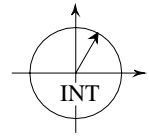


UNIVERSITÄT KARLSRUHE
INSTITUT FÜR NACHRICHTENTECHNIK
Prof. Dr.rer.nat. Friedrich Jondral



Passives Audiotracking

sich bewegender Geräuschquellen

Studienarbeit von

Michael Grimm

Hauptreferent : Prof. Dr.-Ing. Kristian Kroschel
Betreuer : Dipl.-Ing. Dirk Bechler

Beginn : 17.05.02
Abgabe : 31.07.02

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Studienarbeit selbständig und ohne unzulässige fremde Hilfe angefertigt habe. Die verwendeten Literaturquellen sind im Literaturverzeichnis vollständig zitiert.

Karlsruhe, den 31.07.2002

Michael Grimm
Danngrabenweg 7
77815 Bühl

Danksagung

Hiermit möchte ich mich bei allen Personen bedanken, die mir bei der Entstehung dieser Arbeit geholfen haben.

Mein Dank gilt insbesondere Herrn Prof. Kristian Kroschel, der mir diese Studienarbeit ermöglicht hat, und Herrn Dipl.-Ing. Dirk Bechler für die gute Betreuung. Weiterhin danke ich neben allen Mitarbeitern des Instituts auch meiner Familie und meiner Freundin für die Unterstützung.

Inhaltsverzeichnis

1	Einleitung	1
2	Entwurf des Kalman-Filters	5
2.1	Zustandsraumdarstellung	5
2.2	Wahl des Modells	6
2.3	Filterentwurf	9
2.4	Filteralgorithmus	10
3	Parametrisierung des Kalman-Filters	13
3.1	Messrauschen	13
3.2	Systemrauschen	15
3.3	Anfangswerte	15
4	Beurteilung des Filters	19
4.1	Fehlerbetrachtung	19
4.2	Konsistenz	20
4.3	Modellwahrscheinlichkeit	23
5	Adaption an reale Szenarien	27
5.1	Testumgebung	27
5.2	Reine Prädiktion	28
5.3	Trackingpausen	30
5.4	Adaptives Kalman-Filter	31
6	Ergebnisse	35
6.1	Zusammenfassung	35
6.2	Tracking einer ruhenden Geräuschquelle	36
6.3	Tracking einer linear bewegten Geräuschquelle	38
6.4	Tracking einer beschleunigten Geräuschquelle	42
7	Ausblick	43
A	Koordinatentransformation	45

Inhaltsverzeichnis

B	Aufnahmen	47
B.1	Übersicht	47
B.2	Testsätze	49
C	Tabellen zur χ^2 -Verteilung	51

Abbildungsverzeichnis

1.1	Einbettung des Kalman-Filters in die Lokalisationsaufgabe	1
1.2	Versuchsaufbau mit Mikrofonarray	2
4.1	Fehlervergleich	20
4.2	Konsistenztest mit unterschiedlichem Messrauschen	22
4.3	Modellwahrscheinlichkeiten	24
5.1	Zeitlicher Abstand zweier Messwerte	28
5.2	Verlauf der Schätzung bei Prädiktion	29
5.3	Erhöhte Fehlervarianz bei Prädiktion	29
5.4	Verlauf der Schätzung bei Trackingpausen	31
5.5	Struktur des adaptiven Kalman-Filters	32
6.1	Sphärische Koordinaten beim Tracken einer ruhenden Geräuschquelle	37
6.2	Koordinatenverläufe der vorgegebenen Trajektorie	39
6.3	Sphärische Koordinaten beim Tracken einer linear bewegten Geräuschquelle	40
6.4	Lokalisation einer linear bewegten Geräuschquelle (vor dem Filtervorgang)	41
6.5	Tracking einer linear bewegten Geräuschquelle (nach dem Filtervorgang)	41
C.1	Untere Quantile der χ^2 -Verteilung	52
C.2	Obere Quantile der χ^2 -Verteilung	53

Tabellenverzeichnis

4.1	Parameter für Konsistenztests	21
B.1	Aufgenommene Szenarien	49
B.2	Testsätze zur Spracherkennung nach Beckmann-Schilling	50

1 Einleitung

Diese Arbeit ist dem Sonderforschungsbereich 588 *Humanoide Roboter - Lernende und kooperierende multimodale Roboter* an der Universität Karlsruhe (TH) zugeordnet, speziell dem Teilprojekt P3 *Akustische Szenenanalyse*. Die darin enthaltenen Hauptaufgaben umfassen die Lokalisation, Trennung und Klassifikation von Geräuschquellen. Dem Bereich der Lokalisation ist diese Arbeit zugeordnet.

Das Ziel ist dabei, dass ein mit einem statischen Mikrofonarray ausgestatteter Roboter die Geräuschquellen in seinem Umfeld lokalisieren und verfolgen kann (Audio-Tracking). Allein mit Hilfe der akustischen Signale soll der Roboter in der Lage sein, sich ein Bild von seiner Umgebung zu machen und eine sogenannte *acoustic map* zu erstellen. Zur Positionsbestimmung einer Geräuschquelle wird die Information aus den Zeitverzögerungen der akustischen Signale an den einzelnen Mikrofonen verwendet, wie es im folgenden Abschnitt dargestellt wird. Die berechnete Position ist jedoch mit Fehlern behaftet und weicht von der tatsächlichen Position der Quelle ab. Im statischen akustischen Umfeld kann zur Lokalisation eine Mittel- oder Medianwertbildung über große Zeitfenster herangezogen werden. Bei bewegten Geräuschquellen kann diese Mittelung nicht mehr verwendet werden, um die Geräuschquelle akustisch zu verfolgen. Aus diesem Grund muss eine kontinuierliche Filterung der Positionsschätzung erfolgen. Der Entwurf des dazu verwendeten *Kalman-Filters* und die dadurch erreichte Glättung der Positionsschätzungen soll in dieser Arbeit untersucht werden.

