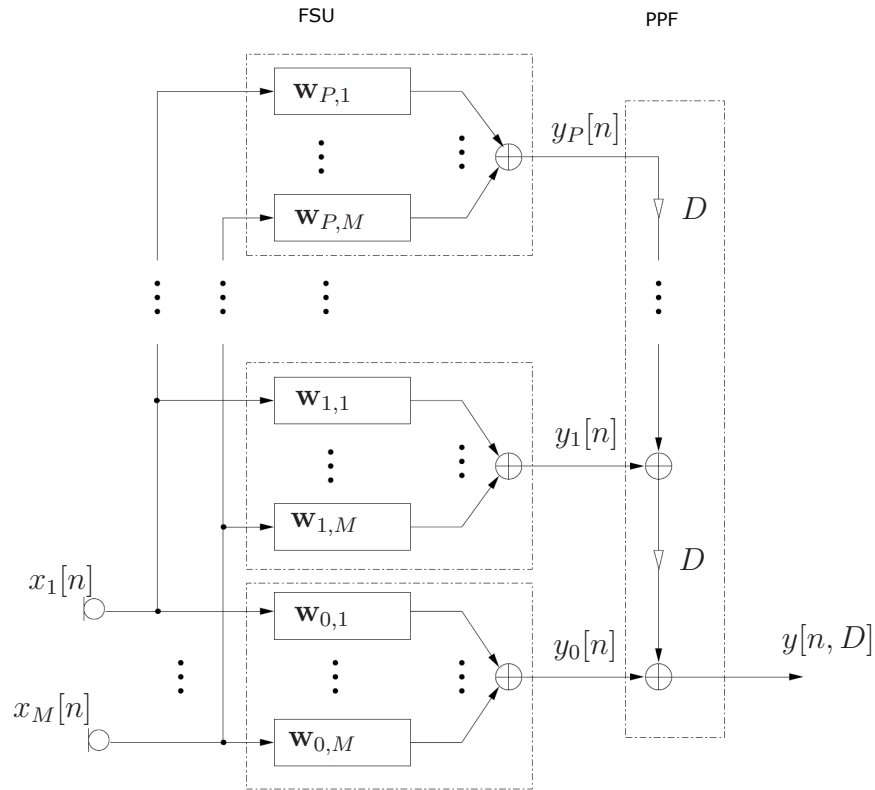


Abbildung 2.5: Schematischer Aufbau eines Polynombeamformer [3].

wie folgt schreiben:

$$y[n, D] = \sum_{p=0}^P D^p \sum_{m=1}^M w_{p,m} * x_m[n]. \quad (2.10)$$

Diese Gleichung wird, analog zu Gleichung (2.5), folgendermaßen im Frequenzbereich definiert:

$$Y(\omega) = \sum_{p=0}^P D^p \sum_{m=1}^M W_{p,m}(\omega) X_m(\omega). \quad (2.11)$$

Die Übertragungsfunktion des Polynombeamformers lässt sich analog zu der Übertragungsfunktion eines FSB aus Gleichung 2.8, als

$$B_D(\omega, \vartheta) = \sum_{p=0}^P D^p \sum_{m=1}^M W_{p,m}(\omega) e^{-j\omega\tau(\vartheta)}, \quad (2.12)$$

formulieren. Die Funktion des PPF kann nun so interpretiert werden, dass dieser zwischen den $P + 1$ festen Konfigurationen interpoliert, um so eine neue Blickrichtung zu verstärken. Abhängig von der Beschaffenheit der verwendeten Mikrofonanordnung und der festen Stützstellen, kann die Variable D einen bestimmten Blickbereich $[0^\circ, \vartheta_{max}]$ abdecken. Die Variable D ist definiert als

$$D = \frac{\vartheta_{des} - \vartheta_{max}/2}{\vartheta_{max}/2}, 0^\circ \leq \vartheta_{des} \leq \vartheta_{max}, \quad (2.13)$$

wobei ϑ_{des} die gewünschte Blickrichtung darstellt. Aus Gleichung 2.13 lässt sich der Wertebereich für D auf $-1 \leq D \leq 1$ bestimmen. Im nächsten Kapitel werden die konkreten Entwürfe der Beamformer, welche zu dem betrachteten Ansatz führen, vorgestellt.

Kapitel 3

Robuster auf kopfbezogene Übertragungsfunktionen basierender Polynombeamformer

In diesem Kapitel werden die Hintergründe dieser Arbeit beleuchtet. Als erstes wird das Optimierungsproblem eines robusten freifeldbasierten Beamformer geschildert, welcher als FSB entworfen wurde. Aus diesem Ansatz wurde der robuste freifeldbasierte Polynombeamformer entwickelt, dessen Konzept im nächsten Teil beschrieben wird. Diese Ansätze wurden auf den Roboter NAO angewendet und getestet. Verglichen mit den Versuchen im Freifeld führten diese jedoch nicht zu zufriedenstellenden Resultaten. Aus diesem Grund musste der Einfluss des Roboterkopfes, auf das Schallfeld, berücksichtigt werden. Um diesem Einfluss entgegen zu wirken, wurden HRTFs gemessen und beim Entwurf berücksichtigt. Mit diesen Messungen wurde der robuste freifeldbasierte Beamformer zum robusten HRTF-basierten Beamformer erweitert, welcher im darauf folgenden Abschnitt vorgestellt wird. Dieser liefert, angewendet auf den Roboter NAO, zufriedenstellende Ergebnisse, woraus sich die Motivation für den robusten HRTF-basierten Polynombeamformer ergab. Dieser Ansatz, dessen Anwendung in dieser Arbeit untersucht wird, wird im letzten Teil dieses Kapitels vorgestellt.

3.1 Robustes Superdirektives Beamforming unter Freifeldannahme

Im Folgenden wird das Optimierungsproblem des robusten, supergerichteten, freifeldbasierten Beamformer vorgestellt. Dieser Entwurf basiert auf dem Beamformerproblem, welches in [5] vorgestellt wird. Supergerichtete Beamformer haben den Vorteil, dass sie trotz kleiner Mikrofonanordnungen ein hohes Richtwirkungsmaß erzielen können [5]. Das betrachtete Beamformer-Modell wird, analog zu Gleichung 2.8, durch dessen Übertragungsfunktion $B(\omega, \vartheta)$ beschrieben. Der Nachteil supergerichteter Beamformer ist, dass sie sehr sensitiv auf weißes Rauschen und kleine Fehler in der Mikrofonanordnung reagieren. Da diese Fehler von Mikrofon zu Mikrofon nahezu unkorreliert sind, können diese ebenfalls als weißes Rauschen modelliert werden. Um den Beamformer robust gegenüber weißem Rauschen zu entwerfen, wird eine WNG Bedingung berücksichtigt. Das WNG lässt sich frequenzabhängig durch,

$$A(\omega) = \frac{|\mathbf{w}_f^T(\omega)\mathbf{g}_{des}(\omega)|^2}{\mathbf{w}_f^H(\omega)\mathbf{w}_f(\omega)}, \quad (3.1)$$

berechnen, wobei

$$\mathbf{g}_{des}(\omega) = [\exp(-j\omega\tau_0(\vartheta_{des})), \dots, \exp(-j\omega\tau_{N-1}(\vartheta_{des}))]^T \quad (3.2)$$

den Ausrichtungsvektor in Richtung ϑ_{des} und $(\cdot)^H$ den hermiteschen Operator beschreibt. Im ersten Schritt des Entwurfsverfahrens, wird eine optimale Übertragungsfunktion $\hat{B}(\omega, \vartheta)$ formuliert, welcher sich die gesuchte Übertragungsfunktion $B(\omega, \vartheta)$ möglichst genau annähern soll. Diese gewünschte Übertragungsfunktion beschreibt ein frequenzinvariantes (Frequency Invariant (FI)) Verhalten in Blickrichtung ϑ_{des} . In Abbildung 2.2 (a), wird beispielhaft eine solche gewünschte Übertragungsfunktion veranschaulicht. Um dieses Problem numerisch lösen zu können, werden der betrachtete Frequenzbereich in Q Frequenzen $\omega_q, q = 0, \dots, Q - 1$ und der betrachtete Winkelbereich, abhängig vom Einfallswinkel der Signale in Bezug auf den Beamformer, in N Winkel $\vartheta_n, n = 0, \dots, N - 1$ diskretisiert. Das Problem lässt sich nun, analog zu Gleichung

(2.8), wie folgt beschreiben:

$$\hat{B}(\omega_q, \vartheta_n) \stackrel{!}{=} \sum_{m=1}^M W_m(\omega_q) e^{-j\omega_q \tau_m(\vartheta_n)}. \quad (3.3)$$

Führt man die Matrix $\mathbf{G}(\omega_q)$ als $[\mathbf{G}(\omega_q)]_{nm} = \exp(-j\omega_q \tau_m(\vartheta_n))$ ein, lässt sich das Problem in Matrixschreibweise als

$$\hat{\mathbf{b}}(\omega_q) \stackrel{!}{=} \mathbf{G}(\omega_q) \mathbf{w}_f(\omega_q), \quad (3.4)$$

darstellen, wobei $\hat{\mathbf{b}}(\omega_q) = [\hat{B}(\omega_q, \vartheta_0), \dots, \hat{B}(\omega_q, \vartheta_{N-1})]$ repräsentiert. Dieses Problem muss für jede Frequenz ω_q gelöst werden. Da die Anzahl der diskreten Winkel N in der Regel größer ist als die Anzahl der Mikrofone M , ist das Problem überbestimmt. Aus diesem Grund wird eine Lösung mit kleinstem quadratischen (Least-Squares (LS)) Fehler gesucht. Das überbestimmte Problem lässt sich als,

$$\min_{\mathbf{w}_f(\omega_q)} \|\mathbf{G}(\omega_q) \mathbf{w}_f(\omega_q) - \hat{\mathbf{b}}(\omega_q)\|_2^2 \quad (3.5)$$

beschreiben. Um Signale aus der gewünschten Blickrichtung nicht zu verzerren, wird dem Optimierungsproblem die Bedingung der Verzerrungsfreiheit,

$$\mathbf{w}_f^T(\omega_q) \mathbf{g}_{des}(\omega_q) \stackrel{!}{=} 1, \quad (3.6)$$

hinzugefügt. Da supergerichtete Beamformer anfällig gegenüber weißem Rauschen und Fehlern in der Mikrofonanordnung sind, wird das frequenzabhängige WNG durch eine untere Schranke beschränkt. Aus Gleichung (3.1) wird folgende Bedingung definiert:

$$\frac{|\mathbf{w}_f^T(\omega_q) \mathbf{g}_{des}(\omega_q)|^2}{\mathbf{w}_f^H(\omega_q) \mathbf{w}_f(\omega_q)} \geq \gamma > 0, \quad (3.7)$$

welche bei der Optimierung berücksichtigt wird. Durch Lösen des Optimierungsproblems (3.5) unter den Bedingungen (3.6) und (3.7), erhält man die Konfiguration eines RLSFI-Beamformer. Da dieses Problem im Frequenzbereich optimiert wurde, müssen die Ergebnisse zur Anwendung in den Zeitbereich transformiert werden. Dies wird mittels einer FIR Approximation der Länge L erreicht. Der nächste Abschnitt behandelt die Entwicklung des RLSFI-Polynombeamformers.

