

TECHNISCHE UNIVERSITÄT BERLIN

Fakultät IV - Elektrotechnik und Informatik

Institut für Telekommunikationssysteme
Fachgebiet Nachrichtenübertragung

Studienarbeit

von

Rüdiger Knörig

Matr.-Nr. 178769

*„Einsatz des Root-Music-Algorithmus zur akustischen
Lokalisierung und Verfolgung multipler Quellen“*

Hochschullehrer: Prof. Dr.-Ing. Peter Noll

Betreuer: Dipl.-Ing. Markus Schwab

Zusammenfassung

Ziel dieser Studienarbeit ist die Implementierung und Evaluierung des Breitband-Root-MUSIC¹-Algorithmus zur Lokalisierung multipler akustischer Quellen durch ULA²-Mikrofonarrays im Rahmen des Sensing-People-Projektes.

Aus einer genaueren Beschreibung der Problemstellung heraus werden zunächst mögliche Lösungsansätze diskutiert. Im Anschluß daran werden die theoretischen Grundlagen des Unterraum-Verfahrens *MUSIC* zuerst unter der Annahme schmalbandiger Signale, danach mit einem anderen Ansatz auch für breitbandige Signale erarbeitet. Neben den Verfahren an sich wird dabei auch auf Einschränkungen bzw. Problemfälle eingegangen, welche für den praktischen Einsatz der Algorithmen von Relevanz sind. Zum Abschluß des theoretischen Teils dieser Ausarbeitung wird die Ableitung des *root-MUSIC*-Verfahrens aus dem grundlegenden *MUSIC*-Algorithmus erläutert.

Nach einer Beschreibung der Implementierung (mit Schwerpunkt auf die Signalverarbeitung) und Abschätzung der Komplexität folgt ihre Evaluierung unter idealen Bedingungen, für einen simulierten Problemfall und mit den Signalen eines realen Mikrofonarrays.

¹Multiple signal classification

²uniform linear array

Inhaltsverzeichnis

1	Einleitung	5
2	Theoretische Einführung	6
2.1	Problemstellung	6
2.1.1	Lösungsansätze	6
2.2	Schmalband-MUSIC	8
2.2.1	Das Signalmodell	8
2.2.2	Notwendigkeit der Bandbreitenbegrenzung	10
2.2.3	Herleitung	10
2.2.4	Zusammenhang zwischen der Mikrofonanzahl und der Anzahl lokalisierbarer Quellen.	15
2.2.5	Schätzung der Korrelationsmatrix	15
2.3	Das Fernfeld-Signalmodell	16
2.3.1	Abschätzung der nötigen Entfernung	16
2.3.2	Parametrisierung für den Schmalband-Ansatz	17
2.4	Breitband-MUSIC	19
2.4.1	Theoretische Herleitung	19
2.4.2	Spatial smoothing	23
2.5	Root-MUSIC	23
2.5.1	Theoretische Herleitung	24
2.5.2	Berechnung des Polynoms	25
3	Die Implementierung	27
3.1	Beschreibung	27
3.2	Parameterisierung	31
3.3	Rechenzeitaufwand	32
4	Evaluierung	35
4.1	ideale Signale	35
4.2	Problemfall Raumechos	38
4.3	reale Signale	44
4.4	Zusammenfassung	46
4.5	Ausblick	46

A Anhang	49
A.1 Beweise	49

Symbolverzeichnis

- I** Einheitsmatrix ($i_{jk} = \begin{cases} 0 & i \neq j \\ 1 & i \equiv j \end{cases}$)
- R_{XY}** Korrelationsmatrix $E\{\vec{X} \cdot \vec{Y}^H\}$.
- A^H** hermitische Matrix von **A**; transponierte Matrix der konjugiert-komplexen Elemente von **A**
- RxyDim** Dimension der spektralen Kovarianzmatrix.
- σ_N^2 Varianz / Leistung des Rauschens $n_k(t)$ an den Mikrofonen.
- Θ Winkel zwischen Wellenvektor \vec{k} und dem Mikrofonarray (s. Abbildung 2.6)
- \vec{k} Wellenvektor, $|\vec{k}| = \frac{\omega}{c}$.
- $\vec{n}(t)$ Vektor der additiven Rauschkomponenten der Mikrofon-signale zum Zeitpunkt t
- $\vec{s}(t)$ Vektor der Quellensignale zum Zeitpunkt t
- $\vec{x}(t)$ Vektor der Mikrofon-signale zum Zeitpunkt t
- d* Abstand der Mikrofone im ULA
- $E\{\dots\}$ Erwartungswert-Operator
- M* Anzahl der Mikrofone
- $n_k(t)$ Das vom *k*-ten Mikrofon zum Zeitpunkt t aufgenommene additives Rauschen.
- Q* Anzahl der Quellen
- $s_k(t)$ Signal der *k*-ten Quelle zum Zeitpunkt t.
- $x_k(t)$ Das vom *k*-ten Mikrofon zum Zeitpunkt t aufgenommenen Signal.

z^* z konjugiert-komplex

CPSM Cross Power Spectrum Matrix - engl. Akronym für die Kreuzleistungsdichtespectralmatrix

DOA Direction Of Approach, s. Einfallswinkel Θ

FFT Fast Fourier Transformation - Ein von Carl Friedrich Gauß entwickelter Algorithmus zur effizienten Berechnung der diskreten Fouriertransformation.

GCC Generalized cross Correlation

MUSIC multiple signal classification

SVD Singular Value Decomposition - Zerlegung einer $m \times n$ -Matrix \mathbf{A} gemäß

$$\mathbf{A} = \mathbf{U} \cdot \lambda \cdot \mathbf{V}^T \quad (1)$$

mit

$$\begin{aligned} \mathbf{U} &: \text{Matrix der linken Eigenvektoren von } \mathbf{A} \\ \lambda &: \text{Diagonalmatrix der Eigenwerte } \lambda_i \\ \mathbf{V} &: \text{Matrix der rechten Eigenvektoren von } \mathbf{A} \end{aligned} \quad (2)$$

ULA uniform linear array - Mikrofonarray mit Orientierung der Mikrofone entlang einer Linie in äquidistanten Abständen

Kapitel 1

Einleitung

Diese Studienarbeit entstand im Rahmen der Beteiligung der TU Berlin am Sensing People Projekt (SPP). Im Rahmen dieses Projekts sollen Methoden erarbeitet werden, um Personen durch audiovisuelle Sensoren erfassbar zu machen. Das Fachgebiet Nachrichtenübertragung der TU Berlin erforscht dabei Methoden zur Richtungsdetektion und zum Tracking¹ von sprechenden Personen mit Mikrofonarrays. Angestrebt wird eine Lösung, welche

- Mehrere Personen simultan erfassen kann.
- Eine hohe Ortsauflösung besitzt.
- Robust gegenüber Störgeräuschen ist.
- Robust gegenüber Reflektionen ist.
- Robust gegenüber Abweichungen in der Charakteristik und genauen Platzierung der Sensoren ist.
- In Echtzeit auf Hardware mit akzeptablen Preisen zu implementieren ist.

Literaturrecherchen ergaben die im folgenden Kapitel angerissenen Lösungsansätze. Von diesen wurde die GCC²-Methode aufgrund ihrer Einfachheit zuerst implementiert. Allerdings hat dieses Verfahren, wie im nächsten Kapitel dargestellt wird, Schwächen in der Erkennung mehrerer gleichzeitig aktiver Sprecher. Aus diesem Grund wurde das komplexere Unterraumverfahren *Root-MUSIC* im Rahmen dieser Studienarbeit implementiert und evaluiert, um es auf seine Eignung zur Lösung des gestellten Problems untersuchen zu können.

¹im Sinne der kontinuierlichen Erfassung der Positionen

²*generalized cross correlation*

Kapitel 2

Theoretische Einführung

2.1 Problemstellung

Im Rahmen dieser Studienarbeit soll das Problem der robusten Lokalisierung von akustischen Quellen mit ULA-Mikrofonarrays durch Anwendung des *root-MUSIC*-Algorithmus gelöst werden. Die Anzahl und die Signalstatistik der Quellen sind dabei als unbekannt anzunehmen. Schematisch ist dieses Problem in Abbildung 2.1 dargestellt. Aus Gründen der Übersichtlichkeit wurde in dieser Darstellung nur ein ULA-Mikrofonarray und für dieses auch nur die direkten Signalfade zwischen den Quellen und dem ersten Mikrofon eingezeichnet. Die Ausbreitung des Schalls wird vereinfachend unter Fernfeldbedingungen (s. Abschnitt 2.3) modelliert, die Lokalisierung beschränkt sich dann auf die Bestimmung des Einfallswinkels Θ (im folgenden mit **DOA**¹ bezeichnet). Die Lokalisierung der Quelle im Raum kann durch Kreuzpeilungen zwischen mehreren Mikrofonarrays erfolgen.

2.1.1 Lösungsansätze

Für das gegebene Problem sind folgende Lösungsansätze bekannt:

Korrelationsbasierende Verfahren Diese Verfahren bestimmen die Verschiebungen der Mikrofonsignale zueinander, welche aus den Maxima der Kreuzkorrelation bestimmt werden können, und rechnen diese in die gesuchten Einfallswinkel um. Zur Erzielung besserer Ergebnisse wird die Kreuzkorrelation meist im Frequenzbereich gemäß $r_{xy}(k) = \mathcal{F}^{-1} \{X(j\omega) \cdot Y^*(j\omega)\}$ bestimmt. Die Signalspektren werden vorher in geeigneter Weise transformiert, um die durch Quellensignalkorrelationen entstandene Maxima zu minimieren (sog. GCC² - Verfahren). Nachteilig ist, daß der Fall mehrerer gleichzeitig aktiver Quellen im Ansatz nicht berücksichtigt wird - zwar

¹Direction Of Approach

²Generalized Cross Correlation

sind in diesem Fall Maxima an den entsprechenden Stellen zu erwarten, nur müssen diese von durch Korrelationen der Quellensignale untereinander entstandenen Peaks ($\hat{=}$ lokale Maxima) separiert werden.

Räumliche Suche mittels Beamformer Das Mikrofonarray wird als SAA³ verwendet, d.h. die Mikrofon-signale werden derart nachbearbeitet, daß die Gesamtanordnung wie ein Richtmikrofon mit steuerbarer Richtcharakteristik wirkt. Die Lokalisierung von Quellen geschieht hierbei durch die Ermittlung der Maxima der Signalleistung des Beamformer-Ausgangssignals in Abhängigkeit vom Peilwinkel Θ . Entscheidend für diesen Ansatz ist die räumliche Trennschärfe des Beamformers. Eine schmalere „Keule“ wird meist um den Preis von Nebenmaxima erkauft. Zudem ist die Funktion der räumlichen Selektivität frequenzabhängig. Da menschliche Sprache recht breitbandig ist, müsste dies durch entsprechende Korrekturmaßnahmen berücksichtigt werden.

Parametrische Methoden Parametrische Verfahren wie der bekannte *maximum-likelihood*-Ansatz bedingen Vorwissen über die Signale und ihre Eigenschaften. Vorteilhaft ist eine geringere Varianz in den Positionsschätzungen, wobei aber auch eine Konkretisierung auf die durch die genaueren Annahmen beschriebene Situation und damit ein Verlust an Flexibilität (und Robustheit gegenüber Abweichungen zu den Annahmen) in Kauf genommen werden muß.

Unterraumbasierende Verfahren (*subspace*-Verfahren, z.B. MUSIC) Unterraum-Verfahren versuchen, orthogonale Signal- und Rauschräume für quellenpositionsabhängige Größen (z.B. die auf Seite 9 eingeführten Peil- bzw. *steering*-Vektoren) zu bilden. Durch Kolinearitätsmessungen mit den Basisvektoren dieser Räume kann ein durch Positionsangaben parameterisierter generischer Vektor einem dieser Räume zugeordnet werden. Vorteilhaft bei diesen Verfahren ist seine Flexibilität, da kaum Annahmen über die Signalausbreitung bzw. die Signale an sich gemacht werden, und eine höhere Trennschärfe gegenüber den Beamformern.

Für diese Studienarbeit wurde eine Variante des MUSIC-Algorithmus, also ein Vertreter der Subspace-Verfahren, ausgewählt. Für die Verwendung von MUSIC zur Lösung der gestellten Aufgabe gibt es zwei verschiedene Ansätze: Schmal- und Breitband-MUSIC. Die theoretischen Grundlagen beider Ansätze sollen im folgenden genauer dargestellt werden.

³synthetic aperture antenna

2.2 Schmalband-MUSIC

2.2.1 Das Signalmodell

Zur Erschließung dieses Lösungsansatzes ist es sinnvoll, die Entstehung der von den Mikrofonen gelieferten Eingangssignale zu modellieren. Dabei soll das in Abbildung 2.1 dargestellte Szenario als Grundlage dienen. Für die einleitenden Betrachtungen sollen folgende Vereinfachungen gelten:

- Es werden nur *direkte Signalpfade* betrachtet; auf Reflektionen wird zu einem späteren Zeitpunkt eingegangen.
- Die *Quellensignale* sollen als *unkorreliert* angenommen werden.
- Die Rauschkomponenten der Mikrofonsignale werden durch additives unkorreliertes Rauschen modelliert.

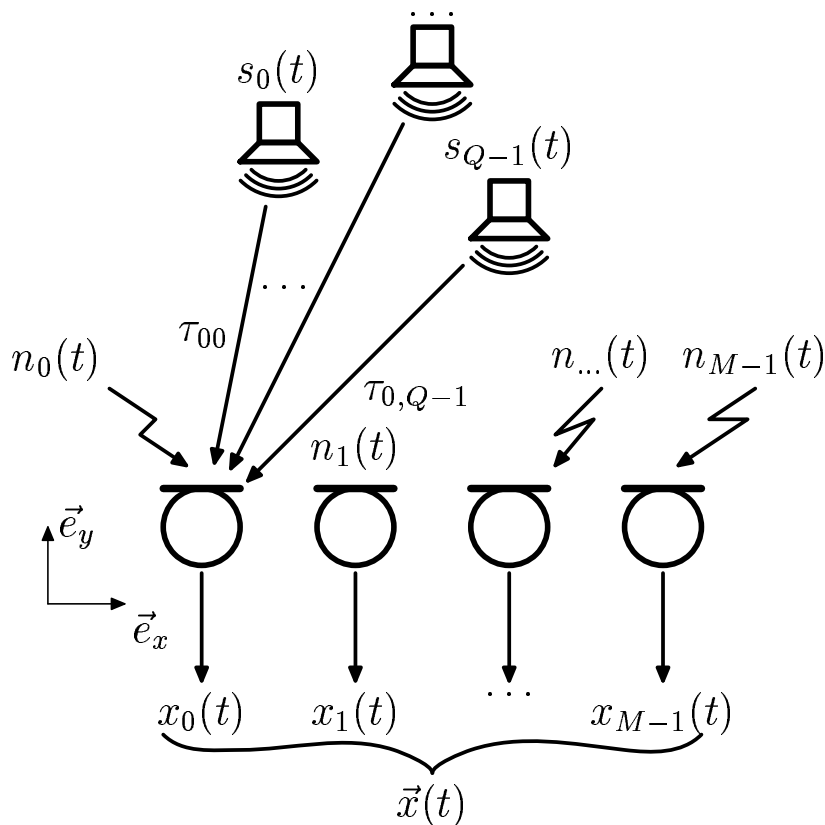


Abbildung 2.1: Das Umgebungsmodell

Mit diesen Annahmen läßt sich das Ausgangssignal des k -ten Mikrofons als Summe der mit α_{lk} gedämpften und mit der Laufzeit τ_{lk} verzögerten Quellensignale

plus einem Rauschsignal $n_k(t)$ beschreiben (s. Abb. 2.1):

$$x_k(t) = \sum_{l=0}^{Q-1} \alpha_{lk} \cdot s_l(t - \tau_{lk}) + n_k(t) \quad (2.1)$$

$$= \sum_{l=0}^{Q-1} \alpha_{lk} \cdot s_l(t) * \delta(t - \tau_{lk}) + n_k(t) \quad (2.2)$$

$$\text{mit } Q \hat{=} \text{Anzahl der Quellen.} \quad (2.3)$$

Transformiert man Gleichung 2.3 in den Frequenzbereich, so erhält man

$$\begin{aligned} X_k(j\omega) &= \sum_{l=0}^{Q-1} \alpha_{lk} \cdot \vec{S}_l(j\omega) \cdot e^{-j\omega\tau_{lk}} + N_k(j\omega) \\ &= N_k(j\omega) + \vec{S}(j\omega) \cdot [\alpha_{0k} \cdot e^{-j\omega\tau_{0k}}, \alpha_{1k} \cdot e^{-j\omega\tau_{1k}}, \dots, \alpha_{(Q-1)k} \cdot e^{-j\omega\tau_{(Q-1)k}}] \end{aligned} \quad (2.4)$$

wobei

$$\vec{S} = \{S_0(j\omega), S_1(j\omega), \dots, S_{Q-1}(j\omega)\} \quad (2.5)$$

der Vektor der Quellensignalspektren ist. Vorteilhaft bei dieser Darstellung ist, daß das Ausgangssignal der Mikrofone nicht mehr durch Faltung wie in Gleichung 2.3 sondern durch multiplikative Verknüpfung gebildet wird; somit läßt sich Gleichung 2.4 zusammenfassend für alle Mikrofonsignalspektren auch durch eine Matrix-Vektor-Multiplikation ausdrücken.

$$\vec{X}(j\omega) = \mathbf{A} \cdot \vec{S}(j\omega) + \vec{N}(j\omega) \quad (2.6)$$

mit

$$\mathbf{A} = \begin{pmatrix} e^{-j\omega\tau_{00}} & e^{-j\omega\tau_{01}} & \dots & e^{-j\omega\tau_{0(Q-1)}} \\ e^{-j\omega\tau_{10}} & e^{-j\omega\tau_{11}} & \dots & e^{-j\omega\tau_{1(Q-1)}} \\ \dots & \dots & \dots & \dots \\ e^{-j\omega\tau_{(M-1)0}} & e^{-j\omega\tau_{(M-1)1}} & \dots & e^{-j\omega\tau_{(M-1)(Q-1)}} \end{pmatrix}, \text{ wenn } \alpha_{lk} \approx 1 \quad (2.7)$$

Man beachte, daß der k -te Spaltenvektor aus \mathbf{A} die Reaktion des Arrays auf das Signal der k -ten Quelle beschreibt; man könnte deshalb auch von „Peilvektoren“ sprechen (diese entsprechen in ihrer Bedeutung den von den Beamformern her bekannten *steering*-Vektoren [KV96]).