Abbildung 1.1 zeigt die gesamte Lokalisationsaufgabe für unseren Fall, bei dem fünf Mikrofone

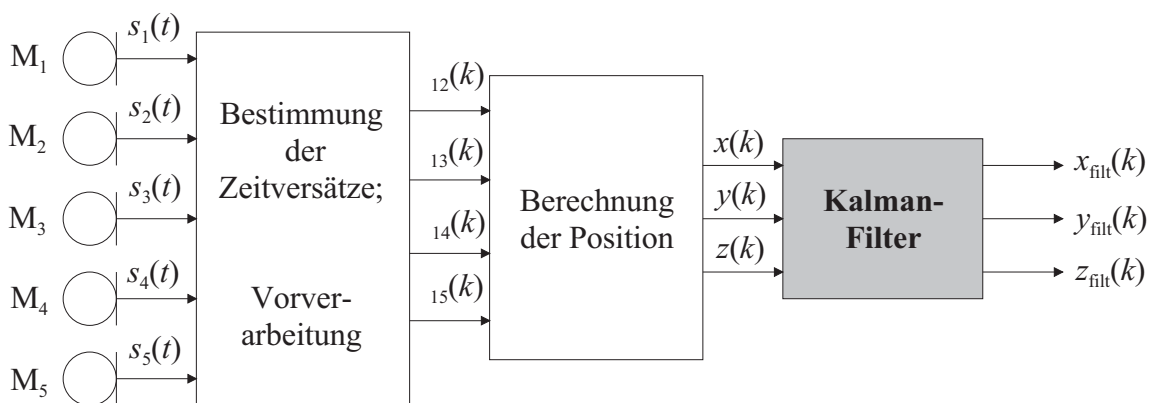


Abbildung 1.1: Einbettung des Kalman-Filters in die Lokalisationsaufgabe

in einer Doppel-Tetraederanordnung verwendet werden. Die Mikrofonanordnung zeigt Abbil-

1 Einleitung

Abbildung 1.2. Die akustischen Signale kommen zunächst an den Mikrofonen M_i , $i = 1, \dots, 5$, an und werden von diesen aufgenommen ($s_i(t)$). Die Lokalisation der Quelle erfolgt dann entspre-



Abbildung 1.2: Versuchsaufbau mit Mikrofonarray

chend der populärsten und in Bezug auf Echtzeitfähigkeit leistungsfähigsten Technik in zwei Schritten [BS97a],[BS97b]:

1. Bestimmung der Zeitversätze des akustischen Signals bei der Ankunft an den Mikrofonpaaren
2. Berechnung der Position aus den Zeitversatzschätzungen und der Geometrie der Mikrofonanordnung.

Das Mikrofon M_1 dient dabei als Referenz-Mikrofon, auf das die anderen Signale bezogen werden, um insgesamt vier Zeitversätze zu bestimmen. Die am häufigsten verwendete Methode zur Schätzung dieser Zeitversätze ist eine modifizierte Form der Kreuzkorrelationsfunktion (*GCC, generalized cross-correlation*), bei der ein Prewhitening durchgeführt wird (*PHAT, Phase Transformation*) [KC76], [OS97]. Die Stelle des Maximums τ_{1j} , $j = 2, \dots, 5$ entspricht dann der

berechneten Zeitversatzschätzung *TDE* (*time delay estimate*). Der tatsächliche Zeitversatz wird in der Literatur als *TDOA* (*time difference of arrival*) bezeichnet, um ihn von dem Schätzwert zu unterscheiden. Diese Zeitversatzschätzungen werden dann über ein kurzes Zeitfenster gemittelt, da zum Tracken von Sprechern mindestens vier Messwerte pro Sekunde erforderlich sind [SK92] und so zusätzliche Verlässlichkeit der TDEs erreicht werden kann.

Die TDEs und die Information über die Mikrofonarray-Geometrie erlauben die Berechnung der Quellenposition. Denn durch jeden Zeitversatz ist im Raum eine hyperboloide Fläche festgelegt, auf der sich die Geräuschquelle befinden muss. Durch Schnittmengenbildung kann dieser Lösungsraum zunächst auf Peilgeraden (*bearing lines*) und letztlich auf die tatsächliche Position im Raum begrenzt werden. Die funktionale Abhängigkeit der Position von den TDEs ist äußerst nichtlinear, zu deren Lösung verschiedene Möglichkeiten bestehen. Der verwendete *OSLS*-Algorithmus (*One Step Least Squares Locator*) stellt eine suboptimale *closed-form* Lösung dar, die für die Berechnung der Position verwendet werden kann [HE00]. Dieser Algorithmus minimiert den quadratischen Fehler bei der Umrechnung, reduziert den numerischen Aufwand aber beträchtlich, ohne die statistische Genauigkeit zu vermindern und erreicht somit Echtzeitfähigkeit.

Auf Grund von Störeinflüssen wie Rauschen, Reflexionen an den Wänden, einer begrenzten Zeitauflösung durch die Abtastung der Audio-Signale und Systemfehler bei der Array-Anordnung streuen die TDEs und damit die geschätzten Koordinaten im Vergleich zu den tatsächlichen Positionen der Quellen. Im Folgenden wird dafür der Entwurf eines Kalman-Filters dargestellt, das die berechnete Positionsschätzung räumlich glättet und so ein robustes Tracken einer bewegten Geräuschquelle ermöglicht.

Nach dem grundsätzlichen Entwurf des Filters in Kapitel 2 wird in Kapitel 3 ausführlich auf die Parameter eingegangen, da diese die Güte des Filters wesentlich beeinflussen. Im Anschluss daran werden in Kapitel 4 verschiedene Kriterien zur Beurteilung des Filters dargestellt. In Kapitel 5 wird noch einmal konkret der Praxisbezug hergestellt und beschrieben, welche Änderungen dafür notwendig sind. Abschließend werden die Ergebnisse zusammengetragen und diskutiert.

1 Einleitung

2 Entwurf des Kalman-Filters

In diesem Abschnitt soll der grundsätzliche Entwurf eines Kalman-Filters zum passiven Audio-Tracken vorgestellt werden. Dabei liegt der Schwerpunkt eher darauf, die Ansätze und Strukturen zu erläutern, als eine komplette Herleitung zu vollziehen. Dafür wird auf die Literaturquellen [Kro96] und [Bro83] verwiesen. Die Anwendung, eine bewegte Geräuschquelle rein akustisch zu verfolgen, soll dabei nicht aus dem Blickfeld verloren gehen.

2.1 Zustandsraumdarstellung

Aufgrund der Instationarität des Bewegungs- und Störprozesses reichen die Korrelationsfunktionen wie beim Einsatz von Wiener Filtern für eine vollständige Beschreibung nicht aus. Deshalb wird stattdessen die Zustandsraumdarstellung verwendet, welche bei Abtastung mit der Abtastperiode T in zeitdiskreter Form durch zwei Gleichungen gegeben ist:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{B}u(k) \quad (2.1)$$

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{n}(k) \quad (2.2)$$

Gleichung (2.1) wird als *Systemgleichung* bezeichnet. Darin ist

- $\mathbf{x}(k)$ der *Zustandsvektor* der Dimension N , der den Systemzustand zum Zeitpunkt kT beinhaltet
- \mathbf{A} die zeitinvariante *Übergangsmatrix* der Dimension $N \times N$ (Systemmatrix)
- $u(k)$ weißes Rauschen mit der Varianz σ_U^2 , das den Systemfehler repräsentiert
- \mathbf{B} ein zeitinvarianter Vektor der Dimension $N \times 1$, der den Systemfehler auf die einzelnen Elemente des Zustandsvektors abbildet (*Systemrauschkoppelmatrix*)

2 Entwurf des Kalman-Filters

In der *Messgleichung* (2.2) haben die einzelnen Elemente folgende Bedeutung:

- $\mathbf{y}(k)$ ist der M -dimensionale *Messvektor* zum Zeitpunkt kT , der die aus den Zeitversätzen berechnete Position der Geräuschquelle enthält
- \mathbf{C} ist die zeitinvariante *Beobachtungsmatrix* der Dimension $M \times N$, die den Systemzustand auf den Messvektor abbildet
- $\mathbf{n}(k)$ ist weißes Rauschen mit der Kovarianzmatrix $\mathbf{R}_{NN}(k)$, das den Messfehler repräsentiert.