3.2 Robustes Polynombeamforming unter Freifeld- annahme

In vielen Beamformeranwendungen, wie zum Beispiel im Bereich robotischen Hörens, wird vorausgesetzt, dass die Blickrichtung des Beamformers dynamisch einstellbar ist. Dies wird gefordert, da sich die interagierenden Quellen bewegen könnten. Das in Kapitel 3.1 vorgestellte Beamformermodell wird für eine feste Blickrichtung ϑ_{des} entworfen und müsste für eine andere Blickrichtung neu entworfen werden. In diesem Kapitel wird, basierend auf dem Ansatz des RLSFI-Beamformers, der RLSFI-Polynombeamformer vorgestellt. Bei diesem Entwurf wird sich auf das in [3] beschriebene Beamformerkonzept bezogen. Abbildung 2.5 zeigt den schematischen Aufbau eines solchen Polynombeamformer, welcher analog zu Gleichung (2.12), durch seine Übertragungsfunktion $B_D(\omega, \vartheta)$ beschrieben wird. Der Polynombeamformer besteht aus $P + 1$ FSUs, welche jeweils für verschiedene Blickrichtungen $\vartheta_{des,i}, i = 0, \dots, I - 1$ stehen. Die Ausgänge dieser FSUs werden im PPF, analog zu Gleichung (2.10), zum Ausgangssignal $y[n, D]$ zusammengefasst. Die Variable D beschreibt nach Gleichung (2.13) die gewünschte Blickrichtung ϑ_{des} . Im ersten Schritt werden optimal gewünschte, frequenzinvariante (FI) Übertragungsfunktionen $\hat{B}_{des,i}(\vartheta, \vartheta_{des,i})$, definiert. In diesem Verfahren werden, analog zur Methode aus Kapitel 3.1, der betrachtete Frequenzbereich in Q Frequenzen $\omega_q, q = 0, \dots, Q - 1$ und der betrachtete Winkelbereich in N Winkel $\vartheta_n, n = 0, \dots, N - 1$ diskretisiert. Das Problem lässt sich nun, analog zu Gleichung 2.12, wie folgt,

$$\hat{B}_{des,i}(\vartheta_n, \vartheta_{des,i}) \stackrel{!}{=} \sum_{p=0}^P D_i^p \sum_{m=1}^M W_{m,p}(\omega_q) e^{-j\omega_q \tau_m(\vartheta_n)}, \quad (3.8)$$

erweitern. Verwendet man, analog zu Gleichung 3.4, die Matrix $\mathbf{G}(\omega_q)$, lässt sich das Problem als,

$$\hat{\mathbf{b}}_{des,i} \stackrel{!}{=} \mathbf{G}(\omega_q) \mathbf{W}_f(\omega_q) \mathbf{d}_i, \quad \forall i = 0, \dots, I - 1, \quad (3.9)$$

formulieren, wobei $\hat{\mathbf{b}}_{des,i} = [\hat{B}_{des,i}(\vartheta_0, \vartheta_{des,i}), \dots, \hat{B}_{des,i}(\vartheta_{N-1}, \vartheta_{des,i})]^T$, $[\mathbf{W}_f(\omega_q)]_{mp} = W_{m,p}(\omega_q)$ und $\mathbf{d}_i = [D_i^0, \dots, D_i^P]^T$ darstellen. Da die Anzahl der diskreten Winkel in

der Regel größer ist als das Produkt aus der Anzahl der Mikrofone und der Anzahl der FSUs, $N \geq M(P + 1)$, ist jedes der i Probleme überbestimmt und es wird eine Lösung mit kleinstem quadratischen Fehler (LS) gesucht. Das Optimierungsproblem dieses Polynombeamformer lässt sich als,

$$\min_{\mathbf{w}_f(\omega_q)} \sum_{i=0}^{I-1} \|\mathbf{G}(\omega_q) \mathbf{W}_f(\omega_q) \mathbf{d}_i - \hat{\mathbf{b}}_{des,i}\|_2^2 \quad (3.10)$$

formulieren, welches für jede Frequenz $\omega_q, q = 0, \dots, Q - 1$ gelöst werden muss. Das Problem kann, durch Umformen, in die Form des Problems 3.5 gebracht werden. Wird die Matrix $\mathbf{W}_f(\omega_q)$ durch den Vektor

$$\mathbf{w}_f(\omega_q) = [W_{1,0}(\omega_q), \dots, W_{1,P}(\omega_q), W_{2,0}(\omega_q), \dots, W_{M,P}(\omega_q)]^T \quad (3.11)$$

dargestellt, erhält man folgende Beziehung

$$\begin{aligned} \mathbf{W}_f(\omega_q) \mathbf{d}_i &= (\mathbf{I}_M \otimes \mathbf{d}_i^T) \mathbf{w}_f(\omega_q) \\ &= \mathbf{D}_i \mathbf{w}_f(\omega_q), \end{aligned} \quad (3.12)$$

wobei \otimes das Kroneckerprodukt bezeichnet und \mathbf{I}_M eine $M \times M$ Einheitsmatrix darstellt. Das Optimierungsproblem 3.10 wird zu

$$\min_{\mathbf{w}_f(\omega_q)} \sum_{i=0}^{I-1} \|\mathbf{G}(\omega_q) \mathbf{D}_i \mathbf{w}_f(\omega_q) - \hat{\mathbf{b}}_{des,i}\|_2^2 \quad (3.13)$$

umgeformt. Werden die Matrix $\mathcal{G}(\omega_q) = [\mathbf{G}(\omega_q) \mathbf{D}_0, \dots, \mathbf{G}(\omega_q) \mathbf{D}_{I-1}]^T$ und $\hat{\mathbf{b}}_{des} = [\hat{\mathbf{b}}_{des,0}^T, \dots, \hat{\mathbf{b}}_{des,I-1}^T]^T$ eingeführt, lässt sich das Problem zu

$$\min_{\mathbf{w}_f(\omega_q)} \|\mathcal{G}(\omega_q) \mathbf{w}_f(\omega_q) - \hat{\mathbf{b}}_{des}\|_2^2 \quad (3.14)$$

umformen. Um Signale, aus den gewählten Blickrichtungen $\vartheta_{des,i}$, nicht zu verzerren, werden dem Optimierungsproblem folgende I Bedingungen der Verzerrungsfreiheit hinzugefügt:

$$\mathbf{g}_{des,i}^T(\omega_q) \mathbf{D}_i \mathbf{w}_f(\omega_q) = 1 \quad \forall i = 0, \dots, I - 1. \quad (3.15)$$

Das WNG wird auch für diesen Beamformer als wichtiges Qualitätskriterium angesehen. Daher wird bei der Optimierung eine Schranke γ festgelegt, welche vom WNG nicht unterschritten werden darf. Diese Schranke wird als I Bedingungen

$$\frac{|\mathbf{g}_{des,i}^T(\omega_q)\mathbf{D}_i\mathbf{w}_f(\omega_q)|^2}{\|\mathbf{D}_i\mathbf{w}_f(\omega_q)\|_2^2} \geq \gamma > 0 \quad \forall i = 0, \dots, I - 1 \quad (3.16)$$

definiert. Mit dieser Schranke wird der Polynombeamformer als robust bezeichnet. Die Bedingung der Verzerrungsfreiheit und die WNG-Schranke können jedoch nur von den gewählten Blickrichtungen $\vartheta_{des,i}$ genau erfüllt werden. Wird eine Blickrichtung D_{new} gewählt, welche nicht mit einer der Blickrichtungen $\vartheta_{des,i}$ korrespondiert, können diese Bedingungen verletzt werden. Im nächsten Abschnitt, wird der Einfluss des Kopfes des Roboters NAO auf das Schallfeld bei der Berechnung des Beamformers aus Kapitel 3.1 berücksichtigt.