2.2.2 Notwendigkeit der Bandbreitenbegrenzung

Gleichung 2.4 setzt implizit voraus, daß zu jedem Zeitpunkt t_0 die Spektren der Mikrofon-signale, abgesehen von der additiven Rauschkomponente $N_k(j\omega)$, Überlagerungen von phasenverschobenen Versionen der Quellensignalspektren $S_l(j\omega)$ sind. Damit diese Annahme haltbar ist, dürfen die Quellensignale sich nicht zu schnell ändern, müssen also dementsprechend schmalbandig sein. Exakter formuliert muß

$$\vec{S}(j\omega)|_{t_0} \approx \vec{S}(j\omega)|_{t_0+\Delta t} \text{ gelten.} \quad (2.8)$$

Dabei ist Δt die Zeit, welche eine Wellenfront maximal benötigt, um nach dem Passieren des ersten Mikrofons des Arrays das letzte Mikrofon zu erreichen – unter der Annahme von Fernfeldbedingungen also $\Delta t_{\max} = M \cdot d/c$ (s. Abbildung 2.6). Die maximale Änderungsfrequenz und damit auch die Bandbreite der Quellensignale beträgt somit

$$B_{\max} = \frac{1}{\Delta t_{\max}} = \frac{c}{M \cdot d}. \quad (2.9)$$

Für ein Mikrofonarray mit 4 Mikrofonen im Abstand von 10 cm und einer Schallgeschwindigkeit von $c = 300 \text{ m} \cdot \text{s}^{-1}$ liegt die maximale Bandbreite demnach bei 750 Hz.

Umgesetzt werden kann die Bandbreitenbegrenzung durch:

1. Filterung im Zeitbereich durch ein Bandpaßfilter.
2. Multiplikation der Spektren mit einem Bandpaßspektrum, also die klassische Filterung im Frequenzbereich.
3. Verwendung der Analogie von Teilbandcodierung und Transformationscodierung: Die FFT-Koeffizienten lassen sich auch als Ausgänge einer Filterbank sehen; verwendet man nur einen Koeffizienten aus dem FFT-Spektrum, so läßt sich dieser auch als komplexer Ausgangswert eines Bandpaßfilters sehen. Die weiteren Betrachtungen laufen dann quasi im Zeitbereich ab, d.h. von diesem Bandpaßsignal wird die Kovarianzmatrix ermittelt (welcher dann der im folgenden angesprochenen spektralen Kovarianzmatrix entspricht).

2.2.3 Herleitung

Der MUSIC-Ansatz

Der MUSIC-Algorithmus beruht als *subspace*-Algorithmus auf der Zerlegung des Vektorraumes einer quellenpositionsabhängigen Größe $\vec{a}(\vec{\Theta})$ in einen Signalraum und einem dazu orthogonalen Rauschraum. Bei $\vec{a}(\vec{\Theta})$ kann es sich z.B. um die auf

Seite 9 angesprochenen Peil- oder *steering*-Vektoren handeln. Ist dies gelungen, kann durch Überprüfung der Orthogonalität zwischen durch $\vec{\Theta}$ parameterisierten Vektoren $\vec{a}(\vec{\Theta})$ und den Basisvektoren des Rauschraumes auf die Existenz einer Quelle an der durch $\vec{\Theta}$ spezifizierten Stelle geschlossen werden ($\vec{\Theta}$ beschreibt die Position der Quelle, z.B. als kartesisches Koordinatenpaar oder als Kombination von Entfernung und Winkel zum Mikrofonarray). Die Abbildung $\vec{\Theta} \rightarrow \vec{a}(\vec{\Theta})$ ergibt sich aus dem Signalausbreitungsmodell, welches Verzögerung (und ggf. Dämpfung) der an den einzelnen Mikrofonen ankommenden Signale beschreibt, wenn $\vec{\Theta}$ als Quellenposition angenommen wird. Ein Beispiel für eine solche Parametrisierung finden sich in Abschnitt 2.3.2, wo die für den Breitband-MUSIC Algorithmus relevante alleinige Parametrisierung durch den Einfallswinkel unter der Annahme von Fernfeldbedingungen (Wellenfronten nahezu planar / Einfallrichtung für alle Mikrofone gleich; ist bei entsprechendem Abstand zur Quelle gegeben) erläutert wird.

Theoretische Herleitung

Die Subraumzerlegung ergibt sich aus einer im folgenden genauer erläuterten Eigenwertzerlegung der spektralen Kovarianzmatrix

$$\mathbf{Cov}_{XX} = E\{(\vec{X}(j\omega) - \vec{\mu}_X) \cdot (\vec{X}(j\omega) - \vec{\mu}_X)^H\} \quad (2.10)$$

der Mikrofon-signale. Unter der Annahme

$$\vec{\mu}_X = E\{\vec{X}(j\omega)\} \stackrel{!}{=} 0 \quad (2.11)$$

läßt sich diese auch durch die spektrale Kreuzkorrelationsmatrix der Mikrofon-signale

$$\mathbf{R}_{XX} = E\{\vec{X}(j\omega) \cdot \vec{X}(j\omega)^H\} \quad (2.12)$$

approximieren. Setzt man Gleichung 2.6 in Gleichung 2.12 ein, so erhält man

$$\begin{aligned} \mathbf{R}_{XX} &= E\left\{(\mathbf{A} \cdot \vec{S}(j\omega) + \vec{N}(j\omega)) \cdot (\mathbf{A} \cdot \vec{S}(j\omega) + \vec{N}(j\omega))^H\right\} \\ &= E\{\mathbf{A} \cdot \vec{S}(j\omega) \cdot \vec{S}^H(j\omega) \cdot \mathbf{A}^H(j\omega)\} + \underbrace{E\{\mathbf{A} \cdot \vec{S}(j\omega) \cdot \vec{N}^H(j\omega)\}}_{\equiv 0} \\ &\quad + \underbrace{E\{\vec{N}(j\omega) \cdot \mathbf{A}^H \cdot \vec{S}^H(j\omega)\}}_{\equiv 0} + E\{\vec{N}(j\omega) \cdot \vec{N}^H(j\omega)\} \end{aligned} \quad (2.13)$$

da $n(t)$ weißes Rauschen ist und somit

$$\begin{aligned} E\{s(t) \cdot n(t)\} = 0 &\Rightarrow E\{S(j\omega) \cdot N(j\omega)\} = 0 \text{ gilt. Damit ist} \\ \mathbf{R}_{XX} &= \mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H + \mathbf{R}_{NN} \end{aligned} \quad (2.14)$$

mit $\mathbf{R}_{NN} = \sigma_N^2 \cdot \mathbf{I}$ folgt schließlich

$$\Rightarrow \mathbf{R}_{XX} = \mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H + \sigma_N^2 \cdot \mathbf{I} \quad (2.15)$$

Da $\mathbf{R}_{XX}(j\omega)$ als Korrelationsmatrix zweifellos eine hermitesche Matrix ist, läßt sie sich nach dem Hauptsatz über Hauptachsentransformationen (s. Kapitel A.1) durch eine Eigenwertzerlegung in die Form

$$\mathbf{R}_{XX} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H \quad (2.16)$$

bringen, wobei \mathbf{U} die orthogonale Matrix der Eigenvektoren $[\vec{v}_1, \vec{v}_2, \dots, \vec{v}_Q]^T$ von \mathbf{R}_{XX} und $\mathbf{\Lambda}$ die Diagonalmatrix der Eigenwerte von \mathbf{R}_{XX} ist. Für letztere ergibt sich nach dem Einsetzen in den Lösungsansatz für Eigenwertprobleme

$$\det(\mathbf{R}_{XX} - \lambda \mathbf{I}) \stackrel{!}{=} 0 \quad (2.17)$$

und mit Einsetzung von Gleichung 2.15

$$\det(\mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H + \sigma_N^2 \cdot \mathbf{I} - \lambda \mathbf{I}) \stackrel{!}{=} 0 \quad (2.18)$$

$$\det(\mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H - (\lambda - \sigma_N^2) \cdot \mathbf{I}) \stackrel{!}{=} 0 \quad (2.19)$$

durch eine Substitution $\lambda' = \lambda - \sigma_N^2 \Leftrightarrow \lambda = \sigma_N^2 + \lambda'$ erhält man mit

$$\det(\mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H - \lambda' \mathbf{I}) \stackrel{!}{=} 0 \quad (2.20)$$

wieder ein Eigenwertproblem. $\mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H$ hat so viele Eigenwerte λ' wie es unkorrelierte Quellensignale Q gibt. Nach der Resubstitution $\lambda = \sigma_N^2 + \lambda'$ erhält man für die Diagonalmatrix $\mathbf{\Lambda}$

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_N^2 + \lambda_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_N^2 + \lambda_2 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \sigma_N^2 + \lambda_Q & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \sigma_N^2 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & \sigma_N^2 \end{pmatrix} \quad (2.21)$$

Die Eigenvektoren \vec{v}_k müssen die Grundgleichung der Eigenwertprobleme

$$\mathbf{A} \cdot \vec{v}_k = \lambda_k \cdot \vec{v}_k \quad (2.22)$$

erfüllen. Da laut Gleichung 2.21 alle $\lambda_i = \sigma_N^2$ für $i > Q$ sind, ist

$$\mathbf{R}_{XX}(j\omega) \cdot \vec{v}_i = \lambda_i \cdot \vec{v}_i \hat{=} \sigma_N^2 \cdot \vec{v}_i \quad \text{für } i > Q \quad (2.23)$$

aber auch

$$\mathbf{R}_{XX}(j\omega) \cdot \vec{v}_i = (\mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H + \sigma_N^2 \cdot \mathbf{I}) \cdot \vec{v}_i = \sigma_N^2 \cdot \vec{v}_i \quad (2.24)$$

$$\Rightarrow (\mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H) \cdot \vec{v}_i + \sigma_N^2 \cdot \vec{v}_i = \sigma_N^2 \cdot \vec{v}_i \quad (2.25)$$

$$\Rightarrow (\mathbf{A} \cdot \mathbf{R}_{SS}(j\omega) \cdot \mathbf{A}^H) \cdot \vec{v}_i = 0 \quad (2.26)$$

$$\Rightarrow \mathbf{A}^H \cdot \vec{v}_i = 0 \quad (2.27)$$

Gleichung 2.27 besagt, daß jeder Eigenvektor \vec{v}_i mit $i > Q$ orthogonal zu den konjugiert-komplexen und transponierten Spaltenvektoren von \mathbf{A} ist; diese sind aber die Peilvektoren zu den jeweiligen Quellen. Im Umkehrschluß läßt sich folgern, daß jeder Peilvektor $\vec{a}(\vec{\Theta})$, welcher auf eine Quelle „zeigt“, orthogonal zu allen Eigenvektoren \vec{v}_i mit $i > Q$ sein muß. Definiert man einen Fehlervektor $\vec{\varepsilon}$,

$$\vec{\varepsilon} = \vec{a}^H(\vec{\Theta}) \cdot (\vec{v}_{Q+1}, \vec{v}_{Q+2}, \dots, \vec{v}_M) = \vec{a}^H(\vec{\Theta}) \cdot \mathbf{U}_N \stackrel{!}{=} \vec{0} \quad (2.28)$$

so läßt sich dies auch mathematisch exakt ausdrücken. Die Matrix \mathbf{U}_N ergibt sich aus der Zusammenfassung aller *nicht* Signalquellen zugeordneten Eigenvektoren. Sie entspricht der Basis der Rauscheigenvektoren; komplementär ergibt die Zusammenfassung der anderen Eigenvektoren die Basis des Signaleigenvektorraums U_S . Setzt man für $\vec{a}(\vec{\Theta})$ generische Peilvektoren ein, welche anhand eines Ausbreitungsmodells für Raumpositionen $\vec{\Theta}$ parameterisiert worden sind, kann Gleichung 2.28 zur Überprüfung auf das Vorhandensein einer Quelle an diesem Ort genutzt werden.

Die Auswertung läßt sich vereinfachen, wenn der Fehlervektor $\vec{\varepsilon}$ auf sein Betragsquadrat reduziert wird:

$$|\vec{\varepsilon}|^2 = \vec{\varepsilon} \cdot \vec{\varepsilon}^H \stackrel{!}{=} 0 \quad (2.29)$$

$$\Rightarrow \left(\vec{a}^H(\vec{\Theta}) \cdot \mathbf{U}_N \right) \cdot \left(\vec{a}^H(\vec{\Theta}) \cdot \mathbf{U}_N \right)^H \stackrel{!}{=} 0 \quad (2.30)$$

$$\Rightarrow \vec{a}^H(\vec{\Theta}) \cdot \underbrace{\mathbf{U}_N \mathbf{U}_N^H}_{\mathbf{\Pi}^\perp} \cdot \vec{a}(\vec{\Theta}) \stackrel{!}{=} 0 \quad (2.31)$$

Das für $\mathbf{U}_N \cdot \mathbf{U}_N^H$ eingeführte Symbol $\mathbf{\Pi}^\perp$ wird in der Literatur (z.B. [KV96]) als Selektor des Rauschunterraums bezeichnet. Die Berechnung dieses Selektors läßt sich durch die Aussage des Satzes über die Hauptachsentransformationen (s. Kapitel A.1), daß \mathbf{U}_N und damit die Eigenvektoren \vec{v}_i untereinander orthogonal

seien, weiter vereinfachen:

$$\mathbf{\Pi}^\perp = \begin{pmatrix} \vec{v}_{Q+1} \\ \vec{v}_{Q+2} \\ \dots \\ \vec{v}_M \end{pmatrix} \cdot (\vec{v}_{Q+1}, \vec{v}_{Q+2}, \dots, \vec{v}_M) \quad (2.32)$$

$$= \begin{pmatrix} |\vec{v}_{Q+1}|^2 & 0 & \dots & 0 \\ 0 & |\vec{v}_{Q+2}|^2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & |\vec{v}_M|^2 \end{pmatrix} \quad (2.33)$$

In der Realität wird durch Verstöße gegen das Signalmodell etc. in vielen Fällen $|\vec{\epsilon}|^2$ sehr klein werden, aber nicht exakt Null werden. Deshalb definiert man das sog. MUSIC-Spektrum in der Form

$$P_M(\vec{\Theta}) = \frac{\vec{a}^H(\vec{\Theta}) \cdot \vec{a}(\vec{\Theta})}{\vec{a}^H(\vec{\Theta}) \cdot \widehat{\mathbf{\Pi}}^\perp \cdot \vec{a}(\vec{\Theta})}; \quad (2.34)$$

der Zählerterm entspricht dem Betragsquadrat der Peilvektoren $\vec{a}(\vec{\Theta})$ und eliminiert die Abhängigkeit des MUSIC-Spektrums von dem Betrag der Peilvektoren. Die Maxima dieses „Spektrums“ geben wahrscheinliche Quellenpositionen an (s. Abbildung 2.2). In dieser Abbildung wurde, gemäß der Aufgabenstellung für die-

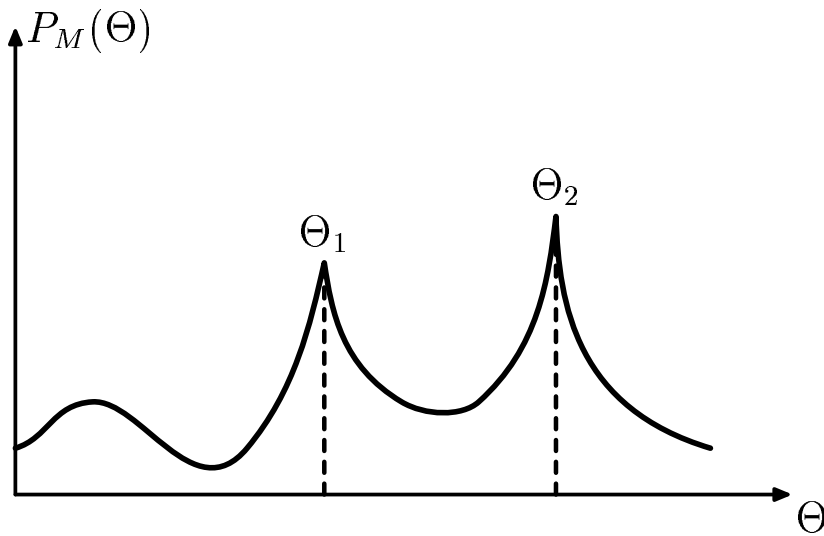


Abbildung 2.2: Das MUSIC-„Spektrum“

se Studienarbeit, $\vec{\Theta}$ auf den Azimutwinkel beschränkt. Angemerkt sei hier noch, daß laut [KV96] eine Variation des MUSIC-Algorithmus existiert, welche schon erkannte Peilvektoren dem Rauschunterraum zuordnet, damit sich die bisher noch nicht erkannten Quellenpeilvektoren stärker im MUSIC-Spektrum hervorheben; dieses Verfahren verbessert auch die Trennschärfe bzw. das Auflösungsvermögen eng beieinanderliegender Peilungen.

Zerlegung in Rausch- und Signaleigenvektorraum.

Die Separierung der Eigenvektoren in den Rauscheigenvektorraum \mathbf{U}_N und den Signaleigenvektorraum \mathbf{U}_S kann durch Analyse der zugehörigen Eigenwerte geschehen. Anhand der Matrix von Gleichung 2.21 erkennt man, daß für diese

$$\lambda_i = \begin{cases} \lambda_{Q_k} + \sigma_N^2 & \vec{v}_i \text{ ist Signaleigenvektor,} \\ \sigma_N^2 & \vec{v}_i \text{ ist Rauscheigenvektor} \end{cases} \quad (2.35)$$

gilt. Abbildung 3.3 zeigt, daß sich die Signaleigenwerte deutlich von den Rauscheigenwerten abheben. Für die Festlegung des Schwellwertes ϵ hat sich die Gleichung

$$\epsilon = \mu_{\vec{\lambda}} + \beta \cdot \sigma_{\vec{\lambda}} \quad (2.36)$$

bewährt; $\vec{\lambda}$ ist der Vektor der Eigenwerte, μ und σ dessen Mittelwert bzw. die Standardabweichung. Für β empfehlen sich Werte zwischen $0.2 \dots 2$. Die Eigenvektoren, deren zugehörige Eigenwerte unterhalb des Schwellwertes liegen, werden dem Rauscheigenvektorraum U_N zugeordnet.