2.2 Wahl des Modells

Das akustische Umfeld des Roboters soll im betrachteten Fall nur aus ruhenden Geräuschquellen und menschlichen Sprechern bestehen, die sich beliebig im Raum bewegen. Damit ergeben sich für das Kalman-Filter drei sinnvolle Bewegungsmodelle.

Stationäres Modell: Als Zustandsvektor wird $\mathbf{x} = (x, y, z)^T$, also nur die Position in kartesischen Koordinaten, gewählt. Dann modelliert $u(k)$ die Geschwindigkeit mit dem Mittelwert $E\{U\} = 0$ und der Varianz $E\{U^2\} = \sigma_U^2$. Es gilt für jede Koordinate $\xi \in \{x, y, z\}$

$$\xi(k+1) = \xi(k) \tag{2.3}$$

und somit für die Matrizen der Systemgleichung

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ und } \mathbf{B} = \begin{pmatrix} T \\ T \\ T \end{pmatrix}.$$

Der Messvektor $\mathbf{y}(k) = (x, y, z)^T$ enthält ebenfalls die Position, so dass in der Messgleichung für die Beobachtungsmatrix \mathbf{C} die Einheitsmatrix zu wählen ist:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Lineares Modell: Dieses Modell wird dem Gehen eines Sprechers am besten entsprechen. Die konstante Geschwindigkeit wird in Form der zeitlichen Ableitungen $\dot{x} = \frac{dx}{dt}$, $\dot{y} = \frac{dy}{dt}$ und $\dot{z} = \frac{dz}{dt}$ in den Zustandsvektor $\mathbf{x} = (x, y, z, \dot{x}, \dot{y}, \dot{z})^T$ mit aufgenommen. Entsprechend den Newton'schen Bewegungsgleichungen berechnet man mit der Abtastzeit T für jede Koordinate $\xi \in \{x, y, z\}$ den Übergang vom Zeitpunkt kT zu $(k+1)T$

$$\xi(k+1) = \xi(k) + T\dot{\xi}(k) \quad (2.4)$$

$$\dot{\xi}(k+1) = \dot{\xi}(k). \quad (2.5)$$

U modelliert in diesem Fall die mittelwertfreie Beschleunigung, sodass sich folgende Matrizen der Systemgleichung ergeben:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \frac{1}{2}T^2 \\ \frac{1}{2}T^2 \\ \frac{1}{2}T^2 \\ T \\ T \\ T \end{pmatrix}.$$

Da mit $\mathbf{y}(k) = (x, y, z)^T$ wieder nur die Position beobachtet werden kann, folgt für die Matrix der Messgleichung

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Beschleunigtes Modell: Entsprechend den anderen beiden Fällen entwirft man auch das Modell für konstante Beschleunigung nach den Newton'schen Bewegungsgleichungen. Durch zweifache zeitliche Ableitung der Position $\ddot{x} = \frac{d^2x}{dt^2}$, $\ddot{y} = \frac{d^2y}{dt^2}$ und $\ddot{z} = \frac{d^2z}{dt^2}$ kann diese in den Zustandsvektor mit aufgenommen werden: $\mathbf{x} = (x, y, z, \dot{x}, \dot{y}, \dot{z}, \ddot{x}, \ddot{y}, \ddot{z})^T$. Beim Übergang von kT zu $(k+1)T$ berechnen sich die einzelnen Elemente und deren erste beiden Ableitungen für $\xi \in \{x, y, z\}$ zu

$$\xi(k+1) = \xi(k) + T\dot{\xi}(k) + \frac{1}{2}T^2\ddot{\xi} \quad (2.6)$$

$$\dot{\xi}(k+1) = \dot{\xi}(k) + T\ddot{\xi}(k) \quad (2.7)$$

$$\ddot{\xi}(k+1) = \ddot{\xi}(k). \quad (2.8)$$

2 Entwurf des Kalman-Filters

Die Änderung der Beschleunigung wird mit U berücksichtigt; sie wirkt sich additiv durch $\frac{1}{6}T^3 u(k)$ auf ξ aus bzw. durch $\frac{1}{2}T^2 u(k)$ auf $\dot{\xi}$ und durch $Tu(k)$ auf $\ddot{\xi}$. Damit kann man die Matrizen folgendermaßen angeben, wenn man berücksichtigt, dass durch die Beobachtungsmatrix wie bei den anderen beiden Modellen nur die Abbildung des Zustandsvektors auf den Messvektor $\mathbf{y}(k) = (x, y, z)^T$ geschieht:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & T & 0 & 0 & \frac{1}{2}T^2 & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 & 0 & \frac{1}{2}T^2 & 0 \\ 0 & 0 & 1 & 0 & 0 & T & 0 & 0 & \frac{1}{2}T^2 \\ 0 & 0 & 0 & 1 & 0 & 0 & T & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \frac{1}{6}T^3 \\ \frac{1}{6}T^3 \\ \frac{1}{6}T^3 \\ \frac{1}{2}T^2 \\ \frac{1}{2}T^2 \\ \frac{1}{2}T^2 \\ T \\ T \\ T \end{pmatrix} \text{ und}$$

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Das beschleunigte Modell findet Verwendung, wenn z.B. beim Sprecher Übergänge zwischen stationärem Verhalten und Gehen mit konstanter Geschwindigkeit auftreten. Es wird sich später zeigen, dass das beschleunigte Modell bei unserer Anwendung selten Verwendung findet, da durch Vorverarbeitung der Mikrofondifferenzsignale zur Bestimmung der TDOA Schätzungen bereits eine Mittelung durchgeführt wird, die die entsprechende Dynamik im Bewegungsablauf reduziert.

Die Dimension des Messvektors ist offensichtlich bei allen drei Modellen

$$M = \dim \mathbf{y}(k) = 3,$$

während sich die Dimension des Zustandsvektors je nach Modell unterscheidet:

$$N = \dim \mathbf{x}(k) = \begin{cases} 3 & \text{beim stationären Modell} \\ 6 & \text{beim linearen Modell} \\ 9 & \text{beim beschleunigten Modell} \end{cases}$$

Der Rechenaufwand zur Invertierung von Matrizen kann durch Berechnung der dritten Potenz der Dimension nach oben abgeschätzt werden. Da für das Kalman-Filter nur eine Inversion der Dimension $M \times M$ nötig ist, ist somit der Rechenaufwand der Ordnung $O(M^3)$ grundsätzlich auch bei einer Umsetzung in ein echtzeitfähiges System problemlos.