3.3 Robustes auf kopfbezogene Übertragungsfunktionen basierendes Beamforming

In Kapitel 3.1 und 3.2 wurden Beamformer unter der Annahme von Freifeldbedingungen entworfen. Da diese, angewendet auf den Roboter NAO, relativ schlechte Ergebnisse erzielen [2], wird in diesem Abschnitt ein Verfahren, anhand des RLSFI-Beamformer, vorgestellt, welches den Kopf des Roboters als Verzerrer berücksichtigt. In Abbildung

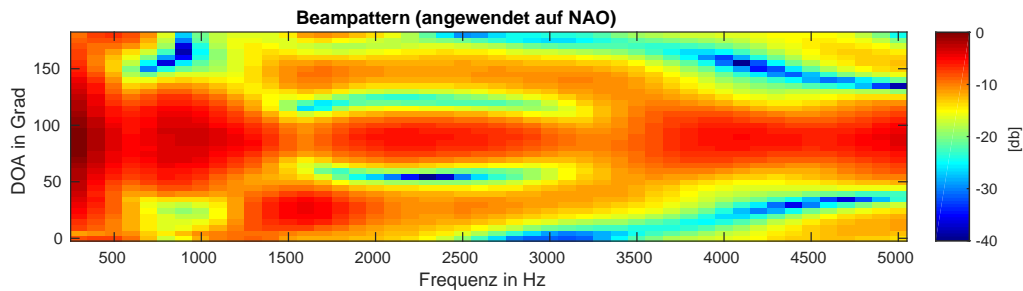


Abbildung 3.1: Beampattern mit $\vartheta_{des} = 90^\circ$ des freifeldbasiertem RLSFI-Beamformer, angewendet auf den Roboter NAO.

3.1 wird das Beampattern des freifeldbasierten RLSFI-Beamformer, angewendet auf den Roboter, gezeigt. Durch den Vergleich zwischen diesem und dem Beampattern von Abbildung 2.2 (b), ist deutlich zu erkennen, dass keine gute räumliche Unterscheidung mehr gegeben ist. Im ersten Schritt des Entwurfes, werden die Übertragungsfunktionen $h_{n,m}(\omega_q)$ zwischen m -ten Mikrofon des Roboters und n -ter Quellenposition, abhängig von dem Einfallswinkel $\vartheta_n, n = 0, \dots, N - 1$ gemessen. Diese Messungen werden für jede Frequenz $\omega_q, q = 0, \dots, Q - 1$, des betrachteten Frequenzbereichs, durchgeführt. Diese Übertragungsfunktionen ersetzen die Freifeldvektoren aus Kapitel 3.1 und das Optimierungsproblem (3.5) lässt sich zu,

$$\min_{\mathbf{w}_f(\omega_q)} \|\mathbf{G}_{\text{HRTF}}(\omega_q)\mathbf{w}_f(\omega_q) - \hat{\mathbf{b}}(\omega_q)\|_2^2 \quad (3.17)$$

umformen, wobei $[\mathbf{G}_{\text{HRTF}}(\omega_q)]_{nm} = h_{n,m}(\omega_q)$ darstellt. Durch Einführen der HRTFs, werden auch die Bedingung der Verzerrungsfreiheit (3.6) zu,

$$\mathbf{w}_f^T(\omega_q)\mathbf{g}_{\text{HRTF}}(\omega_q) \stackrel{!}{=} 1, \quad (3.18)$$

und die WNG-Schranke zu,

$$\frac{|\mathbf{w}_f^T(\omega_q)\mathbf{g}_{\text{HRTF}}(\omega_q)|^2}{\mathbf{w}_f^H(\omega_q)\mathbf{w}_f(\omega_q)} \geq \gamma > 0, \quad (3.19)$$

angepasst, wobei $\mathbf{g}_{\text{HRTF}}(\omega_q) = [h_{des,1}(\omega_q), \dots, h_{des,M}(\omega_q)]^T$ darstellt und $h_{des,m}(\omega_q)$ die gemessene Übertragungsfunktion zwischen gewünschter Richtung und m -ten Mikrofon beschreibt. Im nächsten Teil des Kapitels, wird das Optimierungsproblem des HRTF-basierten RLSFI-Polynombeamformer beschrieben.

3.4 Robuster auf kopfbezogene Übertragungsfunktionen basierender Polynombeamformer

Da die Anwendung der HRTFs auf den RLSFI-Beamformer relativ gute Ergebnisse liefert [2], wird in dieser Arbeit die Anwendung des HRTF-basierten RLSFI-Polynombeamformers getestet und evaluiert. In diesem Teil der Arbeit wird das Optimierungsproblem

dieses Beamformers beschrieben. Dieser wird, analog zu Gleichung (2.12), durch dessen Übertragungsfunktion $B_D(\omega, \vartheta)$ charakterisiert. Statt Freifeldvektoren, werden die HRTFs bei der Berechnung der einzelnen Richtungen mit berücksichtigt, wodurch sich das Optimierungsproblem (3.13) zu

$$\min_{\mathbf{w}_f(\omega_q)} \sum_{i=0}^{I-1} \|\mathbf{G}_{\text{HRTF}}(\omega_q) \mathbf{D}_i \mathbf{w}_f(\omega_q) - \hat{\mathbf{b}}_{des,i}\|_2^2 \quad (3.20)$$

modifizieren lässt. Wird nun die Matrix $\mathcal{G}_{\text{HRTF}}(\omega_q) = [\mathbf{G}_{\text{HRTF}}(\omega_q) \mathbf{D}_0, \dots, \mathbf{G}_{\text{HRTF}}(\omega_q) \mathbf{D}_{I-1}]^T$ eingeführt, lässt sich das Problem, analog zu Problem (3.14), wie folgt umschreiben:

$$\min_{\mathbf{w}_f(\omega_q)} \|\mathcal{G}_{\text{HRTF}}(\omega_q) \mathbf{w}_f(\omega_q) - \hat{\mathbf{b}}_{des}\|_2^2. \quad (3.21)$$

Um Signale aus den gewählten Blickrichtungen $\vartheta_{des,i}$ nicht zu verzerren, werden dem Optimierungsproblem I Bedingungen der Verzerrungsfreiheit,

$$\mathbf{g}_{\text{HRTF},i}^T(\omega_q) \mathbf{D}_i \mathbf{w}_f(\omega_q) = 1 \quad \forall i = 0, \dots, I-1, \quad (3.22)$$

hinzugefügt, wobei $\mathbf{g}_{\text{HRTF},i}(\omega_q) = [h_{des,i,1}(\omega_q), \dots, h_{des,i,M}(\omega_q)]^T$ darstellt. Zusätzlich wird das WNG bei diesem Ansatz, analog zu Bedingung (3.16), durch die untere Schranke γ , gegeben durch,

$$\frac{|\mathbf{g}_{\text{HRTF},i}^T(\omega_q) \mathbf{D}_i \mathbf{w}_f(\omega_q)|^2}{\|\mathbf{D}_i \mathbf{w}_f(\omega_q)\|_2^2} \geq \gamma > 0 \quad \forall i = 0, \dots, I-1, \quad (3.23)$$

beschränkt. In dieser Anwendung können die Bedingung der Verzerrungsfreiheit und der WNG-Schranke nur für die gewählten Stützstellen erfüllt werden. Interpolierte Blickrichtungen, welche nicht in den gewählten Blickrichtungen enthalten sind, könnten diese Bedingungen verletzen. Das nächste Kapitel erläutert den genauen Versuchsaufbau des Experiments und präsentiert die entstandenen Resultate.

Kapitel 4

Evaluierung des robusten auf kopfbezogene Übertragungsfunktionen basierenden Polynombeamformers

In diesem Kapitel werden die Ergebnisse des Experiments vorgestellt. Der erste Abschnitt beschreibt den genauen Versuchsaufbau, welcher zur Durchführung der einzelnen Tests verwendet wurde. Neben der Bewertung des allgemeinen Verhaltens des HRTF-basierten RLSFI-Polynombeamformers, wird ein direkter Vergleich zu dem freifeldbasierten RLSFI-Polynombeamformer und dem HRTF-basierten RLSFI-Beamformer gezogen.

4.1 Aufbau des Experiments

Bevor die Ergebnisse der einzelnen Versuche vorgestellt werden, wird in diesem Teil der Arbeit der Aufbau des Experiments beschrieben. Sämtliche Versuche und die daraus entstandenen Ergebnisse wurden als Simulationen in MATrix LABoratory, The

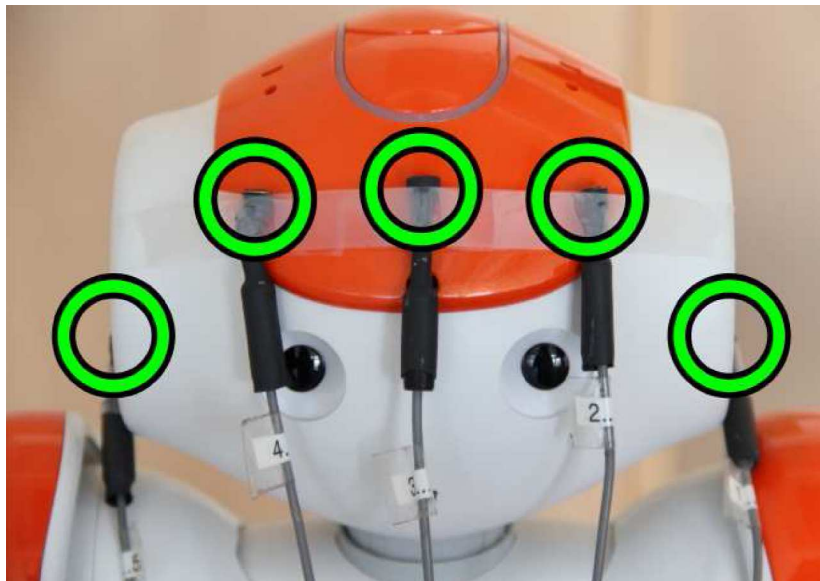


Abbildung 4.1: markierte Mikrofone des Roboters NAO [2].