Eine andere, u.a. in [EDDC] vorgeschlagene Methode, beruht auf der Modellierung der Wahrscheinlichkeitsdichteverteilung der Rauscheigenwerte durch z.B. eine Normalverteilung mit dem Erwartungswert σ_N^2 ; Signaleigenwerte wären in diesem statistischen Modell „Ausrutscher“.

2.2.4 Zusammenhang zwischen der Mikrofonanzahl und der Anzahl lokalisierbarer Quellen.

Das Extrakt der letzten beiden Abschnitte im Kontext ist, daß MUSIC hauptsächlich aus einer Eigenwertzerlegung der Korrelationsmatrix, Zuordnung der Eigenvektoren zum Rausch- oder Signaleigenvektorraum, Berechnung des MUSIC-Spektrums und Ermittlung der Quellenpositionen durch Bestimmung der Maxima besteht. Bei M Mikrofonen ist die Korrelationsmatrix $M \times M$ -dimensional. Unter der Annahme, daß die Korrelationsmatrix vollen Rang hat, existieren M Eigenvektoren. Damit eine Zerlegung stattfinden kann, muß mindestens einer dieser Eigenvektoren dem Rauscheigenvektorraum zugeordnet werden. Die verbleibenden $M - 1$ Eigenvektoren können dann Signaleigenvektoren sein – mit anderen Worten: man benötigt für die Detektion von Q Quellen $Q + 1$ Mikrofone. Das gilt aber nur dann, wenn die Korrelationsmatrix vollen Rang hat, also unter der Annahme von untereinander dekorrelierten Quellen- und Rauschsignalen.

2.2.5 Schätzung der Korrelationsmatrix

Das erste zu lösende Problem besteht in der Gewinnung einer Schätzung für die spektrale Kovarianzmatrix $\hat{\mathbf{R}}_{XX}(j\omega)$, wozu gängige Techniken (s. [Nol00])

verwendet werden können. Mit der Annahme, daß die Kovarianzmatrix in etwa die Kreuzkorrelationsmatrix ist, bieten sich z.B. die folgenden Möglichkeiten an:

- Approximierung des Erwartungswert-Operators durch die Mittelwertbildung über die letzten N Spektren:

$$\hat{\mathbf{R}}_{XX} = E \left[\vec{X}(j\omega) \cdot \vec{X}^H(j\omega) \right] \approx \frac{1}{N} \sum_{k=0}^{N-1} \vec{X}_{n-k}(j\omega) \cdot \vec{X}_{n-k}^H(j\omega) \quad (2.37)$$

- Eine weitere Vereinfachung kann durch Approximation des Mittelwertes mittels des *floating-average*-Verfahrens erreicht werden:

$$\hat{\mathbf{R}}_{XX} := (1 - \alpha) \cdot \hat{\mathbf{R}}_{XX} + \alpha \cdot \left(\vec{X}_{n-k}(j\omega) \cdot \vec{X}_{n-k}^H(j\omega) \right) \quad (2.38)$$

2.3 Konkretisierung des Signalmodells für Fernfeldbedingungen

Unter der Annahme ausreichend großer Abstände zwischen den Quellen und dem Mikrofonarray kann davon ausgegangen werden, daß die einfallenden Schallwellenfronten nahezu planar bzw. die Wellenvektoren \vec{k} an allen Mikrofonen nahezu parallel sind (s. Abbildung 2.6).

2.3.1 Abschätzung der nötigen Entfernung

Für die Abschätzung der nötigen Entfernung zum Erreichen von Fernfeldbedingungen kann folgende Überlegung angestellt werden.

Das entscheidende Kriterium ist die maximale Abweichung von Θ an den einzelnen Mikrofonen. Wie Abbildung 2.6 zeigt, ist bei der Annahme eines „Punktstrahlers“ diese Abweichung an den äußeren Mikrofonen des Arrays maximal. Die maximale Differenz ergibt sich bei $\Theta = 90^\circ$ (für $\Theta = 0^\circ$ bzw. $\Theta = 180^\circ$ ist sie $\Delta = 0^\circ$!). Durch Anwendung der mathematischen Sätze über Winkel an geschnittenen Parallelen findet man einen äquivalenten Winkel zu Δ an der Quellenposition; somit ist die maximale Winkelabweichung bei gegebener Arrayausdehnung d_A und dem Abstand zur Quelle d_S gleich

$$\Delta_{\max} = \frac{1}{2} \arctan \left(\frac{d_A}{2d_S} \right) \quad (2.39)$$

Abbildung 2.4 zeigt die Entwicklung dieser maximalen Winkelabweichung für ein steigendes Verhältnis d_S/d_A . Aus [McC01] konnte folgende Abschätzung für das Verhältnis von Entfernung zum Arraydurchmesser in zusätzlicher Abhängigkeit von der Wellenlänge entnommen werden:

$$d_S > \frac{2d_A^2}{\lambda} \quad (2.40)$$

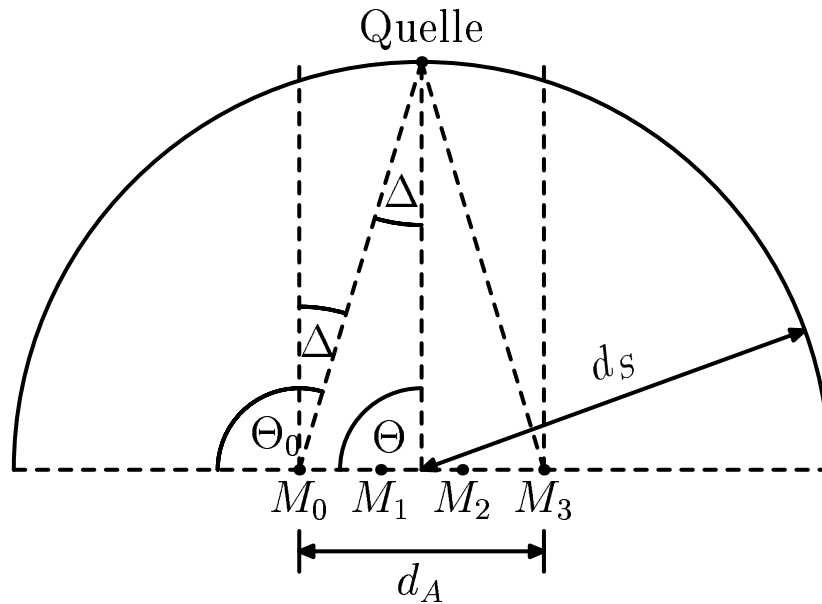


Abbildung 2.3: Abschätzung für Fernfeldbedingungenannahme

Für ein Array mit 4 Mikrofonen im Abstand von $0,2\text{ m}$ und einer angenommenen Schallgeschwindigkeit von $c = 346,25\text{ m} \cdot \text{s}^{-1}$ (Ergebnis aus Gleichung 2.42 für $T = 25^\circ$) ergeben sich –in Abhängigkeit von der Frequenz f – die in Abbildung 2.5 dargestellten Mindestentfernungen. Der nötige Abstand wächst proportional zur Frequenz und erreicht für das Beispiel-Array im für die menschliche Sprache typischen Frequenzbereich beachtliche Werte; eine Reduzierung dieses Abstands ist durch einen kleineren Mikrofonabstand d_A/N möglich, wie die zweite (gestrichelte) Kurve für einen Mikrofonabstand von $0,1\text{ m}$ zeigt.

2.3.2 Parametrisierung für den Schmalband-Ansatz

Unter der Annahme von Fernfeldbedingungen kann der Phasenunterschied zwischen zwei Mikrofonen wie folgt angegeben werden (s. Abbildung 2.6):

$$\begin{aligned}
 \Delta\varphi_{l,l+1} &= \omega \cdot \Delta t = \omega \frac{\Delta s}{c} \\
 &= \omega \frac{d \cos(\Theta)}{c} \\
 \text{mit } k &= |k| = \frac{\omega}{c} \\
 &= kd \cos(\Theta)
 \end{aligned} \tag{2.41}$$

Die Schallgeschwindigkeit c kann durch die aus [YHM01] entnommenen Approximationsformel in Abhängigkeit von der Temperatur T in Grad Celsius bestimmt

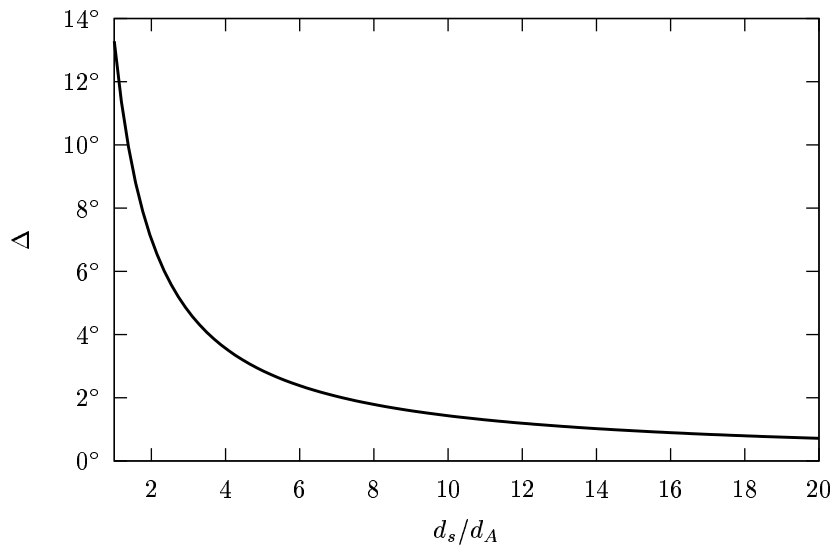


Abbildung 2.4: Maximale Winkelabweichung Δ in Abhängigkeit von d_s/d_A .

werden:

$$c \approx 331 + 0.610 \cdot T \quad [m \cdot s^{-1}] \quad (2.42)$$

Zu beachten ist, daß der in Gleichung 2.41 beschriebene Phasenunterschied durch die Verwendung von k eine Lösung für *eine* Frequenz angibt; bei Verwendung dieses Modells muß das Eingangssignal entsprechend schmalbandig sein. Welchen Einfluß die Frequenz auf die Winkelbestimmung hat, soll die folgende Beispielrechnung erläutern:

Stellt man Gleichung 2.41 nach dem Winkel Θ um, erhält man

$$\hat{\Theta} = \arccos \left(\frac{\varphi \cdot c}{2\pi f_2 \cdot d} \right). \quad (2.43)$$

Setzt man Gleichung 2.41, parameterisiert durch die „richtige“ Frequenz f_1 , in Gleichung 2.43 ein, wobei f_2 die für die Umrechnung der Phasen in die Azimutwinkel verwendete Frequenz ist, so erhält man

$$\hat{\Theta} = \arccos \left(\frac{\frac{2\pi f_1}{c} d \cos(\Theta) \cdot c}{2\pi f_2 \cdot d} \right) = \arccos \left(\frac{f_1}{f_2} \cos(\Theta) \right) \quad (2.44)$$

Die Auswirkungen einer von der tatsächlichen Signalfrequenz abweichenden Frequenz bei der Umrechnung $\varphi \Rightarrow \hat{\Theta}$ ist in Abbildung 2.7 dargestellt; man erkennt, daß schon relativ geringe Frequenzabweichungen zu merklichen Winkelfehlern führen.

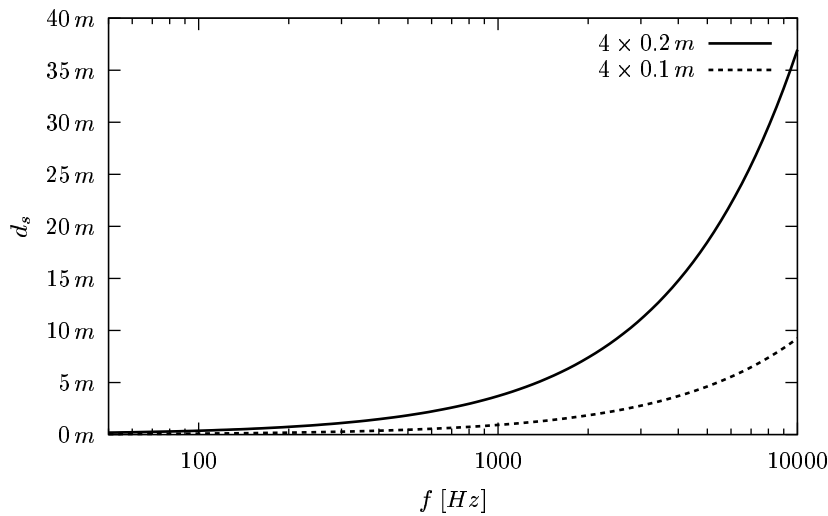


Abbildung 2.5: Abschätzung für die nötige Entfernung nach [McC01].

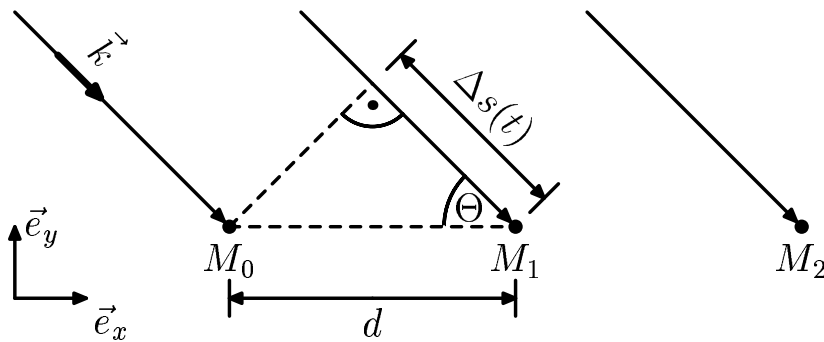


Abbildung 2.6: Signalmodell für Fernfeldbedingungen

2.4 Breitband-MUSIC

Der vorgestellte Schmalband-Ansatz lässt sich schwer mit dem Breitbandcharakter menschlicher Sprache in Einklang bringen. Die Forderung nach Schmalbandigkeit ergab sich dabei schon aus dem Grundansatz (s. Gleichung 2.4). Diese Forderung lässt sich durch einen anderen Grundansatz umgehen: Abbildung 2.6 legt den Gedanken nahe, daß zur Bestimmung des Einfallwinkels schon 2 Mikrofone ausreichen würden wenn, wie hier implizit vorausgesetzt wird, Fernfeldbedingungen herrschen.

2.4.1 Theoretische Herleitung

Der Breitband-Ansatz ist leichter nachvollziehbar, wenn man sich anfangs auf eine Quelle und ein Array mit 2 Mikrofonen beschränkt. Für die Mikrofonsignale

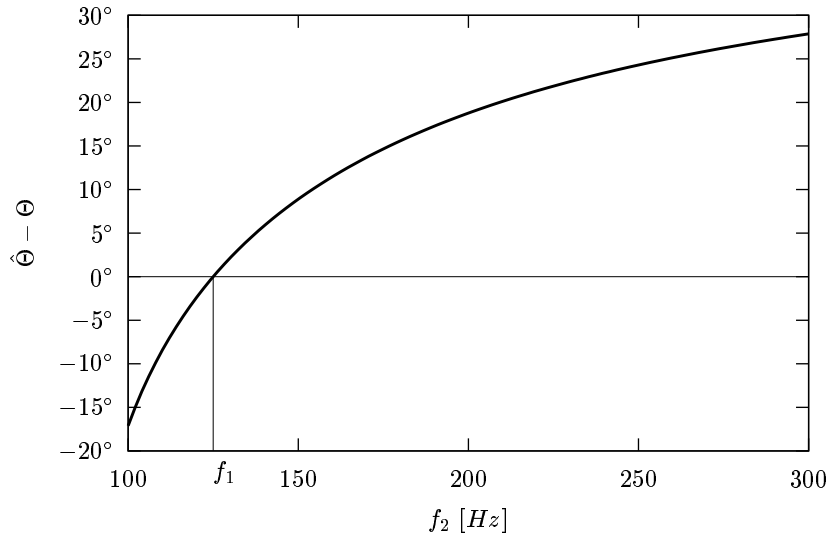


Abbildung 2.7: Differenz des errechneten DOA-Winkels $\hat{\Theta}$ vom tatsächlichen DOA-Winkel $\Theta = 45^\circ$; die Signalfrequenz beträgt $f_1 = 125 \text{ Hz}$.

gilt somit

$$x_0(k) = s_0(k) + n_0(k) \quad (2.45)$$

$$x_1(k) = s_0(k - \Delta k_0) + n_1(k) \quad (2.46)$$

Die relative Verschiebung Δk_0 für das Signal der ersten Quelle berechnet sich (s. Abbildung 2.6) bei einem Einfallswinkel Θ mit

$$\Delta s \hat{=} \cos(\Theta) \cdot d = c \cdot \Delta t \hat{=} c \cdot \frac{\Delta n_0}{f_T} \quad (2.47)$$

zu

$$\Rightarrow \Delta k_0 = f_T \cdot \cos(\Theta) \cdot \frac{d}{c} \quad (2.48)$$

$$\Leftrightarrow \Theta = \cos^{-1} \left(\frac{c \cdot \Delta k_0}{f_T \cdot d} \right) \quad (2.49)$$