Entsprechend der tatsächlichen Bewegung wird das passende Modell jeweils das beste Filterergebnis liefern. In Abschnitt 5.4 wird gezeigt, wie man Kalman-Filter mit verschiedenen Modellen zu einem adaptiven System kombiniert. Zunächst soll aber der Filterentwurf unabhängig vom Modell erläutert werden.

2.3 Filterentwurf

Das Ziel beim Entwurf des Kalman-Filters besteht darin, auf der Grundlage von fehlerbehafteten Messungen $\mathbf{y}(1), \dots, \mathbf{y}(k)$ eine Schätzung $\hat{\mathbf{y}}(k)$ für die tatsächliche Position des durch die Gleichungen (2.1) und (2.2) beschriebenen Systems. Dies geschieht, indem der Zustand des Systems $\mathbf{x}(k)$ fortlaufend geschätzt wird (Bezeichnung $\hat{\mathbf{x}}(k)$) und über die Beobachtungsmatrix \mathbf{C} in die Positionsschätzung abgebildet wird. Als Gütekriterium wählt man den mittleren quadratischen Fehler; dieser soll minimiert werden:

$$E \{ \mathbf{e}(k) \mathbf{e}^T(k) \} = E \{ (\mathbf{x}(k) - \hat{\mathbf{x}}(k)) (\mathbf{x}(k) - \hat{\mathbf{x}}(k))^T \} \rightarrow \min$$

Unter der Voraussetzung, dass sowohl Nutz- und Störprozess als auch System- und Messrauschen unkorreliert sind, wird mit einem Kalman-Filter dieses Ziel optimal erreicht. Die vollständige Herleitung mit dem Orthogonalitätsprinzip kann in [BS88] nachvollzogen werden.

In der Literatur wird dabei zwischen dem *A priori* Schätzwert $\hat{\mathbf{x}}(k|k-1)$ und dem *A posteriori* Schätzwert $\hat{\mathbf{x}}(k|k)$ unterschieden [WB01]. In dieser Notation gibt das erste Argument den Zeitpunkt an, für den der Systemzustand geschätzt werden soll, während das zweite Argument hinter dem Bedingungsstrich den Zeitpunkt angibt, bis zu dem die gestörten Empfangsvektoren zur Verfügung stehen. Damit ist $\hat{\mathbf{x}}(k|k-1)$ der Schätzwert zum Zeitpunkt kT , für den nur Messwerte bis zum Zeitpunkt $(k-1)T$ verwendet werden, wohingegen $\hat{\mathbf{x}}(k|k)$ die Schätzung des Zustands zum Zeitpunkt kT angibt unter der zusätzlichen Berücksichtigung des neuen Messwertes $\mathbf{y}(k)$. Damit kann man folgenden linearen Ansatz machen:

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{W}(k)[\mathbf{y}(k) - \hat{\mathbf{y}}(k|k-1)]. \quad (2.9)$$

In Gleichung (2.9) wird $\mathbf{W}(k)$ als *Filter Gain* und

$$\nu(k) = \mathbf{y}(k) - \hat{\mathbf{y}}(k|k-1) \quad (2.10)$$

als *Innovation* bezeichnet. Die Innovation beinhaltet die mit dem aktuellen Messvektor $\mathbf{y}(k)$ neu hinzugekommene Information über $\mathbf{x}(k)$, die aus den Messvektoren bis zum Zeitpunkt $(k-1)T$ nicht gewonnen werden konnte. Mit der *A priori* Schätzfehlerkovarianz

$$\mathbf{R}_{EE}(k|k-1) = E \{ (\mathbf{x}(k) - \hat{\mathbf{x}}(k|k-1)) (\mathbf{x}(k) - \hat{\mathbf{x}}(k|k-1))^T \} \quad (2.11)$$

2 Entwurf des Kalman-Filters

gilt

$$\mathbf{W}(k) = \mathbf{R}_{EE}(k|k-1)\mathbf{C}^T[\mathbf{C}\mathbf{R}_{EE}(k|k-1)\mathbf{C}^T + \mathbf{R}_{NN}(k)]^{-1}. \quad (2.12)$$

Sowohl der Schätzwert für den Zustand des Systems als auch der Fehler können somit rekursiv berechnet werden.

2.4 Filteralgorithmus

Die Zustandsschätzung und die Fehlerkovarianzmatrix können beim Kalman-Filter in einem Algorithmus sequenziell berechnet werden. Die einzelnen Schritte sollen beim Übergang vom Zeitpunkt kT zu $(k+1)T$ im Folgenden dargestellt werden.

Für die Berechnung der Schätzfehlerkovarianz ist der neue Messwert nicht nötig; deshalb kann diese auch vor der Berechnung des neuen Schätzwerts durchgeführt werden. Für die Ableitung von $\mathbf{R}_{EE}(k+1|k+1)$ aus der vorherigen Kovarianzmatrix $\mathbf{R}_{EE}(k|k)$ berechnet man zunächst die *A priori* Schätzfehlerkovarianz

$$\mathbf{R}_{EE}(k+1|k) = \mathbf{A}\mathbf{R}_{EE}(k|k)\mathbf{A}^T + \mathbf{B}\sigma_U^2\mathbf{B}^T \quad (2.13)$$

und daraus die Kovarianzmatrix der Innovation $\mathbf{R}_{\nu\nu}(k+1) = E\{\nu(k+1)\nu^T(k+1)\}$ mit Gleichung (2.10) zu

$$\mathbf{R}_{\nu\nu}(k+1) = \mathbf{C}\mathbf{R}_{EE}(k+1|k)\mathbf{C}^T + \mathbf{R}_{NN}(k+1). \quad (2.14)$$

Diese Größe wird in Kapitel 4.2 zur Konsistenzbeurteilung herangezogen, um eine Aussage über die Parameterwahl treffen zu können. Damit lässt sich Gleichung (2.12) auch einfacher ausdrücken:

$$\mathbf{W}(k+1) = \mathbf{R}_{EE}(k+1|k)\mathbf{C}^T(\mathbf{R}_{\nu\nu}(k+1))^{-1}. \quad (2.15)$$

Die aktualisierte Schätzfehlerkovarianz zum Zeitpunkt $(k+1)T$ berechnet sich schließlich zu

$$\mathbf{R}_{EE}(k+1|k+1) = \mathbf{R}_{EE}(k+1|k) - \mathbf{W}(k+1)\mathbf{R}_{\nu\nu}(k+1)\mathbf{W}^T(k+1). \quad (2.16)$$