MathWorks (MATLAB) durchgeführt. Eingangssignale bestehen aus einem Nutzsignal und einem Störsignal. Diese werden mit gemessenen Raumimpulsantworten (Room Impulse Responses (RIRs)) gefaltet und simulieren die zeitdiskreten Mikrofonsignale bei einer Abtastrate von $f_s = 16\text{kHz}$. Das Experiment berücksichtigt fünf Mikrofonkanäle, welche die Mikrofone des Roboters NAO simulieren. Abbildung 4.1 veranschaulicht die Positionen der einzelnen Mikrofone des Roboters. Sämtliche eintreffende Eingangssignale werden unter einem Vertikalwinkel von $\Theta = 56^\circ$ und einem Horizontalwinkelbereich von $\vartheta \in (0^\circ, \dots, 180^\circ)$ simuliert. Um das Optimierungsproblem 3.21 numerisch in MATLAB beschreiben zu können, werden der Horizontalwinkelbereich in $\Delta\vartheta = 5^\circ$ Schritten zu $\vartheta \in [0^\circ, 5^\circ, \dots, 180^\circ]$ und der untersuchte Frequenzbereich in $\Delta\omega = 100\text{Hz}$ Schritten diskretisiert. In sämtlichen nachfolgenden Versuchen, wird die WNG-Schranke, welche in Gleichung (3.23) definiert wurde, auf $\gamma = -20\text{dB}$ festgelegt. In diesem Experiment wurde das CVX-Paket von MATLAB herangezogen, um das Optimierungsproblem (3.21), unter den Bedingungen (3.22) und (3.23), zu lösen. Nachdem das Optimierungsproblem im Frequenzbereich gelöst wurde, werden die Koeffizienten der einzelnen FIR-Filter durch inverse zeitdiskrete Fouriertransformation (Inverse Discrete-Time Fourier Transformation (IDTFT)) approximiert. Bei

dieser Transformation werden Filter, der Länge 1024, berechnet. Zusätzlich zu den beschriebenen Beamforming-Charakteristiken aus Kapitel 2.2.1, werden signalabhängige Methoden zum Vergleich der einzelnen Beamformeransätzen angewendet. Zum einen werden die Ausgangssignale der Beamformer durch den Spracherkenner PocketSphinx [8] analysiert, zum anderen wird jeweils segmentweise das frequenzbewertete Signal-Rausch-Verhältnis (frequency-weighted Segmental Signal-to-Noise Ratio (fwSegSNR)) berechnet. Durch Anwendung eines Spracherkenners können die Beamformer, bei gleichen Eingangssignalen, nach ihren jeweiligen Fehlerquoten (Word Error Rate (WER)) verglichen werden. Das fwSegSNR [9] ist definiert als,

$$\text{fwSegSNR} = \frac{10}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^K W(j, m) \log_{10} \frac{|X(j, m)|^2}{(|X(j, m)| - |\hat{X}(j, m)|)^2}}{\sum_{j=1}^K W(j, m)}, \quad (4.1)$$

wobei K der Anzahl an betrachteten, diskreten Frequenzen entspricht, $W(j, m)$ die Gewichtung der j -ten Frequenz, beschrieben in [9], darstellt und M die Anzahl an Segmente repräsentiert, in die das Signal eingeteilt wurde. $|X(j, m)|$ beschreibt das Spektrum des Ausgangssignals des Beamformers an der j -ten Frequenz im m -ten Segment, während $|\hat{X}(j, m)|$ das Spektrum der Nutzsingalkomponenten an diesem Ausgang mit derselben Frequenz darstellt. Größere Werte des fwSegSNR deuten auf eine bessere Qualität des Beamformer hin, da bei kleinerem Unterschied zwischen Ausgangssignal und darin enthaltene Nutzsingalkomponenten, größere Werte bei der Berechnung von Gleichung (4.1) entstehen. Dieser Unterschied entspricht der Rauschkomponente im Signal, welche versucht wird zu minimieren.

4.2 Evaluation der Konfiguration des robusten auf kopfbezogene Übertragungsfunktionen basierenden Polynombeamformers

Im folgenden Abschnitt wird das allgemeine Verhalten des RLSFI-Polynombeamformer ausgewertet. Da ein Polynombeamformer zwischen $P + 1$ Stützstellen interpoliert, wird

zunächst die optimale Anzahl an Stützstellen bestimmt. Dazu werden Konstellationen mit äquidistanten Stützstellen von Polynomgraden $P = [1, 2, \dots, 6]$, betrachtet. Die Konstellation mit $P = 5$ ist in diesem Fall die Ausnahme, da bei gleichem Abstand der Stützstellen, $\Delta\vartheta_{P=5} = 36^\circ$, diese nicht in dem diskretisierten Horizontalwinkelbereich $\vartheta = [0^\circ, 5^\circ, \dots, 180^\circ]$ liegen. Für diese Konstellation wurden nicht äquidistante Stützstellen gewählt, welche mit geringst möglicher Abweichung, zu ihren theoretischen Positionen, in diesem Bereich enthalten sind. Tabelle 4.1 zeigt eine Übersicht über die

Polynomgrad	$\Delta\vartheta$	Position ϑ der einzelnen Stützstellen
1	180°	$[0^\circ, 180^\circ]$
2	90°	$[0^\circ, 90^\circ, 180^\circ]$
3	60°	$[0^\circ, 60^\circ, 120^\circ, 180^\circ]$
4	45°	$[0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ]$
5	(36°)	$[0^\circ, 35^\circ, 70^\circ, 110^\circ, 145^\circ, 180^\circ]$
6	30°	$[0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ]$

Tabelle 4.1: Übersicht über die evaluierten Polynomgrade mit zugehörigen Stützstellen.

evaluierten Polynomgrade mit den jeweils zugehörigen Stützstellen. Um das Verhalten, der einzelnen Konstellationen, über eine große Anzahl an Blickrichtungen zu untersuchen, wird zu jeder diskreten Blickrichtung der mittlere quadratische Fehler (MSE) zwischen angestrebtem und berechnetem Beampattern, unter Verwendung von Gleichung (2.9), ermittelt. Dazu werden zunächst sämtliche Konstellationen mit $P = [1, 2, \dots, 6]$ berechnet. Im Anschluss interpoliert jeder dieser Entwürfe, sämtliche diskrete Blickrichtungen, $\vartheta = [0^\circ, 5^\circ, \dots, 180^\circ]$. Diese Ergebnisse werden mittels ihres MSE-Wertes untersucht. Abbildung 4.2 veranschaulicht die MSE-Auswertung des HRTF-basierten RLSFI-Polynombeamformer mit Polynomgrad $P = 4$. Zusätzlich wurde die Auswertung für den HRTF-basierten RLSFI-Beamformer aufgetragen, welcher für sämtliche, diskrete Blickrichtungen optimiert wurde. Es ist zu erkennen, dass die MSE-Werte beider Entwürfe an den Stützstellen des Polynombeamformer gleich sind und die übrigen Werte des Polynombeamformer in der Regel größer sind. Um einen Vergleich zwischen

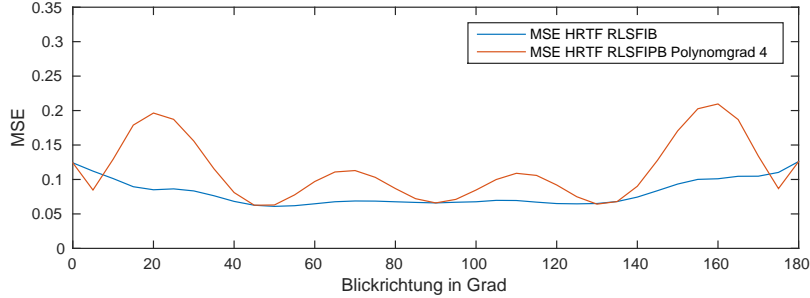


Abbildung 4.2: MSE-Auswertung des HRTF-basierten RLSFI-Polynombeamformer für Polynomgrad 4 und des HRTF-basierten RLSFI-Beamformer.

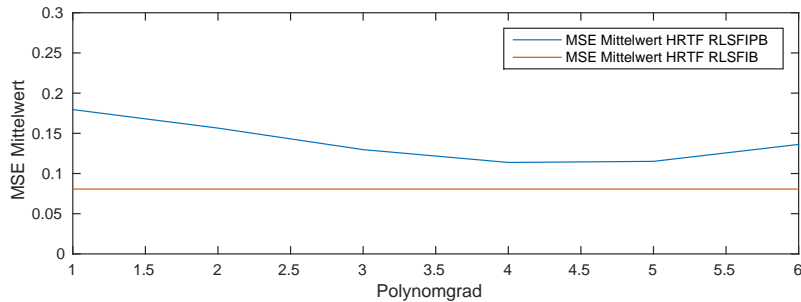


Abbildung 4.3: Durchschnittlicher MSE der einzelnen Konstellationen des HRTF-basierten RLSFI-Polynombeamformer, verglichen mit dem HRTF-basierten RLSFI-Beamformer.

den Konstellationen mit verschiedenen Stützstellen zu ziehen, wird zu jeder Auswertung, der Mittelwert über sämtliche MSE-Werte, MSE Mittelwert $\overline{\text{MSE}}(P)$, gebildet. Diese Mittelung wird wie folgt berechnet:

$$\overline{\text{MSE}}(P) = \frac{1}{N} \sum_{n=0}^{N-1} \text{MSE}_P(\vartheta_n), \quad (4.2)$$

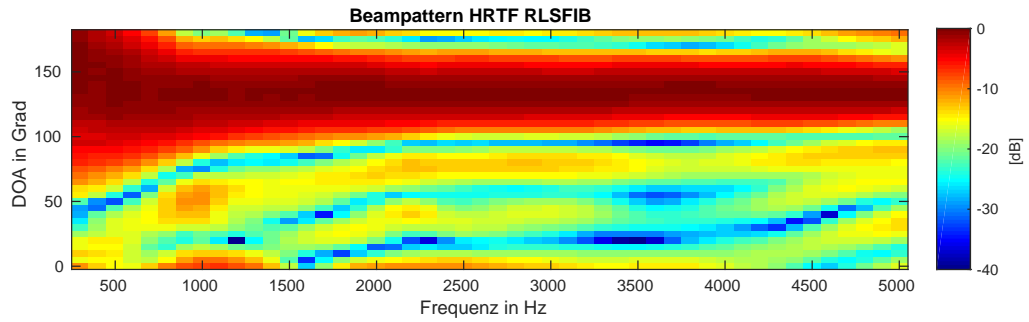
wobei $\text{MSE}_P(\vartheta_n)$ den MSE-Wert der P -ten Konstellation in Blickrichtung ϑ_n beschreibt. Anhand dieser Mittelwerte wird entschieden, welche Konstellation den geringsten durchschnittlichen Fehler aufweist und für die weiteren Tests verwendet wird. In Abbildung 4.3 sind die mittleren Fehler der verschiedenen Entwürfe, zusammen mit dem MSE-Durchschnitt des HRTF-basierten RLSFI-Beamformers, gegeneinander aufgetragen. Die Grafik zeigt, dass der Entwurf für Polynomgrad $P = 4$ die geringste mittlere Abweichung zur angestrebten Lösung aufweist. Zusätzlich ist zu bemerken, dass bei

gleichem Abstand $\Delta\vartheta$ zwischen den Stützstellen, eine größere Anzahl an Stützstellen nicht unbedingt zu einem besseren Ergebnis führt.