Wenn die Quellen- und Rauschsignale als unkorreliert angenommen werden können, dann gilt für die Kreuzkorrelation

$$\begin{aligned} r_{x_0 x_1} &= E \{ (s_0(k) + n_0(k)) \cdot (s_0(k + \Delta k_0) + n_1(k + \Delta k_0)) \} \\ &= A^2 \cdot \delta(n - \Delta k_0) \end{aligned} \quad (2.50)$$

woraus mittels der Wiener-Khinchine-Beziehung (s. Kapitel A.1, Seite 49) das Kreuzleistungsdichtespektrum zu

$$S_{01} \left(jk \frac{2\pi}{SL} \right) \hat{=} \frac{1}{SL} E \left\{ X_0 \left(jk \frac{2\pi}{SL} \right) \cdot X_1^* \left(jk \frac{2\pi}{SL} \right) \right\} = \frac{A^2}{SL} \cdot e^{j\Delta \frac{2\pi}{SL} k_0} \quad (2.51)$$

folgt. Der Betrag des Kreuzleistungsdichtespektrum

$$|S_{01}\left(jk\frac{2\pi}{SL}\right)| = \frac{A^2}{SL} \stackrel{!}{=} \text{const.} \quad (2.52)$$

ist somit konstant. Bildet man nun die Kreuzleistungsdichtespektralmatrix⁴ \mathbf{P} , so erhält man

$$\mathbf{P} = \frac{A^2}{SL} \cdot \begin{pmatrix} e^{j0\frac{2\pi}{SL}\Delta k_0} & e^{j1\frac{2\pi}{SL}\Delta k_0} & \dots & e^{j(SL-1)\frac{2\pi}{SL}\Delta k_0} \\ e^{-j1\frac{2\pi}{SL}\Delta k_0} & e^{j0\frac{2\pi}{SL}\Delta k_0} & \dots & e^{j(SL-2)\frac{2\pi}{SL}\Delta k_0} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j(SL-1)\frac{2\pi}{SL}\Delta k_0} & e^{-j(SL-2)\frac{2\pi}{SL}\Delta k_0} & \dots & e^{j0\frac{2\pi}{SL}\Delta k_0} \end{pmatrix} \quad (2.53)$$

Diese Matrix läßt sich auch durch

$$\mathbf{P} = \vec{e} \cdot A^2 \cdot \vec{e}^H \quad (2.54)$$

mit

$$\vec{e} = \frac{1}{\sqrt{SL}} \left(e^{j0\frac{2\pi}{SL}\Delta k_0}, e^{j1\frac{2\pi}{SL}\Delta k_0}, \dots, e^{j(SL-1)\frac{2\pi}{SL}\Delta k_0} \right) \quad (2.55)$$

ausdrücken. Mit

$$\begin{aligned} \mathbf{P} \cdot \vec{e} &= \vec{e} \cdot A^2 \cdot \underbrace{(\vec{e}^H \cdot \vec{e})}_{\hat{=} |\vec{e}|^2 = 1} \\ \Rightarrow \mathbf{P} \cdot \vec{e} &= A^2 \cdot \vec{e} \end{aligned} \quad (2.56)$$

$$\hat{=} \text{Eigenwertproblem } \mathbf{A} \cdot \vec{v} = \lambda \cdot \vec{v}$$

folgt, daß \vec{e} ein Eigenvektor von \mathbf{P} mit dem zugehörigen Eigenwert A^2 ist.

Berücksichtigt man nun mehrere gleichzeitig aktive Quellen, so kann unter der Annahme der Unkorreliertheit der Quellensignale von einer Superposition der Einzelleistungsdichtespektren zu

$$S_{01}\left(jk\frac{2\pi}{SL}\right) = \sum_{q=0}^{Q-1} A_q^2 \cdot e^{-jk\frac{2\pi}{SL}\Delta n_q} \quad (2.57)$$

ausgegangen werden; somit läßt sich die CPSM-Matrix \mathbf{P} jetzt durch

$$\mathbf{P} = (\vec{e}_0, \vec{e}_1, \dots, \vec{e}_{Q-1}) \cdot \begin{pmatrix} A_0^2 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & A_1^2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & \dots & A_{Q-1}^2 & 0 & \vdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot (\vec{e}_0, \vec{e}_1, \dots, \vec{e}_{Q-1})^H$$

⁴ab hier wie in [EDDC] CPSM

(2.58)

ausdrücken. Diese Darstellung ähnelt der schon aus der Herleitung des Schmalband-MUSIC-Ansatzes bekannten Eigenwertzerlegung $\mathbf{P} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{U}^H$. Es läßt sich auch nachweisen, daß alle \vec{e}_k Eigenvektoren von \mathbf{P} sind, wie oben für den Fall einer Quelle gezeigt. Wie in der Herleitung des Schmalbandansatzes läßt sich auch hier zeigen, daß die den Quellen zugeordneten Eigenvektoren $\vec{v}_k \hat{=} \vec{e}_k, k \in [0 \dots Q - 1]$ orthogonal zu den anderen Eigenvektoren $\vec{v}_k, k \geq Q$ sind:

Da alle Eigenvektoren die Grundgleichung der Eigenwertprobleme $\mathbf{A} \cdot \vec{v} \stackrel{!}{=} \lambda \cdot \vec{v}$ erfüllen müssen, gilt

$$\mathbf{P} \cdot \vec{v}_i \stackrel{!}{=} \lambda_i \cdot \vec{v}_i \quad (2.59)$$

da aber $\lambda_i \equiv 0 \forall k \geq Q$ ist (s. Diagonalmatrix in Gleichung 2.58) folgt

$$\Rightarrow \mathbf{P} \cdot \vec{v}_i \stackrel{!}{=} 0 \quad (2.60)$$

mit der Substitution von \mathbf{P} durch den Ausdruck in Gleichung 2.58 ist aber

$$(\vec{e}_0, \vec{e}_1, \dots, \vec{e}_{Q-1}) \cdot \mathbf{D} \cdot ((\vec{e}_0, \vec{e}_1, \dots, \vec{e}_{Q-1})^H \cdot \vec{v}_k) \stackrel{!}{=} 0 \forall k \geq Q \quad (2.61)$$

$$\Rightarrow (\vec{e}_0, \vec{e}_1, \dots, \vec{e}_{Q-1})^H \cdot \vec{v}_k \stackrel{!}{=} 0 \forall k \geq Q \quad (2.62)$$

Die Gesamtheit der Vektoren \vec{v} ($k \in [0..Q-1]$) bilden somit den Signalunterraum \mathbf{U}_S (*signal subspace*), die verbleibenden Eigenvektoren $\vec{v}_k, \geq Q$ den Rauschunterraum \mathbf{U}_N . Somit kann wie beim Schmalbandansatz ein generischer Vektor $\vec{e}(\Theta)$ durch Feststellung der Orthogonalität zu allen *Nicht*-Signaleigenvektoren den Signaleigenvektorraum zugeordnet werden; idealerweise sollte dann

$$|\vec{e}^H(\Theta) \cdot \mathbf{U}_N|^2 = \vec{e}^H(\Theta) \cdot \underbrace{\mathbf{U}_N \cdot \mathbf{U}_N^H}_{\mathbf{\Pi}^\perp} \cdot \vec{e}(\Theta) \stackrel{!}{=} 0 \quad (2.63)$$

gelten; schließt man Rauschstörungen etc. in die Betrachtungen mit ein, wird der Betrag des Fehlervektors $\vec{e}^H(\Theta) \cdot \mathbf{U}_N$ nicht ganz verschwinden. Deshalb definiert man wieder das MUSIC-Pseudospektrum

$$P_M(\Theta) = \frac{1}{\vec{e}^H(\Theta) \cdot \hat{\mathbf{\Pi}}^\perp \cdot \vec{e}(\Theta)}, \quad (2.64)$$

dessen Maxima wiederum die Quellenpositionen angeben. Nach [EDDC] verbessern sich die Ergebnisse, wenn der Selektor gemäß

$$\hat{\mathbf{\Pi}}^\perp = \mathbf{U}_N \cdot \begin{pmatrix} \frac{1}{\lambda_Q^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_{Q+1}^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\lambda_N^2} \end{pmatrix} \cdot \mathbf{U}_N^H \quad (2.65)$$

berechnet wird. Anschaulich gesehen werden die Basisvektoren des Rauschunterraums mit dem Kehrwert ihrer Eigenwerte gewichtet - je größer die Eigenwerte sind, um so geringer ist die Wahrscheinlichkeit, daß der Basisvektor dem Rauschraum zuzuordnen ist.

2.4.2 Spatial smoothing

Das Breitband-MUSIC-Verfahren ist nicht auf zwei Mikrofone beschränkt; stehen mehr Mikrofone im ULA zur Verfügung, kann dies zu einer Verbesserung der Ergebnisse durch *spatial smoothing* genutzt werden. Dazu bilden immer zwei benachbarte Mikrofone ein Dupel; bei M Mikrofonen können somit $M - 1$ Dupel gebildet werden (s. Abbildung 2.8). Mittelt man

$$\overline{X_1(j\omega) \cdot X_2(j\omega)^*} = \frac{1}{M-1} \sum_{k=0}^{M-1} X_k(j\omega) \cdot X_{k+1}(j\omega)^* \quad (2.66)$$

über alle Dupel, so ist unter der Annahme von Fernfeldbedingungen zu erwarten, daß sich Rauschen und nur einzelne Mikrofone betreffende Störungen herausmitteln.

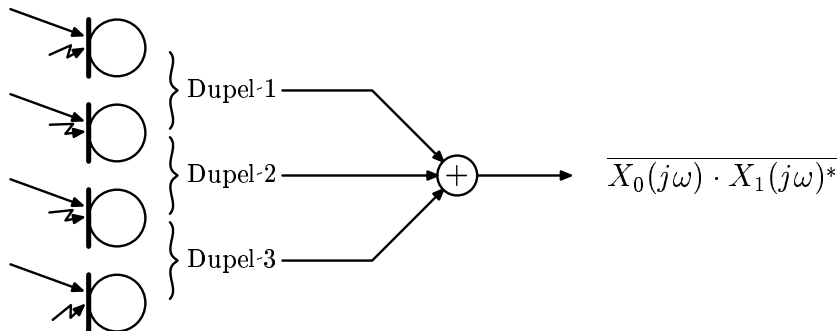


Abbildung 2.8: Prinzip des *spatial smoothings*

2.5 Root-MUSIC

Eine Variante des grundlegenden MUSIC-Algorithmus ist der hier im folgenden genauer beschriebene *root-MUSIC*-Algorithmus. Er reduziert die Maximasuche in einem Spektrum und die Berechnung des Spektrums auf ein polynomiales Nullstellenproblem. Quellen wie [KV96] attestieren ihm zudem eine höhere Trennschärfe im Vergleich zum grundlegenden MUSIC-Algorithmus.

2.5.1 Theoretische Herleitung

MUSIC beruhte auf einer Suche nach den Maxima des durch Gleichung 2.64 beschriebenen Pseudospektrums, welche gleichbedeutend mit den Minima des Nennerterms von Gleichung 2.64 sind. In der digitalen Signalverarbeitung stellt sich oft die ähnliche Aufgabe, die Minima des Betragsfrequenzganges eines Filters $H(jk\Delta\Omega)$ zu bestimmen. Hier kann die Bestimmung des Minimums dadurch erleichtert werden, daß man die z -Transformierte $H(z)$ der Filtertransferfunktion durch die Einführung der Substitution $z = e^{j\frac{2\pi}{SL}}$ aus $H(jk\Delta\Omega)$ bestimmt. Somit wird aus dem Spektrum in Abhängigkeit von $jk\Delta\Omega$ ein Polynom von z . Stellt man $H(z)$ in die Pol/Nullstellenform

$$H(z) = \frac{\prod_{q=0}^{Q-1} (z - z_{0,q})}{\prod_{r=0}^{R-1} (z - z_{x,r})} \quad (2.67)$$

um, so läßt sich der Betragsfrequenzgang unter der Annahme, daß alle Polstellen $z_{x,r} = 0$ sind, als

$$A(z) = |H(z)| = \prod_{q=0}^{Q-1} |z - z_{0,q}| \quad (2.68)$$

darstellen. $A(z)$ ist für eine gegebene Frequenz $z = e^{jk\frac{2\pi}{SL}}$ gleich dem Produkt der Abstände von z zu allen Nullstellen $z_{0,k}$. Das Minimum von $A(z)$ ist für das z zu erwarten, wo die Nullstellen nahe an der „Bahn“ von z , dem Einheitskreis in der z -Ebene, liegen.

Genau diese Überlegung wird auf den Nennerterm von Gleichung 2.64 angewandt; die Umwandlung von

$$P_{RM}(\Theta) = \vec{e}^H(\Theta) \cdot \hat{\mathbf{\Pi}}^\perp \cdot \vec{e}(\Theta) \quad (2.69)$$

in ein Polynom von z geschieht, wenn man in

$$\vec{e} = \left(e^{j0\frac{2\pi}{SL}\Delta k_q}, e^{j1\frac{2\pi}{SL}\Delta k_q}, \dots, e^{j(SL-1)\frac{2\pi}{SL}\Delta k_q} \right) \quad (2.70)$$

$e^{j1\frac{2\pi}{SL}\Delta k_q}$ durch z substituiert und somit

$$\vec{e} = (1, z, z^2, \dots, z^{SL-1}) \quad (2.71)$$

erhält. \vec{e}^H ist mit $e^{j1\frac{2\pi}{SL}\Delta k_q^*} = e^{-j1\frac{2\pi}{SL}\Delta k_q} \Rightarrow z^* = z^{-1}$

$$\vec{e}^H = (1, z^{-1}, z^{-2}, \dots, z^{-(SL-1)}) \quad (2.72)$$

Gleichung 2.69 kann somit als

$$P_{RM}(z) = (1, z^{-1}, z^{-2}, \dots, z^{-(SL-1)}) \cdot \hat{\mathbf{\Pi}}^\perp \cdot (1, z, z^2, \dots, z^{SL-1})^T \quad (2.73)$$

Polynom in z geschrieben werden. Die Nullstellen dieses Polynoms liegen paarweise gespiegelt zum Einheitskreis in der z -Ebene (s. Abbildung 3.4).

Mit diesem Wissen läßt sich der Ablauf von *Root-MUSIC* wie folgt skizzieren:

1. Bestimmung eines Schätzwertes für die Kreuzleistungsdichtespektralmatrix \mathbf{P} .
2. Durchführung einer *Singular Value Decomposition* von \mathbf{P} .

$$\mathbf{P} \xrightarrow{SVD} \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^T \quad (2.74)$$

Die Matrix der linksseitigen Eigenvektoren \mathbf{U} bildet die Basis des Vektorraumes von \vec{e} . \mathbf{D} ist die Diagonalmatrix der Eigenwerte λ , welche positiv und der Größe nach sortiert sind.

3. Abschätzung der Anzahl aktiver Quellen Q aus den Eigenwerten λ (s. Abschnitt 2.2.3).
4. Mit Q kann der Rauschvektorunterraum gemäß $\mathbf{U}_N = \mathbf{U}(1 \dots N, Q \dots N)$ extrahiert werden.
5. Berechnung des Polynoms P_{RM} aus der Selektormatrix $\mathbf{S} = \mathbf{U}_N \cdot \mathbf{U}_N^H$.
6. Berechnung der Nullstellen z_0 des Polynoms, z.B. nach dem Verfahren von Hessenberg (s. Gleichung ??).
7. Bestimmung der Q Nullstellen, welche im Einheitskreis liegen und den größten Betrag haben.
8. Umrechnung der Nullstellen in die zugehörigen DOA-Winkel: Nach der Re-substitution

$$z = e^{j \frac{2\pi}{SL} \Delta k_q} \Leftrightarrow \Delta k_q = \arg\{z\} \cdot \frac{SL}{2\pi} \quad (2.75)$$

und Anwendung von Gleichung 2.49 folgt

$$\Theta = \cos^{-1} \left(\frac{\arg\{z\} \cdot c \cdot SL}{2\pi f_T \cdot d} \right) \quad (2.76)$$

2.5.2 Berechnung des Polynoms

Anhand einer 3×3 -Selektormatrix $\mathbf{\Pi}^\perp$ soll ein Schema zur effizienten Berechnung des Root-MUSIC-Polynoms $P_{RM}(z)$, wie es auch in der Implementation verwendet wird, gezeigt werden.

Schreibt man Gleichung 2.73 für diesen Fall aus, so erhält man

$$P_{RM}(z) = (1, z^{-1}, z^{-2}) \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot (1, z, z^2)^T \quad (2.77)$$

multipliziert man den linken Vektor mit der Matrix, so erhält man

$$P_{RM}(z) = \begin{bmatrix} (a_{11} + a_{21}z^{-1} + a_{31}z^{-2}), \\ (a_{12} + a_{22}z^{-1} + a_{32}z^{-2}), \\ (a_{13} + a_{23}z^{-1} + a_{33}z^{-2}) \end{bmatrix} \cdot \begin{pmatrix} 1 \\ z \\ z^2 \end{pmatrix} \quad (2.78)$$

was nach dem Ausmultiplizieren und Sortieren der Koeffizienten

$$P_{RM}(z) = a_{13}z^2 + (a_{12} + a_{23})z + (a_{11} + a_{22} + a_{33}) + (a_{21} + a_{32})z^{-1} + a_{31}z^{-2} \quad (2.79)$$

ergibt. Sieht man sich diese Koeffizienten genauer an, so erkennt man, daß diese der Summe von Matrixelementen entsprechen, welche auf Parallelen zur Hauptdiagonale der Matrix stehen; diese Parallelen „wandern“ bildlich gesehen von der rechten oberen Ecke (a_{13}) in die linke untere Ecke (a_{31}). Da die Selektormatrix hermetisch ist, reicht es aus, die erste Hälfte des Polynoms zu berechnen - die andere Hälfte ergibt sich gespiegelt zur ersten Hälfte, nur konjugiert-komplex.

Kapitel 3

Die Implementierung

3.1 Beschreibung der Implementierung

Dieser Abschnitt beschreibt die Implementierung in Hinblick auf die Signalverarbeitung – die programmtechnische Struktur der Implementierung ist in der gesondert vorliegenden API-Dokumentation dargelegt.