In ähnlicher Weise kann man den Schätzwert für den Zustand des Systems schrittweise aus dem vorigen ableiten. Zunächst wird der *A priori* Schätzwert

$$\hat{\mathbf{x}}(k+1|k) = \mathbf{A}\hat{\mathbf{x}}(k|k) \quad (2.17)$$

und daraus die geschätzte Position

$$\hat{\mathbf{y}}(k+1|k) = \mathbf{C}\hat{\mathbf{x}}(k+1|k) \quad (2.18)$$

berechnet. Mit der Innovation und dem Filter Gain ergibt sich dann die neue Zustandsschätzung

$$\hat{\mathbf{x}}(k+1|k+1) = \hat{\mathbf{x}}(k+1|k) + \mathbf{W}(k+1)\nu(k+1) \quad (2.19)$$

und die Positionsschätzung

$$\hat{\mathbf{y}}(k+1|k+1) = \mathbf{C}\hat{\mathbf{x}}(k+1|k+1). \quad (2.20)$$

Man kann die Gleichungen (2.19) und (2.16) auch direkt in Abhängigkeit von $\hat{\mathbf{x}}(k|k)$ bzw. $\mathbf{R}_{EE}(k|k)$ schreiben. Noch ungeklärt ist bis zu diesem Zeitpunkt, wie die Anfangswerte $\hat{\mathbf{x}}(0)$ und $\mathbf{R}_{EE}(0)$ zu wählen sind; ebenso sind noch die Größen $\mathbf{R}_{NN}(k)$ und σ_V^2 unbestimmt. Darauf wird im folgenden Abschnitt genauer eingegangen, da sie von entscheidender Bedeutung für den Nutzen des Filters sind.

2 Entwurf des Kalman-Filters

3 Parametrisierung des Kalman-Filters

Für eine gute Glättung der Messdaten ist es wichtig, das Kalman-Filter richtig zu parametrisieren. Dies betrifft neben dem Mess- und Systemrauschen auch die Wahl der Anfangswerte.

3.1 Messrauschen

Das Messrauschen $\mathbf{n}(k)$ in Gleichung (2.2) berücksichtigt den Messfehler, der bei der Bestimmung der Position $\mathbf{y}(k)$ gemacht wird. Die zugehörige Kovarianzmatrix wird mit

$$\mathbf{R}_{NN}(k) = E \{ \mathbf{n}(k) \mathbf{n}^T(k) \} \quad (3.1)$$

bezeichnet.

Um die Größenordnung der einzelnen Elemente zu bestimmen, wurde bei 20 Aufnahmen in Büroumgebung numerisch die Kovarianz des Messfehlers bestimmt. Es stellte sich dabei heraus, dass für die sphärischen Koordinaten Azimut θ , Elevation ϕ und Abstand r die Varianzen bei allen Bewegungsformen der Geräuschquelle zeitlich nahezu konstant bleiben. Im Gegensatz dazu hängen die Varianzen in x-, y- und z-Richtung von der aktuellen Position und damit vom Zeitpunkt der Messung ab. Dies ist damit zu erklären, dass bei der Lokalisation der Geräuschquelle ein Algorithmus verwendet wird, der zur Folge hat (siehe Kapitel 1), dass die Bestimmung von Azimut, Elevation und Abstand weitgehend unkorreliert ist, während die x-, y- und z-Koordinate große Kovarianzen aufweisen. Es wurde beschrieben, dass aus den Zeitversätzen τ_{ij} der Mikrofonpaare $M_i M_j$ in der Ankunft der akustischen Signale an den einzelnen Mikrofonpaaren zunächst Peilgeraden berechnet werden. Bei der Schnittmengenbildung zweier Peilgeraden tritt der sogenannte *Schereneffekt* auf, d.h. schon bei einem kleinen Fehler ändert sich der Abstand zum Schnittpunkt erheblich. Demzufolge ist bei dem Lokalisationsalgorithmus ein großer Fehler für den Abstand und ein kleiner Fehler für Azimut und Elevation zu erwarten. Diese Annahmen werden durch die ermittelten Werte

$$\begin{aligned} 2.0^\circ &\leq \sigma_\Theta \leq 6.4^\circ \\ 0.7^\circ &\leq \sigma_\Phi \leq 2.4^\circ \\ 0.14 \text{ m} &\leq \sigma_R \leq 0.82 \text{ m} \end{aligned} \quad (3.2)$$

bestätigt.

Da sich ein Sprecher in der Regel auf direktem Wege auf etwas zubewegt und nicht entlang

3 Parametrisierung des Kalman-Filters

von Kreisbögen, muss von einer Implementierung des Kalman-Filters in sphärischen Koordinaten abgesehen werden. Vielmehr wäre es wünschenswert, die Kovarianzmatrix von sphärischen Koordinaten

$$\mathbf{R}_{NN}^{(sph)} = \begin{pmatrix} \sigma_{\Theta}^2 & 0 & 0 \\ 0 & \sigma_{\Phi}^2 & 0 \\ 0 & 0 & \sigma_R^2 \end{pmatrix} \quad (3.3)$$

in kartesische zu transformieren. Diese Transformation ist jedoch in geschlossener Form nicht lösbar. Man kann unter der Annahme der statistischen Unabhängigkeit der einzelnen Komponenten aus der gemeinsamen Dichtefunktion in sphärischen Koordinaten

$$f_{\Theta,\Phi,R}(\theta, \phi, r) = f_{\Theta}(\theta) \cdot f_{\Phi}(\phi) \cdot f_R(r) \quad (3.4)$$

zwar nach [Jon00] die gemeinsame Dichtefunktion in kartesischen Koordinaten (x, y, z) berechnen. Diese ergibt sich zu

$$f_{X,Y,Z}(x, y, z) = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_{\Theta} \sigma_{\Phi} \sigma_R} \exp \left[-\frac{1}{2} \left(\left(\frac{\arctan \frac{y}{x}}{\sigma_{\Theta}} \right)^2 + \left(\frac{\arctan \frac{z}{\sqrt{x^2+y^2}}}{\sigma_{\Phi}} \right)^2 + \left(\frac{\sqrt{x^2+y^2+z^2}}{\sigma_R} \right)^2 \right) \right] \cdot \frac{\sqrt{x^2+y^2}}{(x^2+y^2+z^2)|x|}, \quad (3.5)$$

doch weder die einzelnen Randdichten noch die Erwartungswerte für die Kovarianzen sind daraus in geschlossener Form berechenbar. Eine detaillierte Betrachtung findet sich in Anhang A.