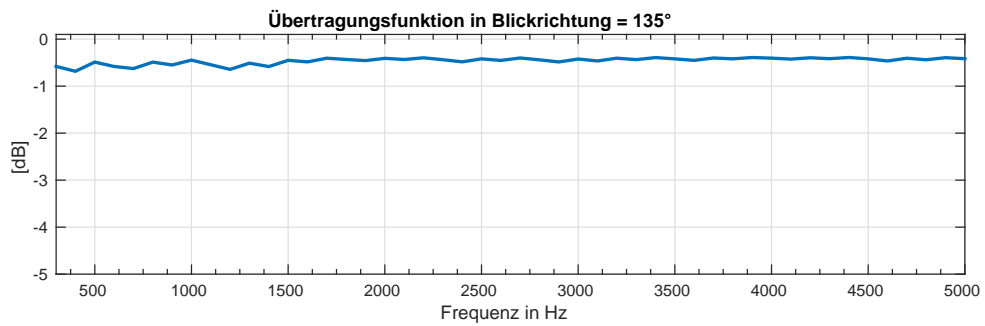
4.3 Signalunabhängige Evaluierung

Dieser Teil der Evaluierung befasst sich mit signalunabhängigen Charakteristiken zur Qualitätsbeurteilung des HRTF-basierten RLSFI-Polynombeamformers. Diese Beurteilung wird anhand eines Vergleichs zwischen diesem und dem HRTF-basierten RLSFI-Beamformer erstellt. Der Vergleich basiert auf den in Kapitel 2.2.1 beschriebenen Merkmalen, wie der Auswertung der Beampattern, dem direkten Vergleich der Übertragungsfunktion in die gewünschte Blickrichtung und des WNG. Zusätzlich werden die Übertragungsfunktionen der Beamformer an bestimmten Frequenzen ausgewertet. Um ein aussagekräftiges Ergebnis zu erhalten, wird dieser Vergleich für zwei verschiedene Blickrichtungen durchgeführt. Dabei entspricht eine dieser Blickrichtungen, einer Stützstelle des HRTF-basierten RLSFI-Polynombeamformers, während die andere zwischen zwei Stützstellen liegt. Der HRTF-basierte RLSFI-Beamformer wird für beide Richtungen optimiert, während der HRTF-basierte RLSFI-Polynombeamformer, mit Polynomgrad $P = 4$ und äquidistanten Stützstellen, beide Blickrichtungen interpoliert. Als Blickrichtungen werden $\vartheta = 135^\circ$, welche einer Stützstelle des HRTF-basierten RLSFI-Polynombeamformers entspricht und $\vartheta = 120^\circ$, welche nicht als Stützstelle enthalten ist, gewählt. Durch diese Wahl der Blickrichtungen wird das Verhalten des HRTF-basierten RLSFI-Polynombeamformers sowohl in eine optimierte Blickrichtung, als auch in eine nicht optimale Blickrichtung, untersucht. Im ersten Schritt werden die Ergebnisse des Versuchs mit Blickrichtung nach $\vartheta = 135^\circ$ analysiert.

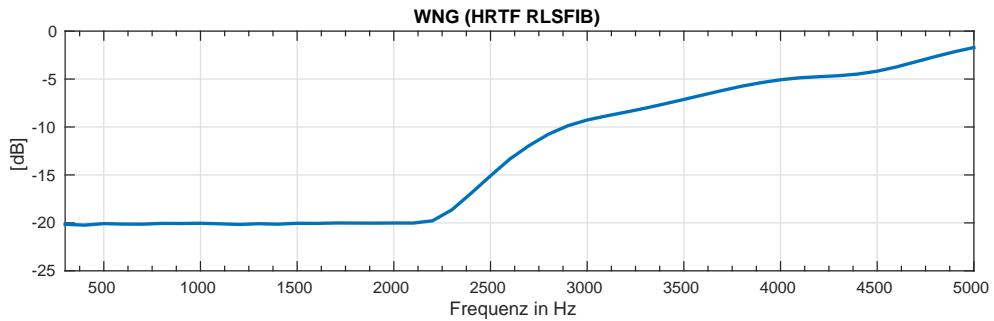
In Abbildung 4.4 werden die Versuchsergebnisse des HRTF-basierten RLSFI-Beamformer und in Abbildung 4.5, die Ergebnisse des HRTF-basierten RLSFI-Polynombeamformer in Blickrichtung von jeweils 135° , dargestellt. Anhand dieser Grafiken, lassen sich optisch kaum Unterschiede zwischen den Ergebnissen erkennen. Sowohl das Beampattern aus Abbildung 4.4 (a), als auch das Beampattern aus Abbildung 4.5 (a) besitzen



(a) Beampattern



(b) Übertragungsfunktion in Blickrichtung



(c) WNG

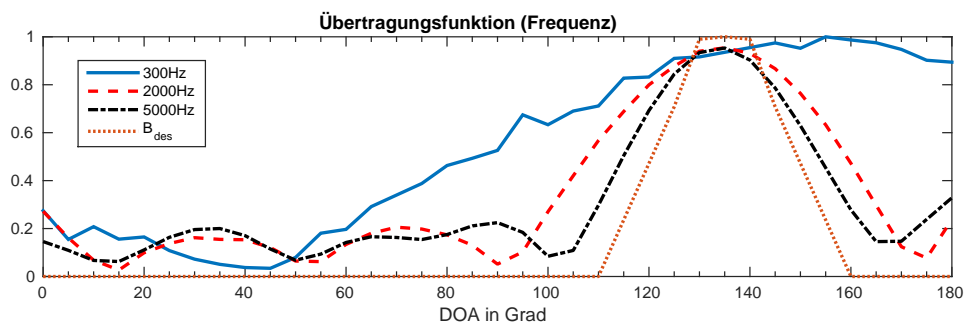
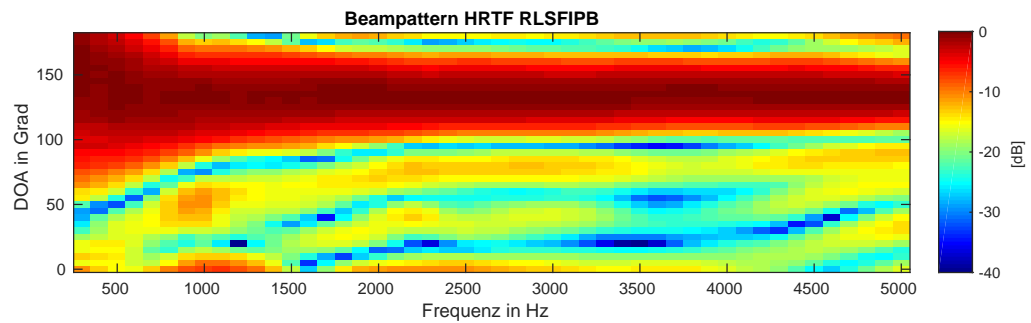
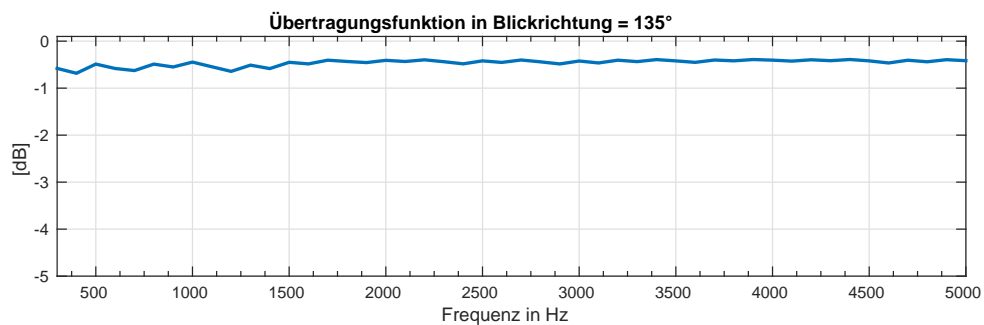
(d) Gewünschte Übertragungsfunktion, verglichen mit realer Übertragungsfunktion, ausgewertet für Frequenzen $\omega = \{300\text{Hz}, 2000\text{Hz}, 5000\text{Hz}\}$

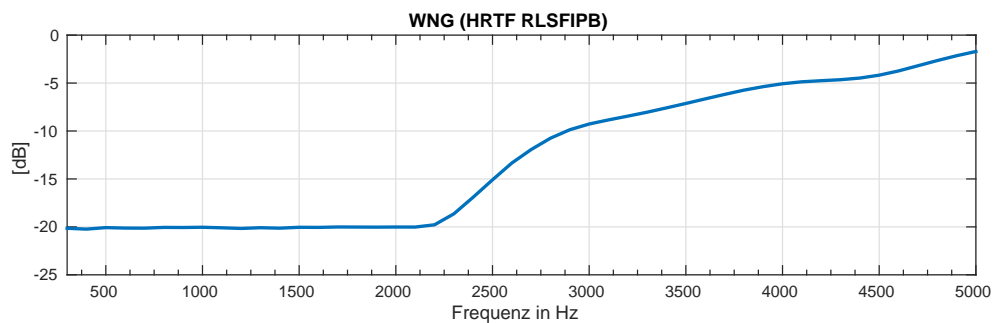
Abbildung 4.4: Versuchsergebnisse des HRTFbasierten RLSFI-Beamformer in Blickrichtung von 135°.