Implementiert wurde die Breitband-Variante des Root-MUSIC-Algorithmus, wobei starke Anleihen an die Ausarbeitung „Robust TDOA Estimation via ROOT-MUSIC“ aus [EDDC] gemacht wurden; Simulationen unter *octave* zeigten, daß die dort vorgeschlagene Bildung der spektralen Kovarianzmatrix zu einer größeren Varianz der Abstände zwischen Nullstellen und Einheitskreis und damit auch zu einem höheren Entscheiderabstand führten. Statt

$$\mathbf{P} = \begin{pmatrix} S_{01}(j0\Delta\Omega) & S_{01}(j1\Delta\Omega) & \dots & S_{01}(jSL\Delta\Omega) \\ S_{01}(-j1\Delta\Omega) & S_{01}(j0\Delta\Omega) & \dots & S_{01}(j(SL-1)\Delta\Omega) \\ \vdots & \vdots & \ddots & \vdots \\ S_{01}(-jSL\Delta\Omega) & S_{01}(-j(SL-1)\Delta\Omega) & \dots & S_{01}(j0\Delta\Omega) \end{pmatrix} \quad (3.1)$$

wurde

$$\mathbf{P} = \begin{pmatrix} S_{01}(j0\Delta\Omega) & S_{01}(j1\Delta\Omega) & \dots & S_{01}(jP_{ord}\Delta\Omega) \\ S_{01}(j1\Delta\Omega) & S_{01}(j2\Delta\Omega) & \dots & S_{01}(j(P_{ord}-1)\Delta\Omega) \\ \vdots & \vdots & \ddots & \vdots \\ S_{01}(jP_{ord}\Delta\Omega) & S_{01}(j(P_{ord}-1)\Delta\Omega) & \dots & S_{01}(jSL\Delta\Omega) \end{pmatrix} \quad (3.2)$$

verwendet. Weitere Analysen zeigten, daß

$$P_{ord} = \frac{bins - 1}{2} \text{ mit } bins \hat{=} \text{Anzahl der verwendeten FFT}^1\text{-Koeffizienten} \quad (3.3)$$

zu den besten Resultaten führten. Da der Rechenaufwand für einen kompletten Analysezyklus recht groß ist, wurde die Analyse in zwei Teile zerlegt.

¹Fast Fourier Transformation

Das „Updaten“ des spektralen Kreuzleistungsdichtespektrum. Dieser Teil der Analyse kann durch die Methode `update()` aufgerufen werden. Als Argument erhält diese Methode eine $N \times SL$ -Matrix, welche für jedes der N Mikrofone einen Block von SL Abtastwerten beinhaltet. Die Blöcke werden mit einem Hammingfenster gefenstert (s. Abbildung 3.1) und per FFT in den Frequenzbereich transformiert. Ab jetzt werden nur noch die für den gewählten Frequenzbereich benötigten FFT-Bins zwischen n_{min} und n_{max} verwendet. Aus den transformierten Spektren wird gemäß Gleichung 2.66 ein durch *spatial smoothing* gemitteltes Spektrum $X_1(j\omega) \cdot X_2(j\omega)^*$ berechnet, welches durch Anwendung des *floating averages* (s. Gleichung 2.38) in den Schätzwert für das spektrale Kreuzleistungsdichtespektrum einfließt. Durch Anwendung der Transformation

$$X'(j\omega) = \frac{X(j\omega)}{|X(j\omega)|^2 + c} \text{ mit } c \ll 1 \quad (3.4)$$

wird das Kreuzleistungsdichtespektrum der Annahme aus Gleichung 2.52 angepaßt. Der Summand c im Nenner soll Divisionen durch Null verhindern.

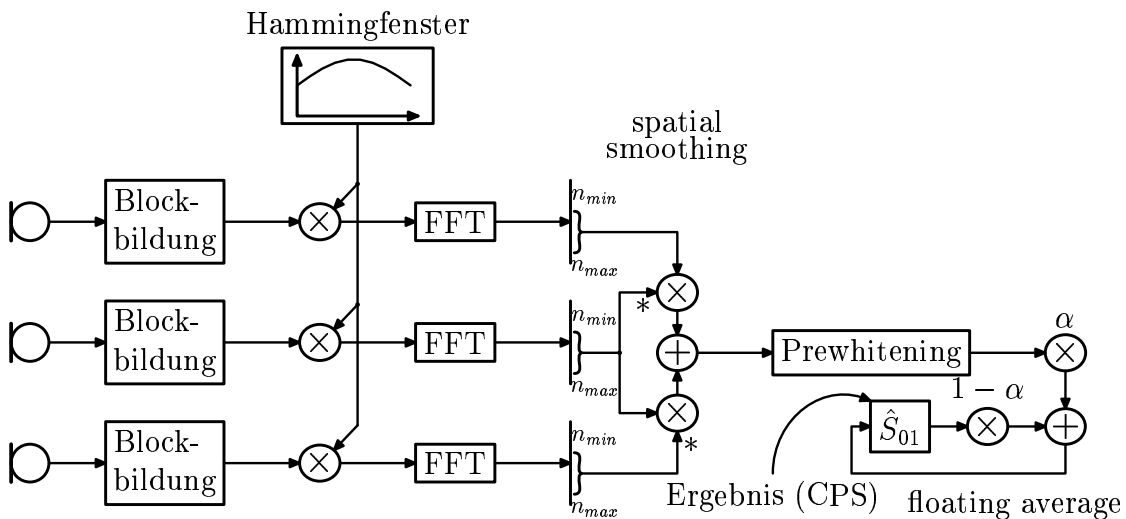


Abbildung 3.1: Datenflußgraph der Methode `update()` für ein Array mit 3 Mikrofonen

Ableitung der DOA-Winkel aus dem spektralen Kreuzleistungsdichtespektrum. Durch Aufruf der parameterlosen Methode `estimate()` wird die Analyse des ermittelten Kreuzleistungsdichtespektrums gestartet. Die Methode liefert eine Liste der DOA-Winkel aller detektierten Quellen zurück.

In einem ersten Schritt wird aus dem spektralen Kreuzleistungsdichtespektrum die CPSM-Matrix nach Gleichung 3.2 gebildet (s. Abbildung 3.2). Diese wird einer *Singular Value Decomposition* in die Matrix \mathbf{U} der Eigenvektoren und dem Vektor $\vec{\lambda}$ der Eigenwerte unterzogen.

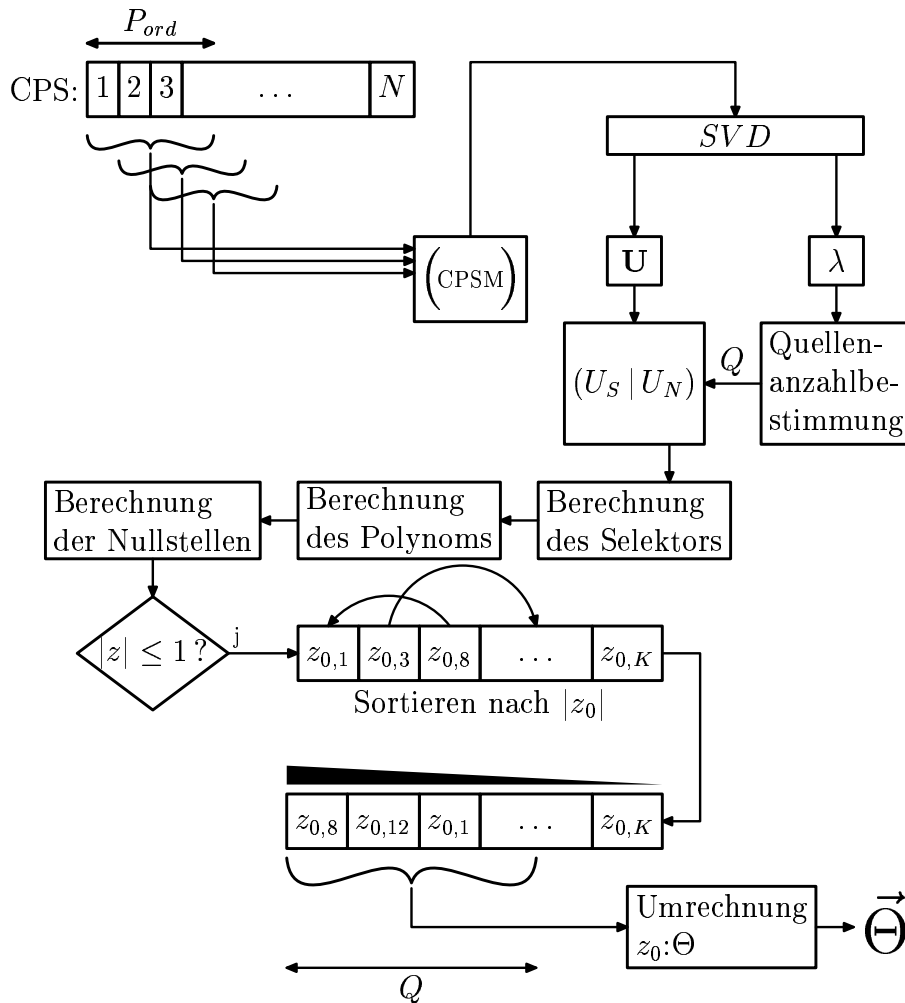


Abbildung 3.2: Datenflußgraph der Methode `estimate()`.

In Abbildung 3.3 ist $\vec{\lambda}$ für den Fall einer simulierten Rauschquelle (gleichverteiltes Rauschen der Varianz $\sigma_Q^2 = 1$) mit und ohne additives Rauschen auf den beiden Mikrofonkanälen dargestellt. Aus $\vec{\lambda}$ kann die Anzahl der aktiven Quellen bestimmt werden; in der Implementierung wurde dazu eine statistische Modellierung genutzt. Die Eigenwerte werden dazu als normalverteilt angenommen, wobei die Rauscheigenwerte „den Regelfall“ und die Signaleigenwerte „statistische Ausreißer“ sind (s. Abbildung 3.3). Da die SVD die Eigenwerte der Größe nach sortiert berechnet, ist die Anzahl der Quellen gleich dem Index Q (Achtung: Indexzählung startet bei 0!) des ersten Eigenvektors, dessen Wahrscheinlichkeit

$$P(\lambda_k) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\lambda_k - \mu)^2}{2\sigma^2}} > P_0 \quad (3.5)$$

(mit $\sigma = \sigma_{\vec{\lambda}}$ und $\mu = \mu_{\vec{\lambda}}$) größer als eine gegebene Mindestwahrscheinlichkeit P_0 ist.

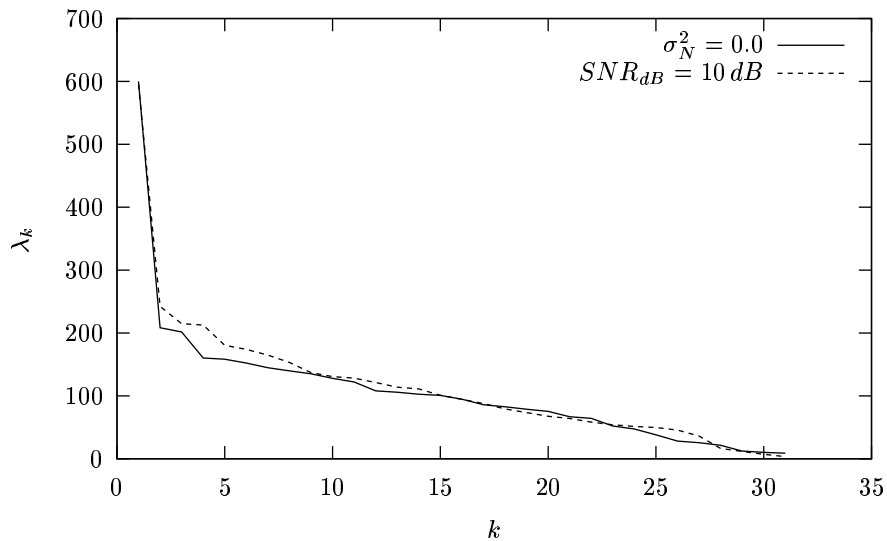


Abbildung 3.3: Verlauf der Eigenwerte bei einer aktiven Rauschquelle ($SL = 128$, $f_{Range} = [0 \dots \frac{f_T}{2}]$)

Nun kann aus der Matrix der Eigenvektoren \mathbf{U} die Matrix $\mathbf{U}_N = \mathbf{U}(Q : P_{ord})$ gewonnen werden. Diese wiederum kann in die Selektormatrix $\mathbf{S} = \mathbf{U} \cdot \mathbf{U}^H$ umgerechnet werden, aus welcher dann nach dem im Abschnitt 2.5.2 erläuterten Verfahren das *root-music*-Polynom $p(z)$ berechnet werden kann. Aus den Koeffizienten

$$p(z) = a_0 z^N + a_1 z^{N-1} + \dots + a_{N-1} z + a_N, \quad (3.6)$$

können die gesuchten Nullstellen z_0 nach dem Verfahren von Hessenberg als Eigenwerte der Matrix

$$\mathbf{H} = \begin{pmatrix} -\frac{a_1}{a_0} & -\frac{a_2}{a_0} & \dots & -\frac{a_N}{a_0} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad (3.7)$$

berechnet werden.

Gemäß der Herleitung von *Root-MUSIC* müssen nun die Q Nullstellen bestimmt werden, die dem Einheitskreis in der z -Ebene am nächsten sind. Eine charakteristische Eigenschaft der Nullstellen des *Root-MUSIC*-Polynoms ist, wie Abbildung 3.4 zeigt, daß die Nullstellen paarweise gespiegelt zum Einheitskreis liegen. Somit kann das Auswahlkriterium für die gesuchten Nullstellen auch in „Finde die betragsgrößten Nullstellen innerhalb des Einheitskreises“ umformuliert werden. Hierfür werden die Indizes und Betragsquadrate der Nullstellen mit

$|z_0| \leq 1$ werden in die Liste `maglist` aufgenommen, welche dann nach den Betragsquadraten sortiert wird. Anhand der „mitsortierten“ Indizes können nun die Q betragsgrößten Nullstellen mit $|z_0| \leq 1$ herausgesucht und nach Gleichung 2.76 in die gesuchten Einfallswinkel umgerechnet werden.

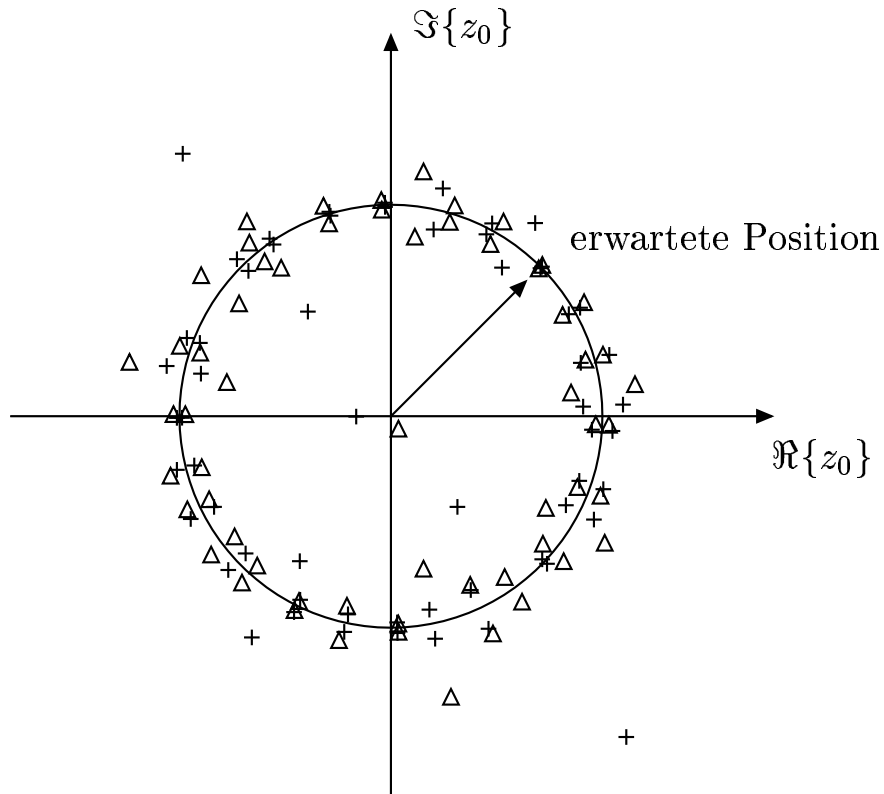


Abbildung 3.4: Position der Nullstellen bei einer aktiven Rauschquelle ($SL = 128$, $f_{Range} = [0 \dots \frac{f_T}{2}]$), $\Theta = 125.25^\circ$, gemessen: 125.24° ($SNR_{dB} = \infty$, Symbol Δ) bzw. 124.98° ($SNR_{dB} = 10 \text{ dB}$, Symbol $+$)

Durch die Auftrennung in die beiden Methodenaufrufe kann die relativ einfache Anpassung der spektralen Kovarianz häufiger erfolgen als die Auswertung, welche allein schon durch die Verwendung von SVD und Eigenwertzerlegung für recht große Matrizen sehr aufwendig ist.

3.2 Wahl der Parameter

In diesem Abschnitt sollen die Erfahrungen bezüglich der Wahl der Parameter, die im letzten Abschnitt eingeführt worden sind, dargelegt werden.

Floating-Average-Faktor α In der Implementierung wird der Erwartungswert des Kreuzleistungsdichtespektrums durch das *floating average* approximiert

(s. Gleichung 2.38). Der *floating average*-Faktor α gibt dabei den relativen Anteil von $X(j\omega)X^*(j\omega)$ am neuen Schätzwert für das Kreuzleistungsdichtespektrum an. Die besten Erfolge wurden für sehr kleine Werte von α erzielt (Der voreingestellte Wert ist $\alpha = 0.05$). So werden spektrale Fluktuationen unterdrückt und somit stabilere Peilungen erzielt. Nachteilig ist dann aber eine langsamere Anpassung an neue Quellen bzw. Positionsänderungen.

Wahrscheinlichkeitsschwelle *prob* Für die obere Wahrscheinlichkeitsschranke für Signaleigenwerte wurde empirisch eine Größenordnung von 10^{-7} gefunden.

Segmentlänge *SL* Die Segmentlänge ist proportional zu der Größe der intern verwendeten Matrizen; einerseits ist bei einer größeren Segmentlänge mit besseren Ergebnissen zu rechnen, andererseits wächst der Rechenaufwand quadratisch mit SL . Die Segmente sollten mindestens 20 ms breit sein, um die Merkmale der menschlichen Sprache in den Spektren zu erfassen.