Für das Kalman-Filter wird nun folgende numerische Näherungslösung verwendet. Für jeden Zeitpunkt kT wird der Messwert $\mathbf{y}(k)$ in sphärische Koordinaten umgewandelt. Mit dem ganzzahligen Parameter $\lambda > 1$ wird daraus die $3 \times \lambda$ -dimensionale Testmatrix

$$\mathbf{Y}(k)^{(sph)} = \underbrace{\begin{pmatrix} \mathbf{y}(k) & \cdots & \mathbf{y}(k) \end{pmatrix}}_{\lambda \text{ Spalten}}$$

generiert. Nachdem zu jeder Zeile von $\mathbf{Y}(k)^{(sph)}$ ein $1 \times \lambda$ -dimensionaler Vektor von Zufallszahlen mit der entsprechenden Varianz aus (3.2) addiert worden ist, werden die einzelnen Spaltenvektoren der verrauschten Matrix $\mathbf{Y}(k)_{noise}^{(sph)}$ wieder in kartesische Koordinaten umgewandelt. Für diese Matrix $\mathbf{Y}(k)_{noise}^{(kart)} = ((y_{ij}))$ mit $1 \leq i \leq 3$ und $1 \leq j \leq \lambda$ können nun numerisch die Elemente der Kovarianzmatrix $\mathbf{R}_{NN}^{(kart)} = ((r_{pq}))$ bestimmt werden. Für ein Element r_{pq} mit $1 \leq p, q \leq 3$ gilt

$$r_{pq} = \frac{1}{\lambda - 1} \sum_{k=1}^{\lambda} (y_{pk} - \bar{y}_p)(y_{qk} - \bar{y}_q), \quad (3.6)$$

wobei die Mittelwerte \bar{y}_i , $1 \leq i \leq 3$, gerade die drei Elemente des Messvektors $\mathbf{y}(k)$ sind. Der Parameter λ gibt also an, auf wie vielen Vektoren die Berechnung der Kovarianzmatrix beruht, und bestimmt nun im Wesentlichen, wie gut die Näherung im Vergleich zur tatsächlichen Kovarianzmatrix wird. Der Aufwand für dieses numerische Verfahren setzt sich aus folgenden Teilen zusammen:

- Transformation des Messvektors in sphärische Koordinaten
- Erzeugen von 3λ Zufallszahlen
- Transformation von λ Vektoren in kartesische Koordinaten
- Kovarianzberechnung nach Gleichung (3.6)

Den größten Anteil hat daran die Kovarianzberechnung und die Transformation von sphärischen in kartesische Koordinaten mit zusammen 14λ Multiplikationen. Deshalb ist es sinnvoll, bei der Implementation in Echtzeit für λ einen an die Systemleistung angepassten Wert zu wählen und zusätzlich diese numerische Berechnung nicht für jeden Zeitpunkt kT durchzuführen, sondern nur in regelmäßigen größeren Abständen. Bei unseren Versuchen wurde die numerische Berechnung der Kovarianzmatrix für jeden Messwert durchgeführt und dabei der Parameter $\lambda = 100$ verwendet.

Insgesamt ist zu bemerken, dass das Filterergebnis sehr stark von einer richtigen Parametrisierung des Messrauschens abhängt.

3.2 Systemrauschen

In der Systemgleichung (2.1) erscheint der additive Term $\mathbf{B}u(k)$ zur Berücksichtigung des Systemrauschens. Darin ist $u(k)$ die Realisation eines weißen gaußschen Zufallsprozesses mit der Varianz σ_U^2 und \mathbf{B} die Systemrauschkoppelmatrix, die die skalare Größe $u(k)$ auf die Dimension N des Zustandsvektors $\mathbf{x}(k)$ abbildet.

Das Systemrauschen kann anschaulich damit erklärt werden, dass das gewählte Bewegungsmodell nicht perfekt eingehalten wird, sondern auch Elemente größerer Dynamik in der Bewegung enthalten sind. Diese sind zwar mittelwertfrei, weisen aber trotzdem eine gewisse Streuung auf. Für das stationäre Modell gilt folglich, dass die Änderung der Position während eines Intervalls T von der Größenordnung $T\sigma_U$ ist. Dieser Wert gilt beim linearen Modell für die Änderung der Geschwindigkeit und beim beschleunigten Modell für die Änderung der Beschleunigung während eines Intervalls T und kann damit als Richtwert für die Wahl von σ_U verwendet werden.

Für das Ergebnis des Kalman-Filters ist die Parametrisierung des Systemrauschens von geringer Bedeutung, da die unterschiedlichen Modellansätze die möglichen Dynamiken weitgehend fehlerfrei erfassen.

3.3 Anfangswerte

Das Kalman-Filter stellt eine rekursive Lösung der Filteraufgabe dar. Mit jedem Schritt wird ein Schätzwert $\hat{\mathbf{x}}(k|k)$ für den Systemzustand bestimmt und der zugehörige Fehler in der Schätzfehlerkovarianzmatrix $\mathbf{R}_{EE}(k|k)$ berechnet. Um den Filtervorgang zu starten, müssen die Anfangswerte $\hat{\mathbf{x}}(0|0)$ und $\mathbf{R}_{EE}(0|0)$ bekannt sein. Beim Tracking einer Geräuschquelle ist aber kein Vorwissen vorhanden, wo sich die Quelle befindet. Deshalb müssen die Anfangswerte aus

3 Parametrisierung des Kalman-Filters

den ersten Messwerten abgeleitet werden. Der Filtervorgang wird damit jedoch erst verzögert beginnen können.

Stationäres Modell: Mit

$$\hat{\mathbf{x}}(0|0) = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = \mathbf{y}(1) \quad (3.7)$$

wird als Anfangszustand der erste Messvektor verwendet. Dieser ist fehlerbehaftet mit dem Messfehler $\mathbf{n}(1)$, sodass für die Fehlerkovarianz des Anfangswertes

$$\mathbf{R}_{EE}(0|0) = \mathbf{R}_{NN}(1) \quad (3.8)$$

gilt.

Lineares Modell: In diesem Fall ist neben der Anfangsposition auch die Anfangsgeschwindigkeit zu bestimmen. Durch einfache Differenzbildung wählt man

$$\hat{\mathbf{x}}(0|0) = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ \dot{x}_0 \\ \dot{y}_0 \\ \dot{z}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{y}(2) \\ \mathbf{y}(2) - \mathbf{y}(1) \end{pmatrix}. \quad (3.9)$$

Da sich das Messrauschen während zweier Messwerte nicht wesentlich ändert, gilt

$$\mathbf{R}_{NN}(2) \approx \mathbf{R}_{NN}(1)$$

und somit

$$\mathbf{R}_{EE}(0|0) = \begin{pmatrix} \mathbf{R}_{NN}(1) & \mathbf{R}_{NN}(1) \\ \mathbf{R}_{NN}(1) & 2\mathbf{R}_{NN}(1) \end{pmatrix}. \quad (3.10)$$

Der Filtervorgang beginnt in diesem Fall erst mit dem dritten Messwert.