(a) Beampattern



(b) Übertragungsfunktion in Blickrichtung



(c) WNG

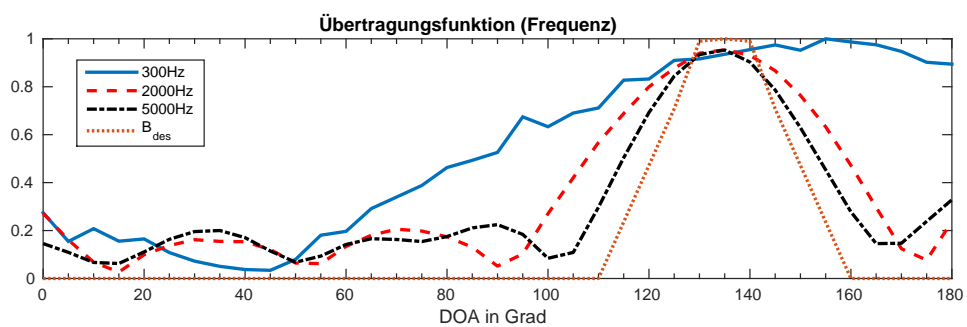
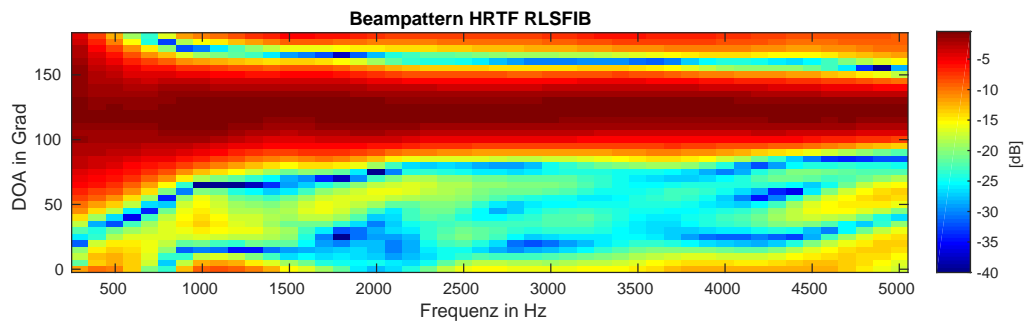
(d) Gewünschte Übertragungsfunktion, verglichen mit realer Übertragungsfunktion, ausgewertet für Frequenzen $\omega = \{300\text{Hz}, 2000\text{Hz}, 5000\text{Hz}\}$

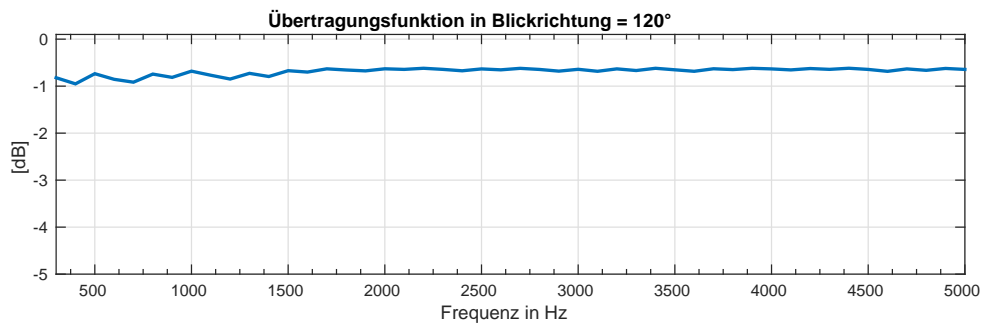
Abbildung 4.5: Versuchsergebnisse des HRTFbasierten RLSFI-Polynombeamformer in Blickrichtung von 135°.

ihr Maximum, entlang der Frequenzachse, bei 135° . Bei niedrigeren Frequenzen, weisen diese einen breiteren Durchlassbereich auf, welcher bei größeren Frequenzen schmaler wird. Dieses Verhalten der Übertragungsfunktion, wird durch Abbildung 4.4 (d) und Abbildung 4.5 (d), bestätigt. Diese zeigen jeweils einen breiteren Durchlassbereich bei 300Hz und einen schmäleren, jedoch mit einigen Nebenmaxima, bei 2000Hz und 5000Hz. Die Übertragungsfunktionen der beiden Entwürfe in die vorgegebene Blickrichtung, dargestellt in Abbildung 4.4 (b) und Abbildung 4.5 (b), bewegen sich zwischen 0dB und -1 dB. Dies weist darauf hin, dass Signale, aus der gewünschten Blickrichtung, nahezu unverzerrt übertragen werden. Die Abweichung, dieser Übertragungseigenschaft, von 0dB, trotz Bedingungen der Verzerrungsfreiheit, definiert in (3.22) und in (3.18), lässt sich auf die Approximation der FIR Filter zurückführen. Das WNG der Versuche, dargestellt in Abbildung 4.4 (c) und Abbildung 4.5 (c), unterschreitet in beiden Fällen nicht die eingestellte WNG-Schranke von -20 dB. Bei niedrigeren Frequenzen befindet sich der WNG-Wert bei -20 dB und steigt für größere Frequenzen auf Werte zwischen 0dB und -5 dB. Ein ähnliches Verhalten der beiden Entwürfe war zu erwarten, da der HRTF-basierte RLSFI-Beamformer für genau diese Blickrichtung optimiert wurde und die gewünschte Blickrichtung als Stützstelle des HRTF-basierten RLSFI-Polynombeamformer enthalten war. Im nachfolgendem Abschnitt werden die Ergebnisse des Versuchs mit Blickrichtung $\vartheta = 120^\circ$ analysiert.

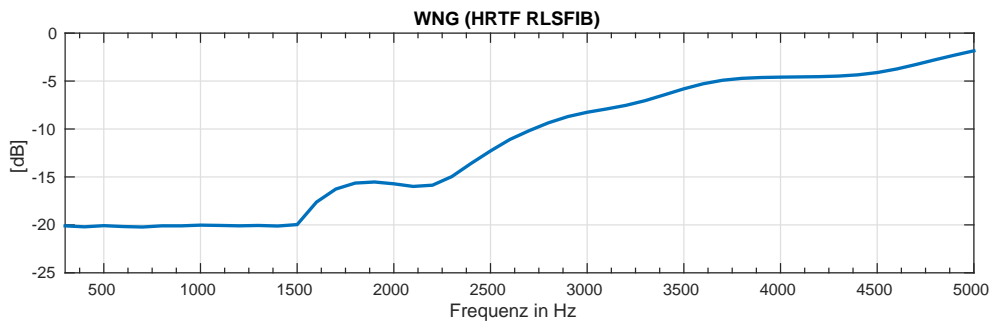
In Abbildung 4.6 und Abbildung 4.7, werden die Versuchsergebnisse des HRTF-basierten RLSFI-Beamformer und des HRTF-basierten RLSFI-Polynombeamformer für die Blickrichtung $\vartheta = 120^\circ$ veranschaulicht. Da diese Blickrichtung keiner optimierten Stützstelle des HRTF-basierten RLSFI-Polynombeamformer entspricht, wird in diesem Versuch dessen Übertragungsverhalten in eine nicht optimale Blickrichtung untersucht. Das Beampattern des HRTF-basierten RLSFI-Beamformers, abgebildet in 4.6 (a), weist ebenfalls, wie in Abbildung 4.4 (a), einen breiteren Durchlassbereich für niedrigere Frequenzen auf, welcher, für größere Frequenzen, schmaler wird. Das Maximum liegt, in diesem Fall, bei $\vartheta = 120^\circ$. Dieses Verhalten wird durch Abbildung 4.6 (d) bestätigt, welche die Übertragungsfunktion, ausgewertet für bestimmte