Untere Frequenzschranke f_{min} Bei der Bearbeitung von realen Signalen ergab sich meist eine starke Peilung für $\Theta = 90^\circ$. Diese ist auf kohärente Anteile des Umgebungsrauschens zurückzuführen, welche wie ein direkt von vorn einfallendes Signal wirkten. In [sig98] wurde die Kohärenz von diffusen Schallfeldern genauer untersucht. Die Funktion der Kohärenz in Abhängigkeit von der Frequenz entspricht einer $si()$ -Funktion, welche mit steigendem Abstand zwischen den Mikrofonen immer schmaler wird. Für ein Mikrofonabstand $d = 20\text{ cm}$ liegt die erste Nullstelle bei $\approx 800\text{ Hz}$, deshalb sollte f_{min} mindestens 500 Hz betragen.

3.3 Abschätzung des Rechenzeitaufwandes

Die Evaluierung der Implementierung unter Berücksichtigung der im letzten Abschnitt gegebenen Empfehlungen führte selbst auf moderner Hardware (Pentium 4 mit 2 GHz Takt) zu Iterationszeiten in der Größenordnung von $\approx 10\text{ s}$. In diesem Abschnitt soll daher die Komplexität der wichtigen (bzw. rechenaufwendigen) Teiloperationen analysiert werden, um Aussagen über die Ursache des hohen Bedarfs an Rechenleistung machen zu können. Einerseits soll so Verständnis für den Rechenzeitaufwand und eine Abschätzmöglichkeit in Abhängigkeit von den Parametern erarbeitet werden, andererseits kann diese Abschätzung als Grundlage von Ansätze zur Verbesserung des Algorithmus dienen.

Ausgangspunkt der Betrachtung ist eine *Root-MUSIC*-Analyse über Segmente der Größenordnung SL , welche von einem Mikrofonarray mit M Mikrofonen stammen.

Die FFT. Die erste komplexe Rechenoperation ist die FFT, welche vom Typ $O(N) = N \cdot \log N$ ist. Da diese für alle Mikrofondatensegmente durchgeführt werden muß, beträgt der Gesamtaufwand

$$O_{FFT}(SL) = M \cdot SL \cdot \log SL. \quad (3.8)$$

Die SVD. Die nächste komplexe Operation ist die *Singular Value Decomposition* der spektralen Kovarianzmatrix, eine Operation des Typs $O(N) = N^2$. Unter der Annahme, daß der gesamte Frequenzbereich $f \in [0 \dots f_T/2]$ genutzt wird, werden $\text{bins} = SL/2$ Bins der FFT-Spektren für die weitere Analyse verwendet. Die Dimension der zu zerlegenden (quadratischen) spektralen Kovarianzmatrix beträgt dann $\text{RxyDim} = \text{bins} - 1/2$, so daß der Aufwand für die SVD

$$O_{SVD}(SL) = \text{RxyDim}^2 \approx \frac{SL^2}{16} \quad (3.9)$$

ist.

Eigenwertzerlegung der Hessenberg-Matrix (EVD). Dies ist der rechenintensivste Teil der Berechnung und somit auch der erste Ansatzpunkt für Optimierungen. Zur Berechnung der Nullstellen muß eine Matrix der Größe $(P - 1) \times (P - 1)$ mit $P = 2 \cdot \text{RxyDim} - 1$ zerlegt werden; da die Eigenwertzerlegung ebenfalls eine Operation des Typs $O(N) = N^2$ ist, beträgt der Aufwand

$$O_{EVD} = (\text{bins} - 1)^2 = \left(\frac{SL}{2} - 2 \right)^2 \quad (3.10)$$

Das Sortieren der Nullstellen. Zuletzt werden die Nullstellen nach ihren Beträgen sortiert. Als Sortieralgorithmus wurde Quicksort ($O(N) = N \cdot \log N$) verwendet. Da die Nullstellen außerhalb des Einheitskreises schon vorher aussortiert worden sind und die Nullstellen paarweise gespiegelt am Einheitskreis auftreten, muß ein Vektor mit $P/2$ Nullstellen sortiert werden. Der Aufwand für diesen Schritt ist somit

$$O_{\text{sort}}(SL) = \frac{P}{2} \cdot \log \frac{P}{2} = \left(\frac{SL}{4} - 1 \right) \cdot \log \left(\frac{SL}{4} - 1 \right). \quad (3.11)$$

Die Darstellung der Entwicklung von $O_{\text{ges.}}(SL)$ in Abbildung 3.5 zeigt, daß $O_{\text{ges.}}(SL) \sim SL^x$ ist; durch eine Polynomapproximation konnte x zu ≈ 0.32 bestimmt werden. In Tabelle 3.1 sind die relativen Anteile der wichtigen Berechnungen am geschätzten Gesamtaufwand zusammengetragen. Diese Abschätzung macht deutlich, daß nahezu 60% der Komplexität von *root-MUSIC* für die Berechnung der Nullstellen (EVD) aufgewandt werden. Somit ist es naheliegend, Optimierungen an diesem Punkt anzusetzen. Allerdings gilt das hier verwendete

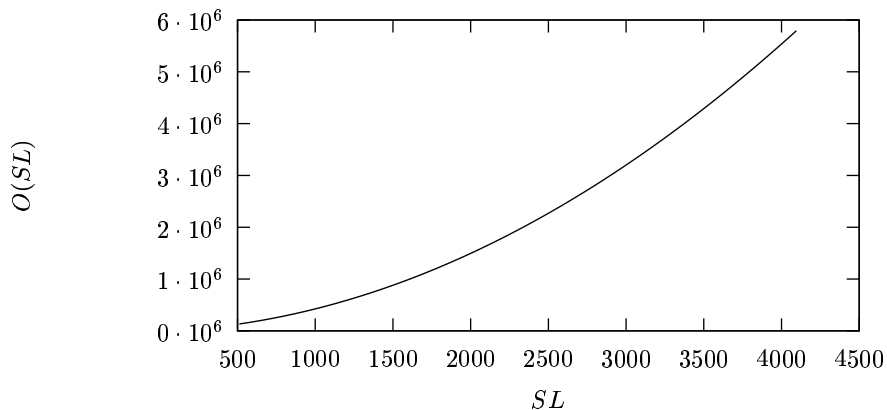


Abbildung 3.5: Abschätzung des Aufwand eines *Root-Music*-Analysezyklus in Abhängigkeit von der Segmentlänge SL für ein Mikrofonarray mit 16 Mikrofonen.

Verfahren von Hessenberg schon als optimale Lösung zur Berechnung der Nullstellen eines komplexen Polynoms, wenn alle Nullstellen berechnet werden sollen.

Die sich laut Tabelle 3.1 ergebende Operation mit der zweitgrößten Komplexität, die FFT, eröffnet größere Chancen zur Optimierung des Laufzeitverhaltens. Wurde die FFT für die Mikrofonsegmente schon im „umgebenden“ Programm berechnet, so kann die nochmalige Berechnung in der Implementierung für den *Root-MUSIC*-Algorithmus unterbleiben. Allerdings spiegeln die Werte in Tabelle 3.1 nicht exakt den realen Rechenzeitbedarf wieder; dieser hängt auch von der Effizienz der jeweiligen Umsetzung ab.

Berechnung	Anteil am Gesamtaufwand
FFT	25.70%
SVD	14.83%
EVD	58.86%
Quicksort	0.32%

Tabelle 3.1: relativer Anteil der rechenaufwendigen Teilberechnungen am Gesamtaufwand für ein Array mit 16 Mikrofonen und $SL = 1024$

Kapitel 4

Evaluierung

Abschließend soll die Leistung der Implementierung bewertet werden. Dazu wurden ein simuliertes Szenario mit drei Rauschquellen, ein reales Szenario mit 4 Sprechern und das Szenario eines Raumes mit Reflektionen an den Wänden untersucht.

4.1 Generiertes Rauschen von mehreren Quellen

Ausgangsbasis für die Untersuchung sind die Daten eines simulierten Arrays mit 2 Mikrofonen im Abstand $d = 0.2 m$, welches Signale von 3 Rauschquellen mit einer Varianz von je $\sigma_Q = 1$ aus $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ empfängt. Variiert wurden bei der Analyse die Varianz des additiven Rauschens σ_N^2 und die Segmentlänge SL . Es wurde der gesamte verfügbare Frequenzbereich $[0 \dots f_T/2]$ ($f_T = 48 kHz$) genutzt. Die Signale selbst wurden durch das in `examples/siggen` liegende Programm `siggen` erzeugt, die Analyse fand mit dem Programm `wbrootmusic` aus dem gleichen Verzeichnis statt. Die Quellensignale entsprechen den in der theoretischen Herleitung gemachten Annahmen über die Unkorreliertheit und der Konstanz des Betrages des Kreuzleistungsdichtespektrums. Auch die Rauschsignale entsprechen der für sie gemachten Annahme der Unkorreliertheit zueinander und zu den Quellensignalen. Somit kann die Implementierung des Verfahrens auf das „theoretische“ Funktionieren geprüft werden. In den Abbildungen 4.1 bis 4.4 sind die Peilungen über die Segmente dargestellt. Die Abbildungen zeigen die erzielten Ergebnisse für unterschiedliche Segmentlängen ($SL = 128$ für die Abbildungen 4.1 und 4.3 bzw. 256 für die Abbildungen 4.2 und 4.4). Für die ersten beiden Abbildungen wurden ungestörte Quellensignale simuliert, für die letzten beiden Abbildungen wurde zusätzlich eine additive Rauschstörung ($SNR = 10 dB$) simuliert. Erwartungsgemäß zeigt sich für die größere Segmentlänge eine geringere Varianz der Peilungen. Die additiven Rauschstörungen haben keine Rückwirkung auf das Ergebnis. Auch dies deckt sich mit den Erwartungen, da die

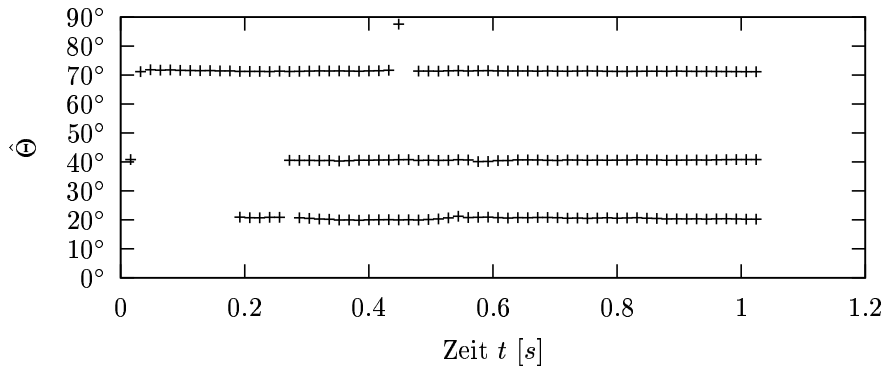


Abbildung 4.1: Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-4}$, $SL = 128$, $SNR_{dB} = \infty$)

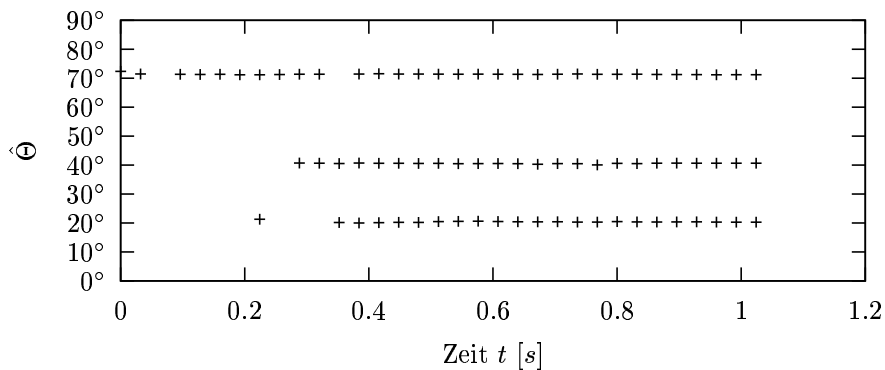


Abbildung 4.2: Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-6}$, $SL = 256$, $SNR_{dB} = \infty$)

Rauschkomponenten weitestgehend unkorreliert sind und eine geringere Leistung als die Quellensignale besitzen.

Auffallend in allen Abbildungen ist das zeitlich versetzte Einsetzen der Peilungen. Dies ist auf die Approximation des Erwartungswertoperators durch das *floating average* bei der Berechnung der spektralen Kovarianzmatrix zurückzuführen – der Rang dieser Matrix baut sich erst nach und nach auf. Bei der Approximierung des Erwartungswertes muß der nichtstationäre Charakter der Quellensignale und Quellenpositionen berücksichtigt werden. Über den Gewichtungsfaktor α des *floating-average*-Verfahrens muß ein Kompromiß zwischen der Forderung nach Stabilität und Vollständigkeit der Peilungen durch eine ausreichend „ausgeformte“ spektrale Kovarianzmatrix und der Forderung nach schneller Anpassung an neue Umgebungen gesucht werden.

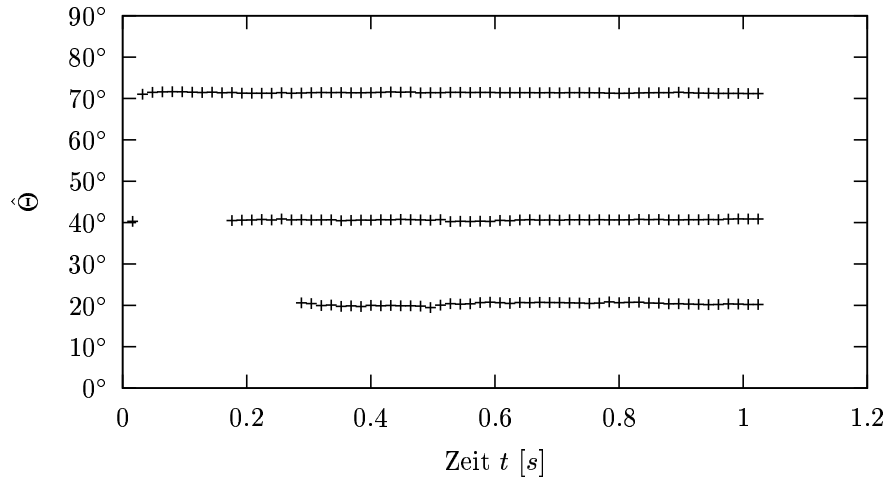


Abbildung 4.3: Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-4}$, $SL = 128$, $SNR_{dB} = 10$ dB)

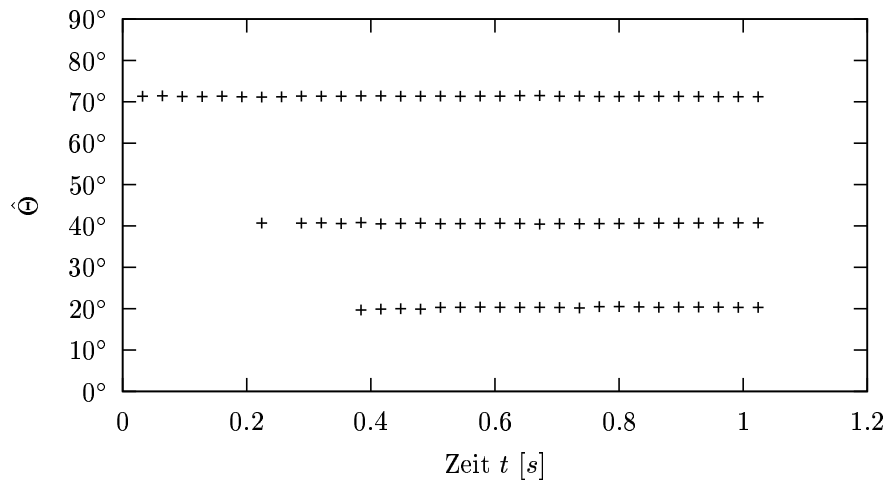


Abbildung 4.4: Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-6}$, $SL = 256$, $SNR_{dB} = 10$ dB)

4.2 Analyse von Signalen mit Raumechos

Nach der Prüfung mit genau den Annahmen entsprechenden Quellensignalen wurde nun das Verhalten des Algorithmus für den Fall untersucht, daß die Quellensignale nicht der Forderung nach De- und Unkorreliertheit entsprechen. Für diesen Test wurde praxisnah ein Raum mit Nachhall simuliert. Für die Generierung der Raumimpulsantworten wurde ein Matlab-Skript aus [Roo] verwendet. Für die Simulation wurde das in Abbildung 4.5 dargestellte Szenario verwendet. Es besteht aus einem Raum mit den Abmessungen $(4\text{ m} \times 5\text{ m} \times 3\text{ m})$ mit 2 Mikrofonen bei $(x, y, z) = (1.9\text{ m}, 0.2\text{ m}, 1.5\text{ m})$ und $(2.1\text{ m}, 0.2\text{ m}, 1.5\text{ m})$. Die Sprecher sind an den Positionen $(3.0\text{ m}, 4.0\text{ m}, 1.8\text{ m})$ ($\hat{=} \Theta_1 = 104.74^\circ$) und $(1.0\text{ m}, 3.0\text{ m}, 1.8\text{ m})$ ($\hat{=} \Theta_2 = 70.35^\circ$). Die mittels eines Simulationsprogramms bestimmten Raumimpulsantworten $h_{11}(k) \dots h_{22}(k)$ sind in den Abbildungen 4.8 bis 4.11 dargestellt. Die Mikrofonsignale wurden durch

$$x_0(n) = n_0(k) + \sum_{k=1}^2 q_k(n) * h_{1k} \quad (4.1)$$

$$x_1(n) = n_1(k) + \sum_{k=1}^2 q_k(n) * h_{2k} \quad (4.2)$$

berechnet, wobei n_1 und n_2 additives gaußverteiltetes Rauschen der Varianz $\sigma_n^2 = 0,01$ sind (dies entspricht bei einer durchschnittlichen Varianz der Sprachblöcke $\sigma_q = 0,04$ einem $SNR_{dB} \approx 6\text{ dB}$). Da die Quellensignale durch Reflektionen über mehrere Pfade zu den Mikrofonen gelangen, sind in deren Ausgangssignalen korrelierte Komponenten enthalten. Das verstößt gegen die in der Herleitung von MUSIC gemachten Annahmen.