Beschleunigtes Modell: Ganz entsprechend setzt man bei diesem Bewegungsmodell

$$\hat{\mathbf{x}}(0|0) = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ \dot{x}_0 \\ \dot{y}_0 \\ \dot{z}_0 \\ \ddot{x}_0 \\ \ddot{y}_0 \\ \ddot{z}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{y}(3) \\ \mathbf{y}(3) - \mathbf{y}(2) \\ [\mathbf{y}(3) - \mathbf{y}(2)] - [\mathbf{y}(2) - \mathbf{y}(1)] \end{pmatrix} = \begin{pmatrix} \mathbf{y}(3) \\ \mathbf{y}(3) - \mathbf{y}(2) \\ \mathbf{y}(3) - 2\mathbf{y}(2) + \mathbf{y}(1) \end{pmatrix} \quad (3.11)$$

und

$$\mathbf{R}_{EE}(0|0) = \begin{pmatrix} \mathbf{R}_{NN}(1) & \mathbf{R}_{NN}(1) & \mathbf{R}_{NN}(1) \\ \mathbf{R}_{NN}(1) & 2\mathbf{R}_{NN}(1) & 2\mathbf{R}_{NN}(1) \\ \mathbf{R}_{NN}(1) & 2\mathbf{R}_{NN}(1) & 4\mathbf{R}_{NN}(1) \end{pmatrix}. \quad (3.12)$$

Man kann die Anfangswerte für die Fehlerkovarianz auch kleiner wählen. Dann ändert sich die Charakteristik des Filterergebnisses dahingehend, dass während des Einschwingvorgangs der neue Messwert im Gegensatz zum Anfangszustand ein kleineres Gewicht in der Positionsschätzung erhält. Dies ist sinnvoll, wenn für die ersten Messwerte eine geringere Fehlerbehaftung zu erwarten ist, als für die folgenden Messwerte. Nach dem *Gesetz der 1. Wellenfront* kann man davon in der Tat ausgehen. Darunter versteht man den Effekt, dass die erste am Mikrofon ankommende Wellenfront den direkten Ausbreitungspfad durchlaufen hat. Das liegt daran, dass bei der ersten Wellenfront keine Mehrwegeüberlagerung durch Reflexion an den Wänden stattfindet. Diese Gesetzmäßigkeit wird auch *precedence effect* genannt [HS97].

Wenn die Geräuschquelle nur kurzzeitig aktiv ist, ist eine richtige Wahl der Anfangswerte wichtiger, als wenn der Trackingvorgang über eine lange Zeit dauert, da diese sich nur auf das Einschwingverhalten des Filters auswirken.

3 Parametrisierung des Kalman-Filters

4 Beurteilung des Filters

In diesem Abschnitt sollen drei Kriterien zur Beurteilung des Kalman-Filters vorgestellt werden. Zunächst soll der Filtergewinn anhand des Fehlers betrachtet werden. Danach wird eine Methode vorgestellt, mit der eine Aussage über die Wahl der Parameter getroffen werden kann. Als letztes wird der Begriff der Modellwahrscheinlichkeit eingeführt, um die Modellbildung zu beurteilen.

4.1 Fehlerbetrachtung

Es ist wünschenswert, die Verbesserung durch den Filtervorgang in einer greifbaren Form darzustellen. Dazu kann der Vergleich von Messfehler $\mathbf{n}(k)$ und Schätzfehler $\mathbf{e}(k) = \hat{\mathbf{x}}(k|k) - \mathbf{x}(k)$ herangezogen werden. Insbesondere der stationäre Fall erlaubt eine direkte Gegenüberstellung, da hier das Messrauschen zeitlich konstant ist. Bedingt durch die konstante Dimension des Messfehlers kann ein Vergleich bei allen Modellen nur für den Positionsfehler durchgeführt werden und *nicht* bei den Systemmodellen höherer Ordnung auch für Geschwindigkeit oder Beschleunigung.

Der Messfehler berechnet sich zu

$$\mathbf{n}(k) = \mathbf{y}(k) - \mathbf{y}(k)_{ideal} = \mathbf{y}(k) - \mathbf{C}\mathbf{x}(k) \quad (4.1)$$

und erfordert wie der Schätzfehler $\mathbf{e}(k)$ das Wissen über den tatsächlichen Systemzustand $\mathbf{x}(k)$. Deshalb können diese Größen nur für fest vorbestimmte Trajektorien der bewegten Geräuschquelle ermittelt werden. Abbildung 4.1 zeigt an einem Beispiel den Verlauf des Mess- und Schätzfehlers für eine lineare Bewegung. Dazu wurde ein gehender Sprecher lokalisiert und das Messergebnis mit zwei Kalman-Filtern verarbeitet, denen eine stationäre bzw. eine lineare Modellannahme zu Grunde lag. Dabei ist es natürlich nicht möglich, beim Gehen eine exakt konstante Geschwindigkeit einzuhalten. Trotzdem wurde an dieser Stelle einer suboptimalen Trajektorie gegenüber simulierten Daten der Vorzug gegeben.

Während die Varianz des Messfehlers zeitlich konstant bleibt, ändert sich die des Schätzfehlers schnell. Der Schätzfehler für den stationären Ansatz steigt auf Grund der falschen Modellannahme auf große Werte an, während er beim linearen Bewegungsmodell unter den Messfehler sinkt. Vor allem in der Umgebung des stark fehlerbehafteten Messwertes bei $t = 4.8$ s ist zu erkennen, dass der Schätzer eine deutlich geringere Varianz aufweist als die Messwerte.