(a) Beampattern



(b) Übertragungsfunktion in Blickrichtung



(c) WNG

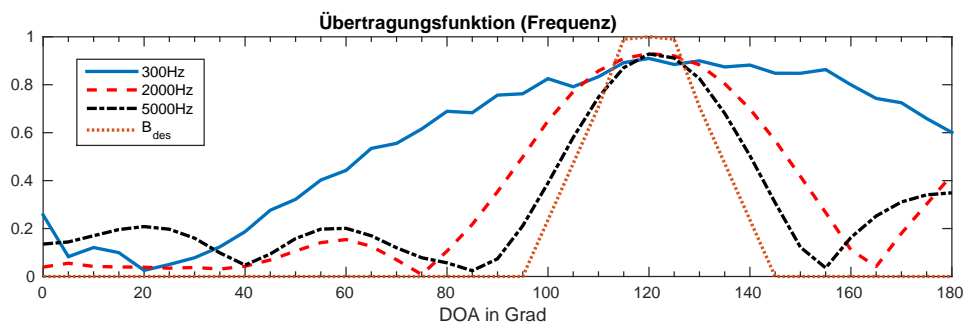
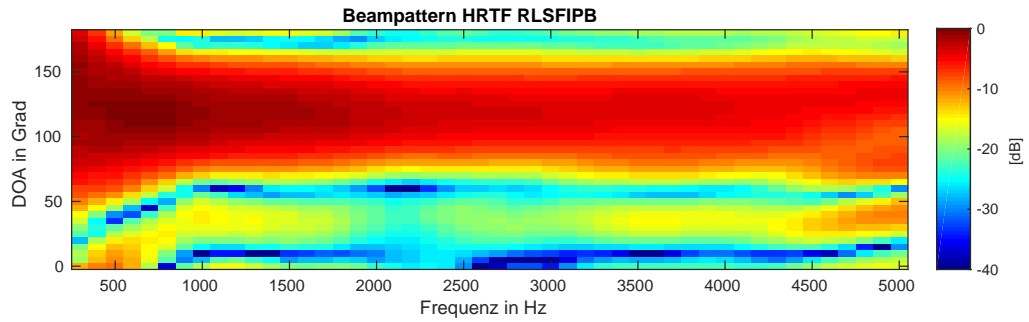
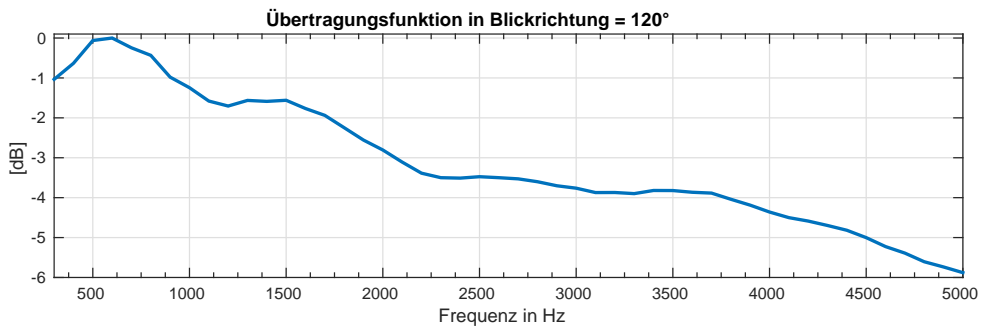
(d) Gewünschte Übertragungsfunktion, verglichen mit realer Übertragungsfunktion, ausgewertet für Frequenzen $\omega = \{300\text{Hz}, 2000\text{Hz}, 5000\text{Hz}\}$

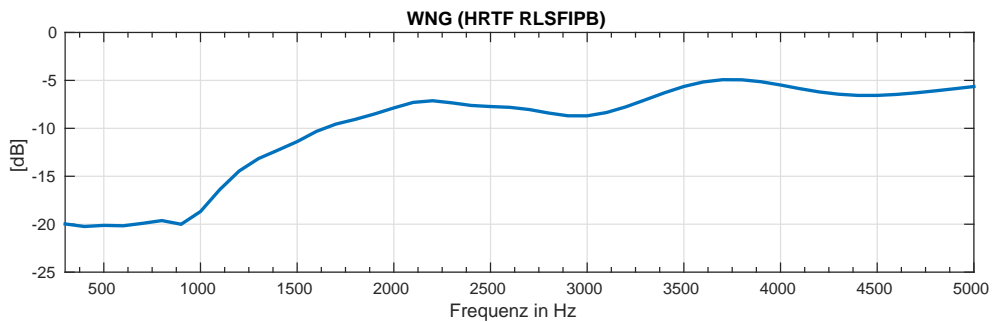
Abbildung 4.6: Versuchsergebnisse des HRTFbasierten RLSFI-Beamformer in Blickrichtung von 120°.



(a) Beampattern



(b) Übertragungsfunktion in Blickrichtung



(c) WNG

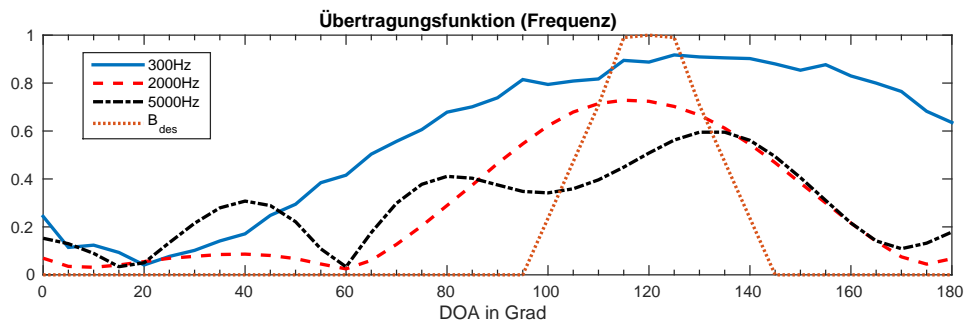
(d) Gewünschte Übertragungsfunktion, verglichen mit realer Übertragungsfunktion, ausgewertet für Frequenzen $\omega = \{300\text{Hz}, 2000\text{Hz}, 5000\text{Hz}\}$

Abbildung 4.7: Versuchsergebnisse des HRTFbasierten RLSFI-Polynombeamformer in Blickrichtung von 120°.

Frequenzen, repräsentiert. Während die Frequenz von 300Hz einen breiteren Durchlassbereich aufweist, besitzen größere Frequenzen einen kleineren Durchlassbereich, jedoch mit zusätzlichen Nebenmaxima. In Abbildung 4.7 wird das Beampattern des HRTF-basierten RLSFI-Polynombeamformer dargestellt. Dieses besitzt ebenfalls einen breiten Durchlassbereich für niedrige Frequenzen, für größer werdende Frequenzen ab 4500Hz wird dieser jedoch diffuser. Das Verhalten dieses Beampatterns, für bestimmte Frequenzen, wird in Abbildung 4.7 (d) veranschaulicht. Signalanteile von 300Hz erfahren einen breiten Durchlassbereich, während Frequenzen von 2000Hz ein schmalerer Durchlassbereich, mit wenig ausgeprägten Nebenmaxima, zugeordnet ist. Verglichen dazu, wird der Durchlassbereich, von Frequenzen um 5000Hz, durch stärker ausgeprägte Nebenmaxima, charakterisiert. Die Übertragungsfunktion des HRTF-basierten RLSFI-Beamformers in die gewünschte Blickrichtung, veranschaulicht durch Abbildung 4.6 (b), bewegt sich zwischen 0dB und -1 dB, was auf eine nahezu verzerrungsfreie Übertragung aus dieser Blickrichtung hindeutet. Die Übertragungsfunktion des HRTF-basierten RLSFI-Polynombeamformers, in die gewünschte Blickrichtung, wird in Abbildung 4.7 (b) dargestellt. Diese zeigt, dass niederfrequente Signalanteile, aus dieser Richtung, nahezu unverzerrt übertragen werden, während Anteile von größer werdender Frequenz, stärker verzerrt werden. Dieses Verhalten zeigt, dass der HRTF-basierte RLSFI-Polynombeamformer für Blickrichtungen, welche nicht als Stützstellen enthalten sind, die Bedingung der Verzerrungsfreiheit für größere Frequenzen nicht mehr optimal erfüllt. Das WNG der beiden Entwürfe, veranschaulicht in Abbildung 4.6 (c) und Abbildung 4.7 (c), unterschreitet in beiden Fällen nicht die eingestellte WNG-Schranke von -20 dB. Während der maximale WNG-Wert des HRTF-basierten RLSFI-Beamformer zwischen 0dB und -5 dB liegt, befindet sich das WNG-Maximum des HRTF-basierten RLSFI-Polynombeamformer bei -5 dB.

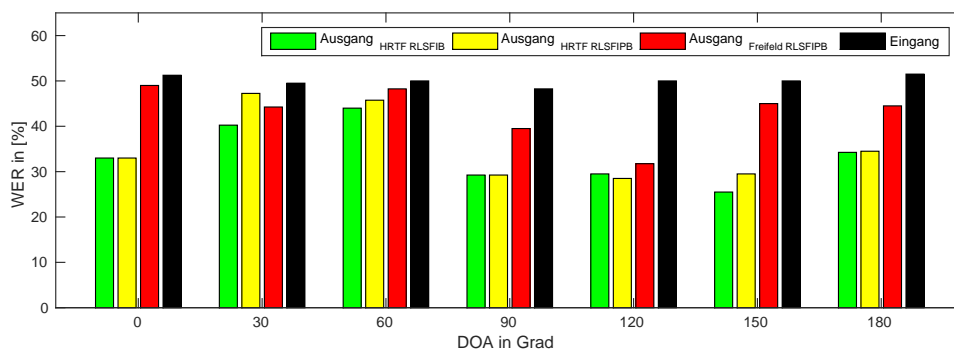
Diese Versuche zeigen, dass der HRTF-basierte RLSFI-Polynombeamformer, für Blickrichtungen, welche unter den gewählten Stützstellen enthalten sind, nahezu gleiche Ergebnisse wie der HRTF-basierte RLSFI-Beamformer liefert. Für Blickrichtungen, welche nicht als Stützstellen enthalten sind, ist nach wie vor eine gute räumliche Un-

terscheidung gegeben, für höherfrequente Signalanteile nimmt die Qualität jedoch ab. Der nachfolgende Abschnitt behandelt die signalabhängige Evaluation des Experiments.