Die Streuung der gemessenen DOA-Werte bedingt eine statistische Auswertung der Ergebnisse. Da sich die beiden Quellen (s. Abbildung 4.5) aus Sicht des Mikrofonarrays jeweils in der linken bzw. rechten Hälfte des Raumes befinden, wurde für Θ_1 und Θ_2 die Erwartungswerte über die entsprechenden Winkelbereiche verwendet. Mit dem in Abbildung 4.7 dargestellten Histogramms als Approximation für $p_\Theta(\Theta)$ können mit

$$\hat{\Theta}_1 = 2 \cdot E \{ \hat{p}_\Theta(\Theta) \mid \Theta \leq 90^\circ \} \approx 2 \sum_{\forall \Theta_n \leq 90^\circ} \hat{p}_\Theta(\Theta_n) \cdot \Theta_n \quad (4.3)$$

und

$$\hat{\Theta}_2 = 2 \cdot E \{ \hat{p}_\Theta(\Theta) \mid \Theta > 90^\circ \} \approx 2 \sum_{\forall \Theta_n > 90^\circ} \hat{p}_\Theta(\Theta_n) \cdot \Theta_n \quad (4.4)$$

Schätzwerte für die gemessenen DOA-Winkel ermittelt werden. Diese ergaben sich zu $\hat{\Theta}_1 = 48.2^\circ$ und $\hat{\Theta}_2 = 107.8^\circ$.

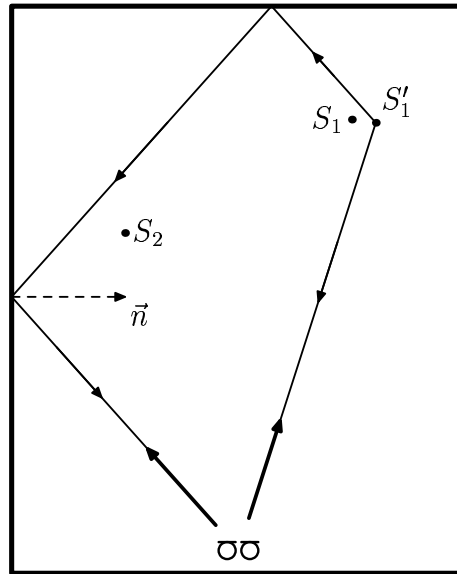


Abbildung 4.5: Szenario für die Analyse von Signalen mit Raumechos

Während $\hat{\Theta}_2$ den Erwartungen entspricht, scheint $\hat{\Theta}_1$ auf dem ersten Blick völlig falsch zu sein. Aus Abbildung 4.5 wird jedoch ersichtlich, daß $\hat{\Theta}_1$ mit großer Wahrscheinlichkeit einer Reflexion des Signals von S_1 (gepeilt an der Position S_1') entspricht.

Durch die Verwendung eines vorgeschalteten LPC-Filters konnte diese Reflexion zwar unterdrückt werden, aber die Peilung für S_2 blieb weiterhin aus.

Die Simulation wurde mit anderen Sprachsignalen und einer blockweisen adaptiven LPC-Vorfilterung (LPC-Filter mit 128 Koeffizienten) wiederholt. Die Ergebnisse sind in Abbildung 4.12 dargestellt, das aus ihnen abgeleitete Histogramm in Abbildung 4.13. Im Bereich $t \in [0 \dots 1.75s]$, wo beide Sprecher aktiv sind (s. Abbildungen 4.14 und 4.15), werden auch beide Quellen erkannt. Das Histogramm (Abbildung 4.15) zeigt aber auch, daß immer noch Peilungen auf Reflexionen auftreten. Somit wurde zwar das Problem der Unterdrückung von Quellen, welches auf die Reduktion des Rangs der spektralen Kovarianzmatrix durch die Korrelationen bedingt ist, nur teilweise gelöst. Deshalb ist eine Nachbehandlung der Positionsdaten, wie sie in der Zusammenfassung noch näher besprochen wird, anzuraten.

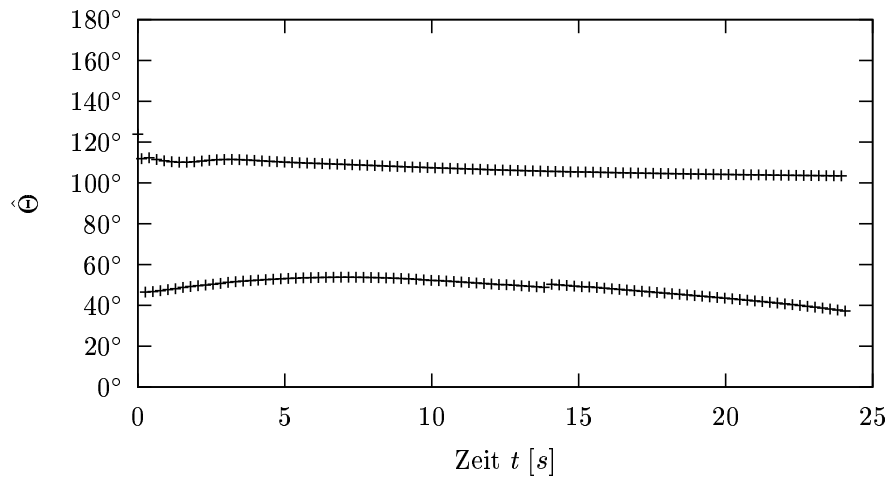


Abbildung 4.6: Ergebnisse für das „verhallte-Räume-Szenario“ ($f_T = 8 \text{ kHz}$, $f = [1 \text{ kHz} \dots 4 \text{ kHz}]$, $\alpha = 0.05$, $prob = 10^{-7}$, $SL = 1024$)

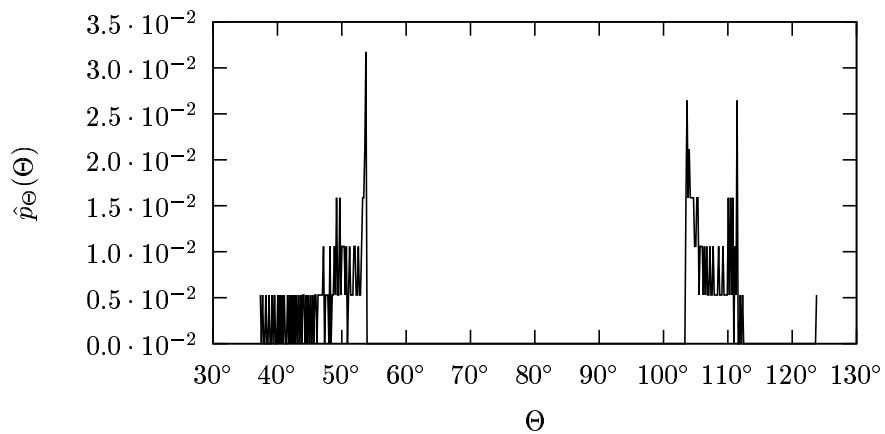
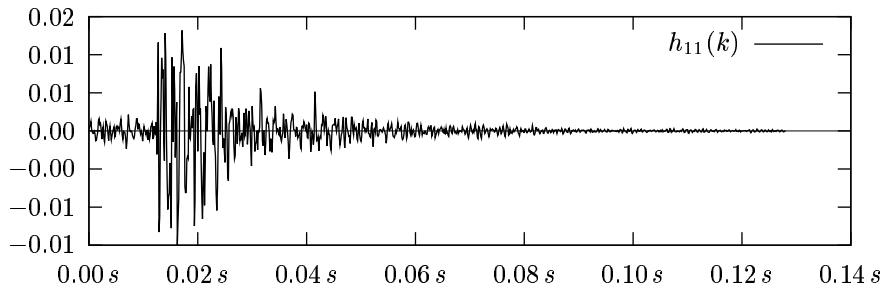
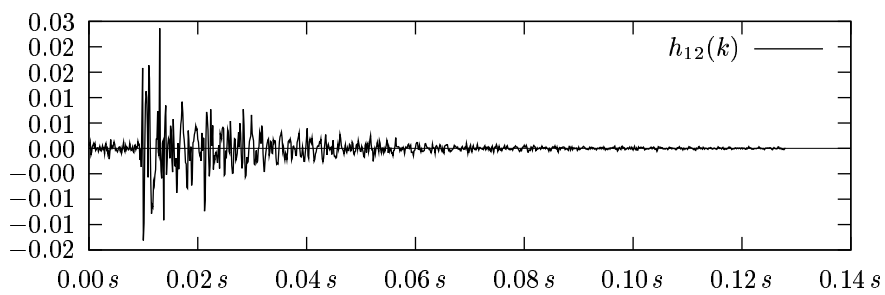
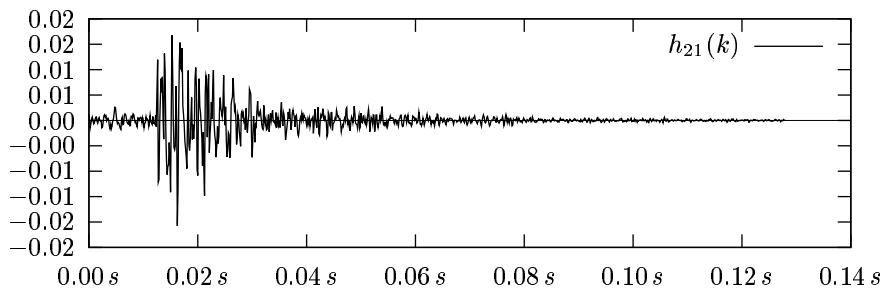
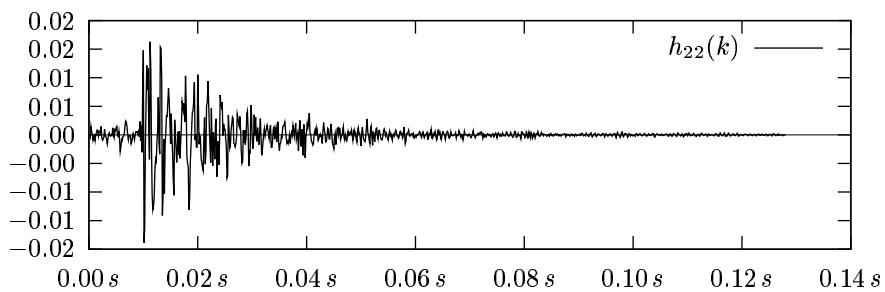


Abbildung 4.7: Histogramm der ermittelten DOA-Winkel für das „verhallte-Räume-Szenario“

Abbildung 4.8: Raumimpulsantwort $S_1 \rightarrow \text{Mic}_1$.Abbildung 4.9: Raumimpulsantwort $S_2 \rightarrow \text{Mic}_1$.Abbildung 4.10: Raumimpulsantwort $S_1 \rightarrow \text{Mic}_2$.Abbildung 4.11: Raumimpulsantwort $S_2 \rightarrow \text{Mic}_2$.

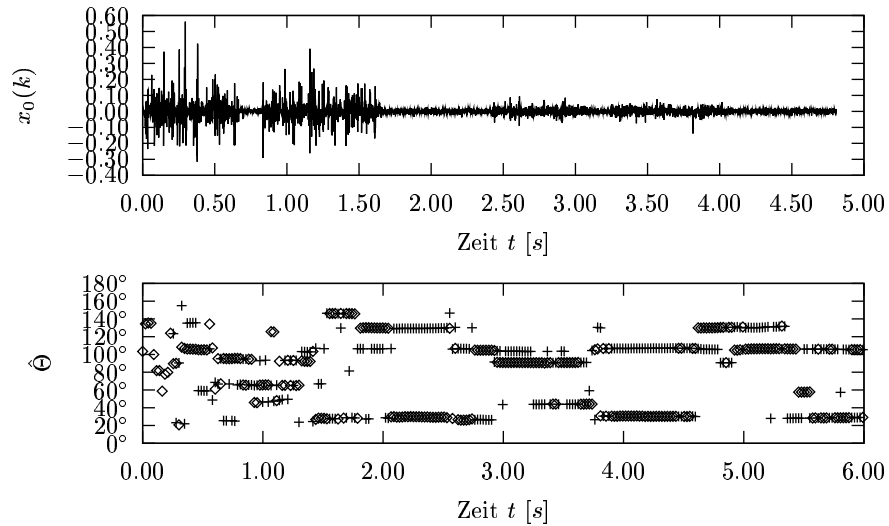


Abbildung 4.12: Ergebnisse für andere Sprecher und blockweise adaptive LPC-Vorfilterung durch ein Filter mit 128 Koeffizienten. ($f_T = 44,1 \text{ kHz}$, $f = [0 \dots 22.05 \text{ kHz}]$, $\alpha = 0.05$, $prob = 10^{-7}$, $SL = 1024$)

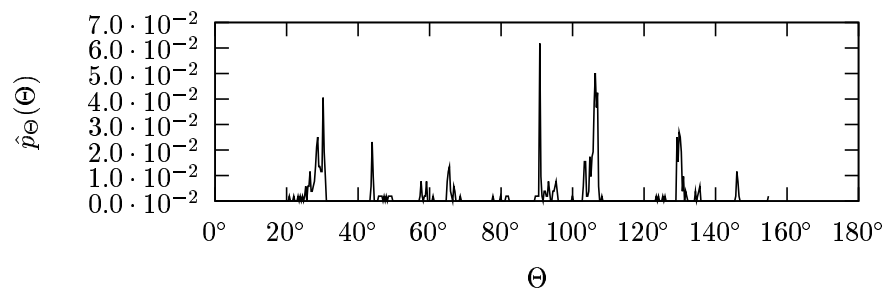
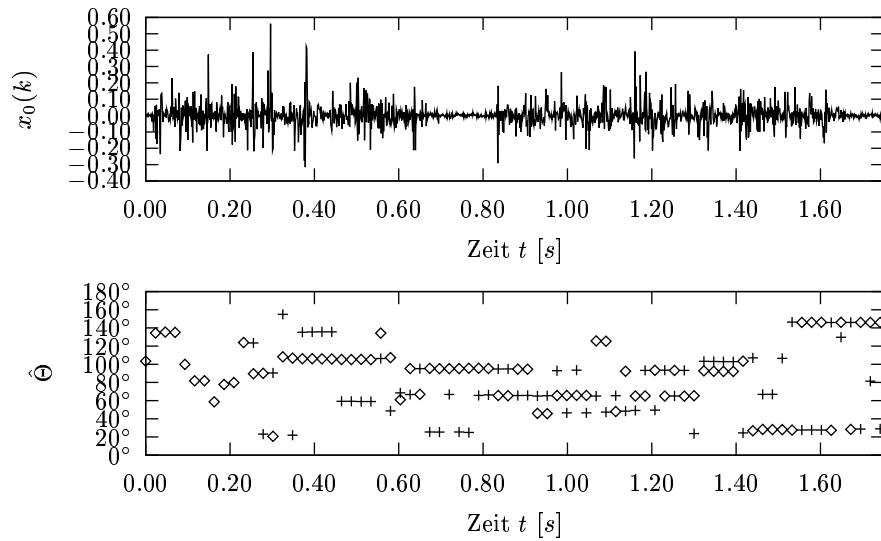
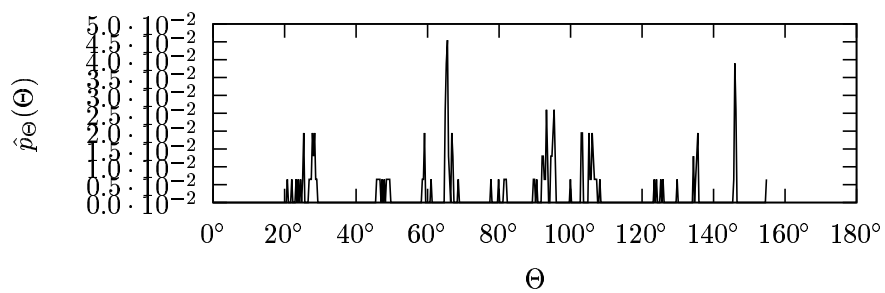


Abbildung 4.13: Histogramm der ermittelten DOA-Winkel für das „verhallte-Räume-Szenario“ mit anderen Sprechern und LPC-Vorfilterung

Abbildung 4.14: Ausschnitt $t \in [0 \dots 1.75 \text{ s}]$ aus Abbildung 4.12Abbildung 4.15: Histogramm für den Ausschnitt $t \in [0 \dots 1.75 \text{ s}]$ aus Abbildung 4.14

4.3 Analyse realer Signale

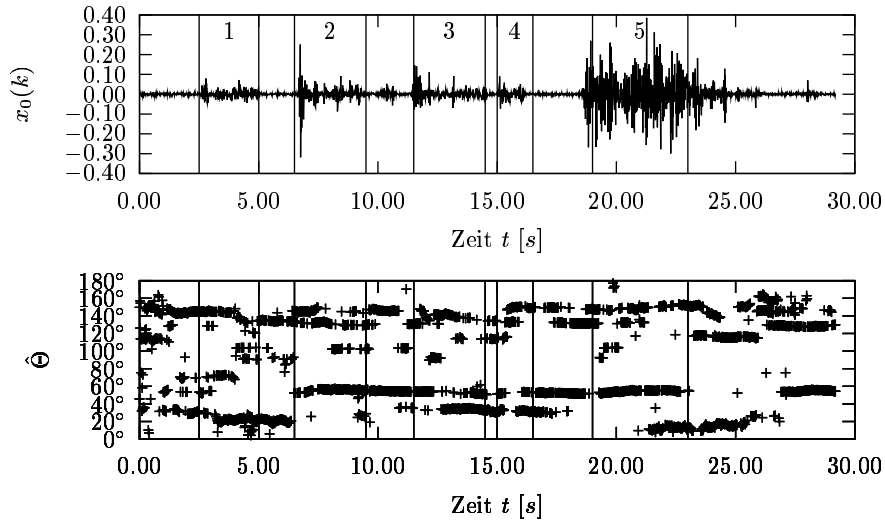


Abbildung 4.16: Ergebnisse für reale Mikrofondaten ($f_T = 48 \text{ kHz}$, $f = [1 \text{ kHz} \dots 20 \text{ kHz}]$, $\alpha = 0.05$, $\text{prob} = 10^{-7}$, $SL = 1024$), Bereich 1-4: Sprecher 1-4, Bereich 5: alle Sprecher aktiv

Abschließend soll das Verhalten der Implementierung für die Signale eines realen Mikrofonarrays getestet werden. Das Testszenario bestand aus einem Mikrofonarray mit 4 Mikrofonen im Abstand von je 20 cm, deren Signale mit $f_T = 48 \text{ kHz}$ abgetastet wurden. Um dieses Mikrofonarray wurden in einer halb-kreisförmigen Anordnungen im Abstand von $\approx 2 \text{ m}$ zum Mikrofonarray 4 Personen platziert. Aus Sicht des Mikrofonarrays befanden sich diese Personen ungefähr bei $\Theta_1 \approx 20^\circ$, $\Theta_2 \approx 60^\circ$, $\Theta_3 \approx 120^\circ$, $\Theta_4 \approx 160^\circ$. Diese Sprecher sprachen zuerst reihum einen kurzen Satz und dann alle zusammen, wobei der Text den Sprechern überlassen war. Die Ergebnisse der Breitband-Root-MUSIC-Analyse sind in Abbildung 4.16 zu sehen.

Anhand der in den Abbildungen 4.17 bis 4.21 dargestellten Histogramme der den jeweiligen Sprechern zugeordneten Signalausschnitten kann eine genauere statistische Auswertung der Ergebnisse vorgenommen werden. Aus diesen wird ersichtlich, daß Sprecher 1 und Sprecher 2 recht gut erfaßt werden. Weitere Maxima an den entsprechenden Positionen deuten darauf hin, daß die anderen Sprecher anscheinend nicht ganz lautlos gewesen sind. Die schlechtesten Ergebnisse ergaben sich für den 3. Sprecher; allerdings ist im Histogramm des direkt folgenden Blocks 4 eine starke Peilung bei $\Theta \approx 135^\circ$ dargestellt, welche vermutlich den 3. Sprecher repräsentiert. Sieht man sich die Darstellung der Peilungen in Abbildung `effig:realspeech` genauer an, so erkennt man eine gewisse Trägheit in der Erkennung der Quellen. Es vergeht einige Zeit, bis eine neue Quelle auftaucht bzw. die Peilung einer verstummten Quelle verschwindet. Dies ist auf das verwendete

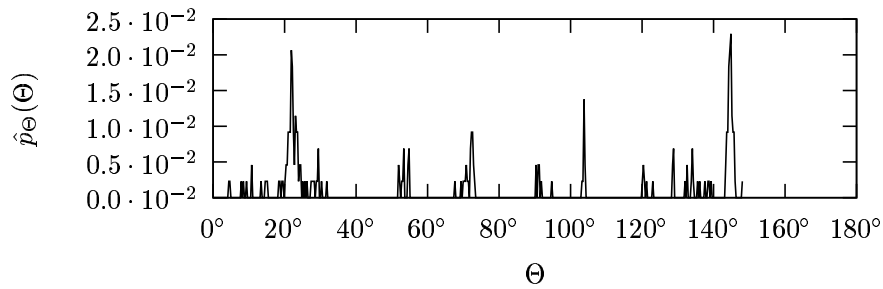


Abbildung 4.17: Histogramm der ermittelten DOA-Winkel für den ersten Sprecher (erwartet: $\Theta \approx 20^\circ$)

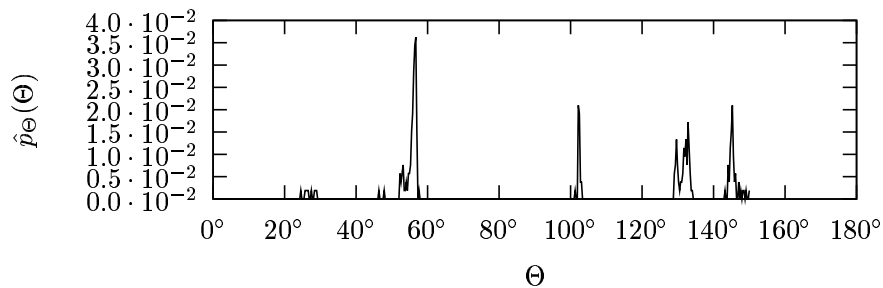


Abbildung 4.18: Histogramm der ermittelten DOA-Winkel für den zweiten Sprecher (erwartet: $\Theta \approx 60^\circ$)

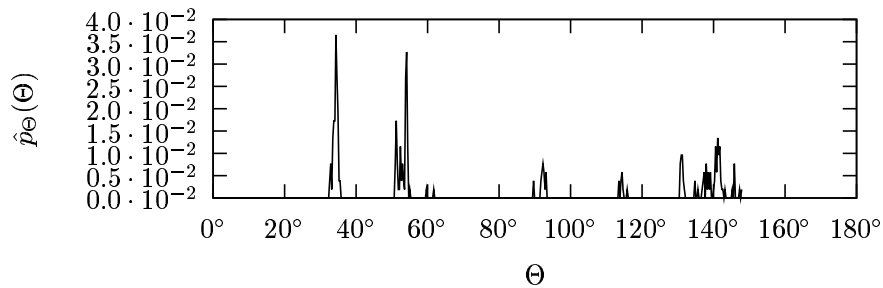


Abbildung 4.19: Histogramm der ermittelten DOA-Winkel für den 3. Sprecher (erwartet: $\Theta \approx 120^\circ$)

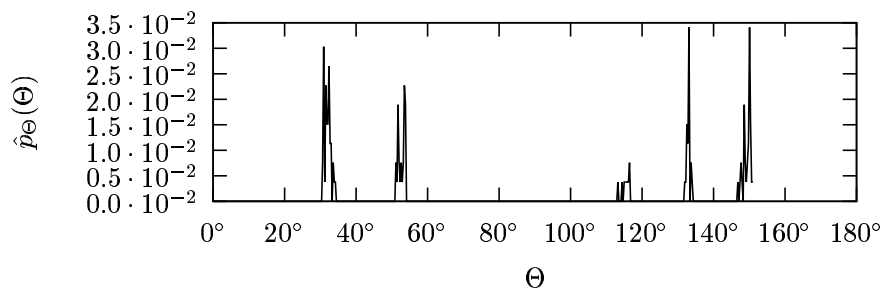


Abbildung 4.20: Histogramm der ermittelten DOA-Winkel für den 4. Sprecher (erwartet: $\Theta \approx 160^\circ$)

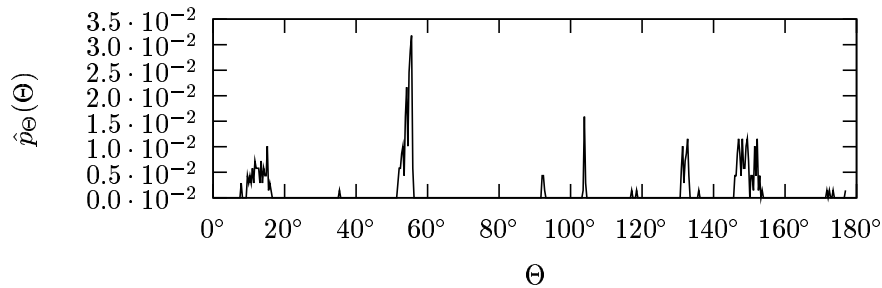


Abbildung 4.21: Histogramm der ermittelten DOA-Winkel für alle Sprecher (erwartet: $\Theta \approx 20^\circ, 60^\circ, 120^\circ$ und 160°)

floating-Average-Verfahren zurückzuführen. Dies würde erklären, warum sich die Peilung auf Sprecher 3 erst später etabliert hat. Sprecher, die über lange Zeit aktiv sind, können die *floating-average*-Approximation der spektralen Kovarianz so stark prägen, daß sie erst nach einiger Zeit „verdrängt“ werden. Diese „Prägung“ bewirkt auch eine schnellere Wiedererkennung von vor kurzer Zeit aktiveren Sprechern (welche sich noch an der gleichen Position befinden) – ihre „Spuren“ in der spektralen Kovarianzmatrix werden quasi nur wieder aufgefrischt. Leider wurde versäumt, die Positionen der Sprecher zum Array genau zu erfassen. Somit sind die angegebenen Peilwinkel nur als Schätzwerte zu betrachten.

4.4 Zusammenfassung

Mit der Evaluierung durch generierte Eingangssignale, die genau den theoretischen Annahmen entsprachen, konnte das Funktionieren der Implementierung verifiziert werden. Es wurden alle Quellen erkannt, die Peilungen waren auch bei einem $SNR_{dB} = 10 \text{ dB}$ stabil.

Im Anschluß daran wurde ein Problemfall modelliert, dessen Auftreten in der Praxis absehbar bzw. nicht unwahrscheinlich ist. Hier zeigte sich die Anfälligkeit des Algorithmus gegenüber Korrelationen in den Eingangssignalen, welche auch durch vorgeschaltete Dekorrelationsfilter nicht völlig behoben werden konnten. Die Evaluation mit den Signalen eines realen Mikrofonarrays zeigte aber, daß der Algorithmus auch ohne zusätzliche Vorfilterung in normalen Büroumgebungen funktioniert.

4.5 Ausblick

Die Evaluierung der realen Mikrofon-signale bzw. des Problemfalls eines Raums mit Nachhall zeigte, daß in der Praxis mit einer Verrauschung der Peilungen und Peilungen auf Reflektionspfade zu rechnen ist. Somit ist eine Nachbearbeitung der durch Kreuzpeilung gewonnenen Positionsdaten unabdingbar. Nach [EDDC]

führen Fehlpeilungen zu dünn besetzten oder unplausiblen Clustern, so daß schon allein durch einfache Plausibilitätsprüfungen und Maßnahmen wie einer Schwellwertsegmentierung der ermittelten Cluster nach ihrer mittleren Besetzung zu wesentlich besseren Positionsdaten führen sollten. Weiterhin wäre die Verwendung eines Sprachdetektors zur Erkennung von Sprachpausen sinnvoll, um in diesen Zeiträumen die Analyse abzuschalten.

Das Problem des hohen Rechenzeitaufwandes konnte zwar auf seine Ursache hin analysiert werden, leider konnte keine Lösung gefunden werden, um die rechenaufwendigste Teilberechnung zu vereinfachen. Deshalb sollte bei der Verwendung der Implementierung der analysierte Frequenzbereich, welcher ja proportional zur Anzahl der FFT-Bins ist, so weit wie möglich reduziert werden. Weiterhin sollte von der Trennung des Analysezykluses in das Updaten des spektralen Leistungsdichtespektrums und der eigentlichen Analyse Gebrauch gemacht werden.

Literaturverzeichnis

- [EDDC] Raffaele Parisi Elio D. Di Claudio. Multi-source localization strategies. In *Array Signal Processing*, pages 194–195.
- [KV96] HAMID KRIM and MATS VIBERG. Two decades of array signal processing research. *IEEE Signal Processing Magazine*, pages 67–94, 6 1996.
- [McC01] Ian A. McCowan. Microphone arrays, a tutorial. 2001.
- [Nol00] Prof. Dr.-Ing. Peter Noll. *Statistische Nachrichtentheorie*. TU Berlin, 1999/2000.
- [Roo] Dsp algorithms. URL: <http://www.dspalgorithms.com>.
- [sig98] *Digitale Sprachsignalverarbeitung*, chapter 12.8.2, pages 411–413. Number ISBN 3-519-06165-1. B.G. Teubner Stuttgart, 1998.
- [YHM01] Gary W. Elko Yiteng Huang, Jacob Benesty and Russel M. Mersereau. Real-time passive source localization: A practical linear-correlation least-squares approach. *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, 9(943-956), 11 2001.

Anhang A

Anhang

A.1 Mathematische Sätze

Hauptsatz über Hauptachsentransformationen. *Der Hauptsatz über Hauptachsentransformationen besagt, daß man zu jeder hermiteschen Matrix \mathbf{A} immer eine orthogonale Matrix \mathbf{U} so finden kann, daß*

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H \quad (\text{A.1})$$

mit

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \vdots & 0 \\ 0 & \lambda_2 & \vdots & 0 \\ \dots & \dots & \vdots & \dots \\ 0 & 0 & \vdots & \lambda_N \end{pmatrix} \quad (\text{A.2})$$

als Diagonalmatrix der Eigenwerte $\lambda_1, \dots, \lambda_N$ von \mathbf{A} .

Wiener-Khintchine-Beziehung. *Die Wiener-Khintchine-Beziehung besagt, daß die Kreuzkorrelation $L_{r_{xy}}(t)$ eines Leistungssignals und sein Leistungsdichtespektrum $S_{xy}(j\omega)$ Fouriertransformierte zueinander sind:*

$$\mathcal{F}\{L_{r_{xy}}(t)\} = S_{xy}(j\omega) \hat{=} \lim_{T \rightarrow \infty} \frac{1}{T} X(j\omega) \cdot Y(j\omega)^* \quad (\text{A.3})$$

Abbildungsverzeichnis

2.1	Das Umgebungsmodell	8
2.2	Das MUSIC-„Spektrum“	14
2.3	Abschätzung für Fernfeldbedingungsannahme	17
2.4	Maximale Winkelabweichung Δ in Abhängigkeit von d_S/d_A	18
2.5	Abschätzung für die nötige Entfernung nach [McC01].	19
2.6	Signalmodell für Fernfeldbedingungen	19
2.7	Differenz des errechneten DOA-Winkels $\hat{\Theta}$ vom tatsächlichen DOA-Winkel $\Theta = 45^\circ$; die Signalfrequenz beträgt $f_1 = 125 \text{ Hz}$	20
2.8	Prinzip des <i>spatial smoothings</i>	23
3.1	Datenflußgraph der Methode <code>update()</code> für ein Array mit 3 Mikrofonen	28
3.2	Datenflußgraph der Methode <code>estimate()</code>	29
3.3	Verlauf der Eigenwerte bei einer aktiven Rauschquelle ($SL = 128$, $f_{\text{Range}} = [0 \dots \frac{f_T}{2}]$)	30
3.4	Position der Nullstellen bei einer aktiven Rauschquelle ($SL = 128$, $f_{\text{Range}} = [0 \dots \frac{f_T}{2}]$), $\Theta = 125.25^\circ$, gemessen: 125.24° ($SNR_{dB} = \infty$, Symbol Δ) bzw. 124.98° ($SNR_{dB} = 10 \text{ dB}$, Symbol $+$)	31
3.5	Abschätzung des Aufwand eines <i>Root-Music</i> -Analysezyklus in Abhängigkeit von der Segmentlänge SL für ein Mikrofonarray mit 16 Mikrofonen.	34
4.1	Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-4}$, $SL = 128$, $SNR_{dB} = \infty$)	36
4.2	Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-6}$, $SL = 256$, $SNR_{dB} = \infty$)	36
4.3	Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-4}$, $SL = 128$, $SNR_{dB} = 10 \text{ dB}$)	37
4.4	Ergebnisse für die simulierten Rauschquellen bei $\Theta_1 = 20^\circ$, $\Theta_2 = 40^\circ$ und $\Theta_3 = 70^\circ$ ($\alpha = 0.01$, $prob = 10^{-6}$, $SL = 256$, $SNR_{dB} = 10 \text{ dB}$)	37
4.5	Szenario für die Analyse von Signalen mit Raumechos	39

4.6	Ergebnisse für das „verhallte-Räume-Szenario“ ($f_T = 8 \text{ kHz}$, $f = [1 \text{ kHz} \dots 4 \text{ kHz}]$, $\alpha = 0.05$, $prob = 10^{-7}$, $SL = 1024$)	40
4.7	Histogramm der ermittelten DOA-Winkel für das „verhallte-Räume-Szenario“	40
4.8	Raumimpulsantwort $S_1 \rightarrow \text{Mic}_1$	41
4.9	Raumimpulsantwort $S_2 \rightarrow \text{Mic}_1$	41
4.10	Raumimpulsantwort $S_1 \rightarrow \text{Mic}_2$	41
4.11	Raumimpulsantwort $S_2 \rightarrow \text{Mic}_2$	41
4.12	Ergebnisse für andere Sprecher und blockweise adaptive LPC-Vorfilterung durch ein Filter mit 128 Koeffizienten. ($f_T = 44,1 \text{ kHz}$, $f = [0 \dots 22.05 \text{ kHz}]$, $\alpha = 0.05$, $prob = 10^{-7}$, $SL = 1024$)	42
4.13	Histogramm der ermittelten DOA-Winkel für das „verhallte-Räume-Szenario“ mit anderen Sprechern und LPC-Vorfilterung	42
4.14	Ausschnitt $t \in [0 \dots 1.75 \text{ s}]$ aus Abbildung 4.12	43
4.15	Histogramm für den Ausschnitt $t \in [0 \dots 1.75 \text{ s}]$ aus Abbildung 4.14	43
4.16	Ergebnisse für reale Mikrofondaten ($f_T = 48 \text{ kHz}$, $f = [1 \text{ kHz} \dots 20 \text{ kHz}]$, $\alpha = 0.05$, $prob = 10^{-7}$, $SL = 1024$), Bereich 1-4: Sprecher 1-4, Bereich 5: alle Sprecher aktiv	44
4.17	Histogramm der ermittelten DOA-Winkel für den ersten Sprecher (erwartet: $\Theta \approx 20^\circ$)	45
4.18	Histogramm der ermittelten DOA-Winkel für den zweiten Sprecher (erwartet: $\Theta \approx 60^\circ$)	45
4.19	Histogramm der ermittelten DOA-Winkel für den 3. Sprecher (erwartet: $\Theta \approx 120^\circ$)	45
4.20	Histogramm der ermittelten DOA-Winkel für den 4. Sprecher (erwartet: $\Theta \approx 160^\circ$)	45
4.21	Histogramm der ermittelten DOA-Winkel für alle Sprecher (erwartet: $\Theta \approx 20^\circ, 60^\circ, 120^\circ$ und 160°)	46