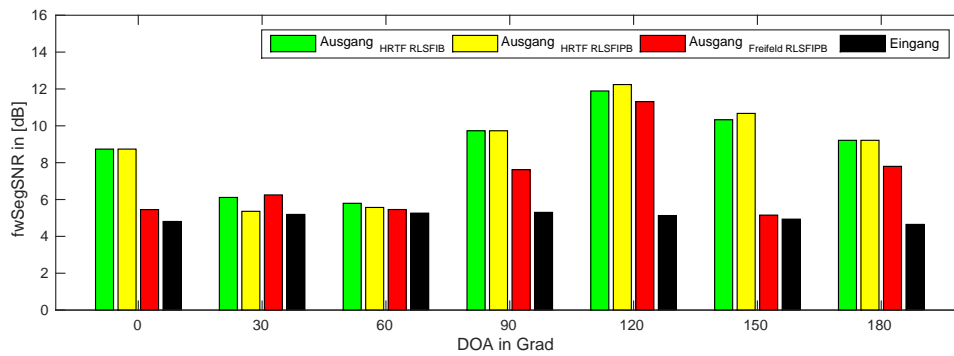
4.4 Signalabhängige Evaluierung

Dieser Teil der Evaluierung befasst sich mit der signalabhängigen Qualitätsbeurteilung des HRTF-basierten RLSFI-Polynombeamformers. Die Beurteilung basiert auf dem Vergleich zwischen diesem, dem HRTF-basierten RLSFI-Beamformer und dem freifeldbasierten RLSFI-Polynombeamformer, dessen Entwurf in Kapitel 3.2 vorgestellt wurde. Dazu werden die HRTF-basierten Entwürfe auf dem Roboter NAO simuliert, während der Versuch, des freifeldbasierten Ansatzes, unter Freifeldbedingungen durchgeführt wird. Als Maß zur Qualitätsbeurteilung werden WER und fwSegSNR herangezogen. Die drei Beamformer werden mit gleichen Signalen an ihren Eingängen angeregt. WER und fwSegSNR werden sowohl an den Ausgängen, als auch am Eingang der Beamformer gemessen, welcher für die betrachteten Entwürfe gleich ist. Um ein aussagekräftiges Ergebnis zu erhalten, werden die Messungen für Signale aus sieben verschiedenen Richtungen, $\vartheta = [0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ]$, durchgeführt.

In Abbildung 4.8, werden die Ergebnisse der WER- und der fwSegSNR-Messung veranschaulicht. Es ist deutlich zu erkennen, dass die WER-Werte, welche in Abbildung 4.8 (a) dargestellt werden, gemessen an den Ausgängen der HRTF-basierten Entwürfe, über sämtliche Blickrichtungen niedriger sind, als am Ausgang des freifeldbasierten RLSFI-Polynombeamformer oder am Eingang. Der Versuch, mit ankommenden Signalen unter einem Vertikalwinkel von $\vartheta = 30^\circ$, ist hier die Ausnahme. Dieses Verhalten schließt auf eine bessere Worterkennungsrate der HRTF-basierten Beamformer. Zusätzlich ist zu vermerken, dass sich die WER-Werte dieser Entwürfe, für die jeweiligen Blickrichtungen, nur gering unterscheiden und für die Blickrichtungen, welche als Stützstellen im HRTF-basierten RLSFI-Polynombeamformer, $\vartheta = [0^\circ, 90^\circ, 180^\circ]$, enthalten sind, kein Unterschied festzustellen ist. Die gemessenen fwSegSNR-Werte, welche in Abbildung



(a) WER



(b) fwSegSNR

Abbildung 4.8: WER und fwSegSNR, ausgewertet am Eingang und an den Ausgängen des HRTF-basierten RLSFI-Beamformer, des HRTF-basierten RLSFI-Polynombeamformer und des Freifeld basierten RLSFI-Polynombeamformer.

4.8 (b) veranschaulicht werden, deuten auf eine ähnliche Qualitätsbeurteilung hin. Die $fwSegSNR$ -Werte, gemessen an den Ausgängen der HRTF-basierten Ansätze, sind über sämtliche Blickrichtungen größer als die gemessenen Werte am Eingang und Ausgang des freifeldbasierten Polynombeamformer. Die Ausnahme bilden in diesem Fall ebenfalls, Signale, welche unter einem Vertikalwinkel von 30° , eintreffen. Auch hier unterscheiden sich die Werte zwischen den HRTF-basierten Entwürfen für die jeweiligen Blickrichtungen nur gering und sind gleich für Blickrichtungen, welche als Stützstellen im HRTF-basierten RLSFI-Polynombeamformer enthalten sind.

Diese Ergebnisse zeigen, dass der HRTF-basierte RLSFI-Polynombeamformer, trotz Interpolationsfehler, ähnlich gute Resultate im Vergleich zu dem HRTF-basierten RLSFI-Beamformer erzielt. Der freifeldbasierte RLSFI-Beamformer erlangte, unter den drei evaluierten Beamformeransätzen, durchschnittlich die schlechtesten Resultate. Das bedeutet, dass die getestete Kombination von HRTFs und RLSFI-Polynombeamformer, verglichen mit dem freifeldbasierten RLSFI-Polynombeamformer unter Freifeldbedingungen, eine Qualitätssteigerung nach sich zieht.

Kapitel 5

Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Polynombeamformer evaluiert, welcher den Kopf des Roboters NAO als Verzerrer mit berücksichtigt. Neben dem Entwicklungsverfahren dieses Beamformers wurden signalabhängige und signalunabhängige Methoden zur Qualitätsbeurteilung von Beamformern vorgestellt. Diese Methoden wurden verwendet, um das Verhalten des HRTF-basierten RLSFI-Polynombeamformer zu testen. Um diese Ergebnisse einordnen zu können, wurde der HRTF-basierte RLSFI-Beamformer als Referenz heran gezogen. Da dieser, für sämtliche Tests, in die jeweiligen Blickrichtungen individuell optimiert wurde, konnten dessen Ergebnisse als Optimalresultate angesehen und mit den Ergebnissen des HRTF-basierten RLSFI-Polynombeamformer verglichen werden. In der signalunabhängigen Evaluation wurde gezeigt, dass der HRTF-basierte RLSFI-Polynombeamformer für niederfrequente Signalanteile ein ähnliches Verhalten wie der Referenzbeamformer zeigte. Für höhere Frequenzen war, im Gegensatz zu den Referenzergebnissen, ein diffuseres Übertragungsverhalten für nicht optimierte Blickrichtungen zu beobachten. Diese Abweichung lässt sich durch einen durch die Interpolation verursachten Fehler begründen. Die signalabhängigen Tests zeigen, dass die Ergebnisse des HRTF-basierten RLSFI-Polynombeamformer nur gering von den Referenzwerten abweichen. Dies lässt darauf schließen, dass dieser Beamformer eine ähnliche Performanz im Vergleich zu dem HRTF-basierten RLSFI-Beamformer liefert. Zusätzlich besitzt dieser die Eigenschaft seine Blickrichtung dynamisch ändern zu können,

welche im Kontext von robotischem Hören, sehr wichtig ist.

In den nächsten Schritten, sollte getestet werden, ob die Performanz des HRTF-basierten RLSFI-Polynombeamformers für die reale Anwendung auf dem Roboter NAO ausreicht oder ob weitere Schritte zur Optimierung der Signalextraktion notwendig sind.

Abbildungsverzeichnis

2.1	Model eines FSB [6].	4
2.2	Optimales und reales Beampattern, entworfen für die Blickrichtung von 90°	7
2.3	$B_{act}(\omega, \vartheta_{des} = 90^\circ)$ von Abbildung 2.2 (b).	8
2.4	WNG von Abbildung 2.2 (b).	9
2.5	Schematischer Aufbau eines Polynombeamformer [3].	10
3.1	Beampattern mit $\vartheta_{des} = 90^\circ$ des freifeldbasiertem RLSFI-Beamformer, angewendet auf den Roboter NAO.	17
4.1	markierte Mikrofone des Roboters NAO [2].	21
4.2	MSE-Auswertung des HRTF-basierten RLSFI-Polynombeamformer für Polynomgrad 4 und des HRTF-basierten RLSFI-Beamformer.	24
4.3	Durchschnittlicher MSE der einzelnen Konstellationen des HRTF-basierten RLSFI-Polynombeamformer, verglichen mit dem HRTF-basierten RLSFI-Beamformer.	24
4.4	Versuchsergebnisse des HRTF-basierten RLSFI-Beamformer in Blickrichtung von 135°	26
4.5	Versuchsergebnisse des HRTF-basierten RLSFI-Polynombeamformer in Blickrichtung von 135°	27
4.6	Versuchsergebnisse des HRTF-basierten RLSFI-Beamformer in Blickrichtung von 120°	29

4.7	Versuchsergebnisse des HRTFbasierten RLSFI-Polynombeamformer in Blickrichtung von 120°	30
4.8	WER und fwSegSNR, ausgewertet am Eingang und an den Ausgängen des HRTF-basierten RLSFI-Beamformer, des HRTF-basierten RLSFI- Polynombeamformer und des Freifeld basierten RLSFI-Polynombeamformer.	33

Tabellenverzeichnis

- 4.1 Übersicht über die evaluierten Polynomgrade mit zugehörigen Stützstellen. 23

Literaturverzeichnis

- [1] Aldebaran Robotics. *NAO NEXT Gen H25 Datasheet*, December 2011.
- [2] H. Barfuss, C. Hümmer, A. Schwarz, and W. Kellermann. Hrtf-based robust least-squares frequency-invariant beamforming. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2015.
- [3] E. Mabande and W. Kellermann. Design of robust polynomial beamformers as a convex optimization problem. In *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel-Aviv, Israel, August 2010.
- [4] I.A. McCowan. *Robust Speech Recognition using Microphone Arrays*. PhD thesis, Queensland University of Technology, Australia, 2001.
- [5] E. Mabande, A. Schad, and W. Kellermann. Design of robust superdirective beamformers as a convex optimization problem. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 77–80, Taipei, Taiwan, April 2009.
- [6] Michael Bürger. Sectorial optimization of robust polynomial beamformers for uniformly spaced arrays, September 2011. SIM Project.
- [7] Edwin T. Mabande. *Robust Time-Invariant Broadband Beamforming as a Convex Optimization Problem*. PhD thesis, University of Erlangen-Nuremberg, Multimedia Communications and Signal Processing, 2014.
- [8] D. Huggins-Daines, M. Kumar, A. Chan, A.W. Black, M. Ravishankar, and A.I. Rudnicky. Pocketsphinx: A free, real-time continuous speech recognition system

- for hand-held devices. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 1, pages I–I, May 2006.
- [9] Yi Hu and P.C. Loizou. Evaluation of objective quality measures for speech enhancement. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(1):229–238, Jan 2008.