4 Beurteilung des Filters

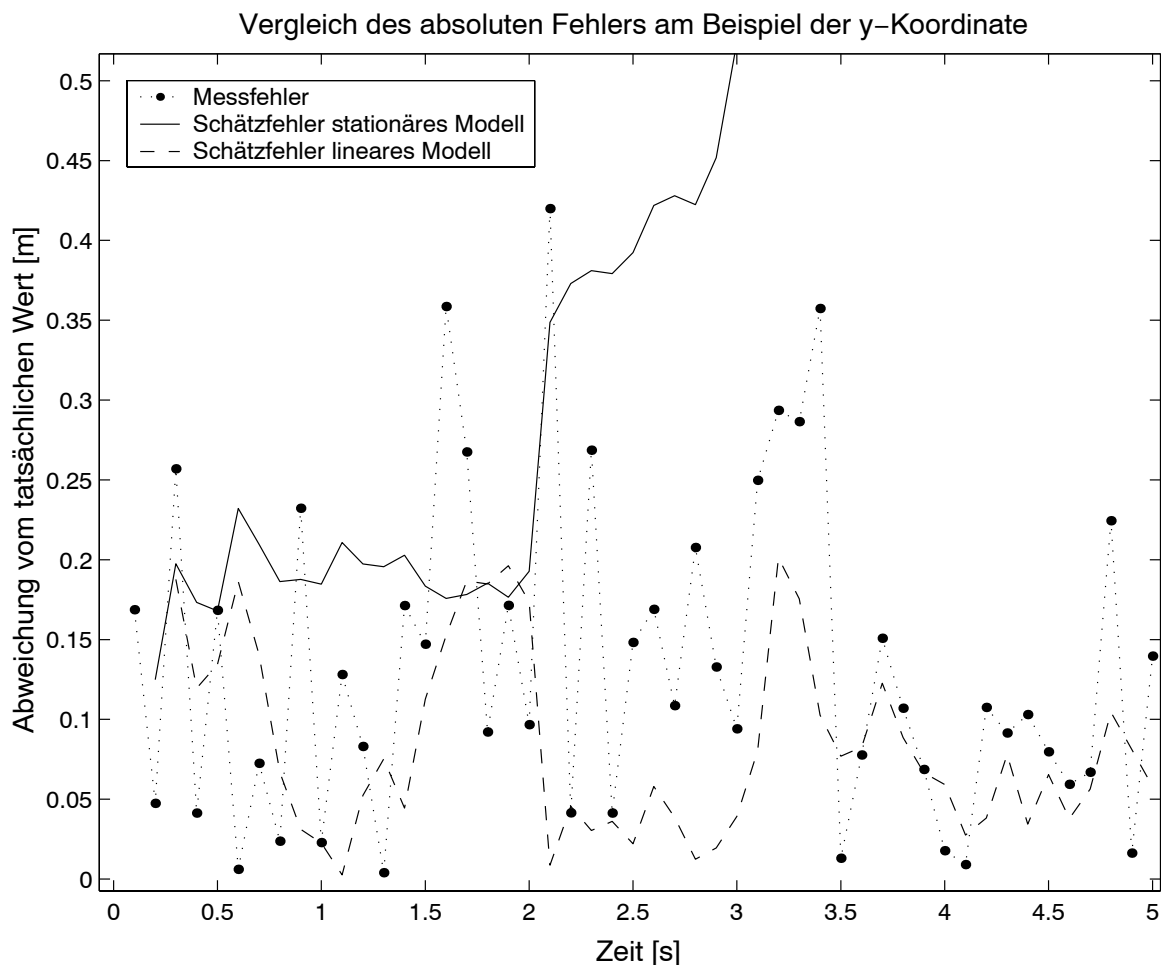


Abbildung 4.1: Fehlervergleich

4.2 Konsistenz

Die Parameter des Kalman-Filters haben einen wesentlichen Einfluss auf die Qualität der Schätzung des Systemzustandes. Selbst wenn die Modellannahme die Bewegung der Geräuschquelle hinreichend genau beschreibt, gibt es noch die in Kapitel 3 diskutierten Parameter $\mathbf{R}_{NN}(k)$, σ_U^2 , $\hat{\mathbf{x}}(0|0)$ und $\mathbf{R}_{EE}(0|0)$, die das Filter als richtig voraussetzt. Falls diese Parameter nicht mit den realen Werten des betrachteten Systems übereinstimmen, so kann dies zu großen Ungenauigkeiten bei der Schätzung führen. Im Folgenden wird nun ein Verfahren vorgestellt, mit welchem es möglich ist, die Parameterwahl zu prüfen: der *Konsistenztest*.

Dazu betrachtet man die Innovation, die als Differenz zwischen dem aktuellen Messwert und dem *a priori* geschätzten Messwert definiert ist:

$$\nu(k) = \mathbf{y}(k) - \hat{\mathbf{y}}(k|k-1) = \mathbf{y}(k) - \mathbf{C}\hat{\mathbf{x}}(k|k-1). \quad (4.2)$$

Falls das Systemmodell exakt stimmt und das Kalman-Filter richtig parametrisiert ist, sollte nach Einschwingen des Filters die Innovation gegen Null konvergieren.

Für den Mittelwert und die Varianz der Innovation berechnet man nach [Spo99]

$$E \{ \nu(k) \} = \mathbf{0} \quad (4.3)$$

$$\mathbf{R}_{\nu\nu}(k) = E \{ \nu(k)\nu^T(k) \} = \mathbf{R}_{NN}(k) + \mathbf{C}\mathbf{R}_{EE}(k|k-1)\mathbf{C}^T. \quad (4.4)$$

Im Idealfall folgt dann der Skalar $\nu^T(k)\mathbf{R}_{\nu\nu}^{-1}(k)\nu(k)$ einer χ^2 -Verteilung mit $M = 3$ Freiheitsgraden [BS88]. Durch Vergleich von Mittelwert und Varianz der Größe $\nu^T(k)\mathbf{R}_{\nu\nu}^{-1}(k)\nu(k)$ mit den bekannten Werten für eine χ_M^2 -verteilte Zufallsvariable [LW85] lässt sich leicht eine Aussage machen, ob $\nu^T(k)\mathbf{R}_{\nu\nu}^{-1}(k)\nu(k)$ einer χ^2 -Verteilung folgt und damit die Parameter des Kalman-Filters richtig gewählt wurden.

In der Praxis wird man diesen Konsistenztest allerdings nicht aus einem einzelnen Vektor $\nu(k)$ aufbauen. Stattdessen generiert man die *Fading-memory Likelihood Funktion* $\rho(k)$, in die neben der Innovation zum aktuellen Zeitpunkt kT auch vergangene Werte eingehen, die mit einer Konstanten $\mu < 1$ ausgeblendet werden:

$$\rho(k) = \mu\rho(k-1) + \nu^T(k)\mathbf{R}_{\nu\nu}^{-1}(k)\nu(k). \quad (4.5)$$

Die effektive Fensterlänge von ρ umfasst $(1 - \mu)^{-1}$ Innovationsvektoren, und ρ folgt näherungsweise einer χ^2 -verteilten Zufallsvariable mit $M(1 + \mu)/(1 - \mu)$ Freiheitsgraden, dem Mittelwert $M/(1 - \mu)$ und der Varianz $2M/(1 - \mu^2)$.

Ein sinnvoller Wert ist $\mu = \frac{2}{3}$. Damit erhält man für die effektive Fensterlänge den Wert von $(1 - \mu)^{-1} = 3$, und ρ folgt bei richtiger Parametrisierung der Statistik einer χ^2 -verteilten Zufallsvariable mit $M(1 + \mu)/(1 - \mu) = 15$ Freiheitsgraden. Es lässt sich damit das 95%-Konfidenzintervall [6.2621, 27.4884] angeben. Falls der Wert von $\rho(k)$ in diesem Intervall liegt, folgt $\rho(k)$ hinreichend gut einer χ_{15}^2 -verteilten Zufallsvariablen und die Parameter sind gut gewählt. Tabelle 4.1 enthält eine Übersicht über den Zusammenhang der Parameter für denkbare Fensterlängen. Die Quantile der χ^2 -Verteilung entnimmt man Anhang C.

Effektive Fensterlänge	Ausblendkonstante μ	Anzahl der Freiheitsgrade	95%-Konfidenzintervall	
			Untergrenze	Obergrenze
1	0	3	0.2158	9.3484
2	0.5000	9	2.7004	19.0228
3	0.6667	15	6.2621	27.4884
4	0.7500	21	10.2829	35.4789
5	0.8000	27	14.5734	43.1945
10	0.9000	57	38.0267	79.7522
15	0.9333	87	63.0894	114.6929
20	0.9500	117	88.9551	148.8288

Tabelle 4.1: Parameter für Konsistenztests

4 Beurteilung des Filters

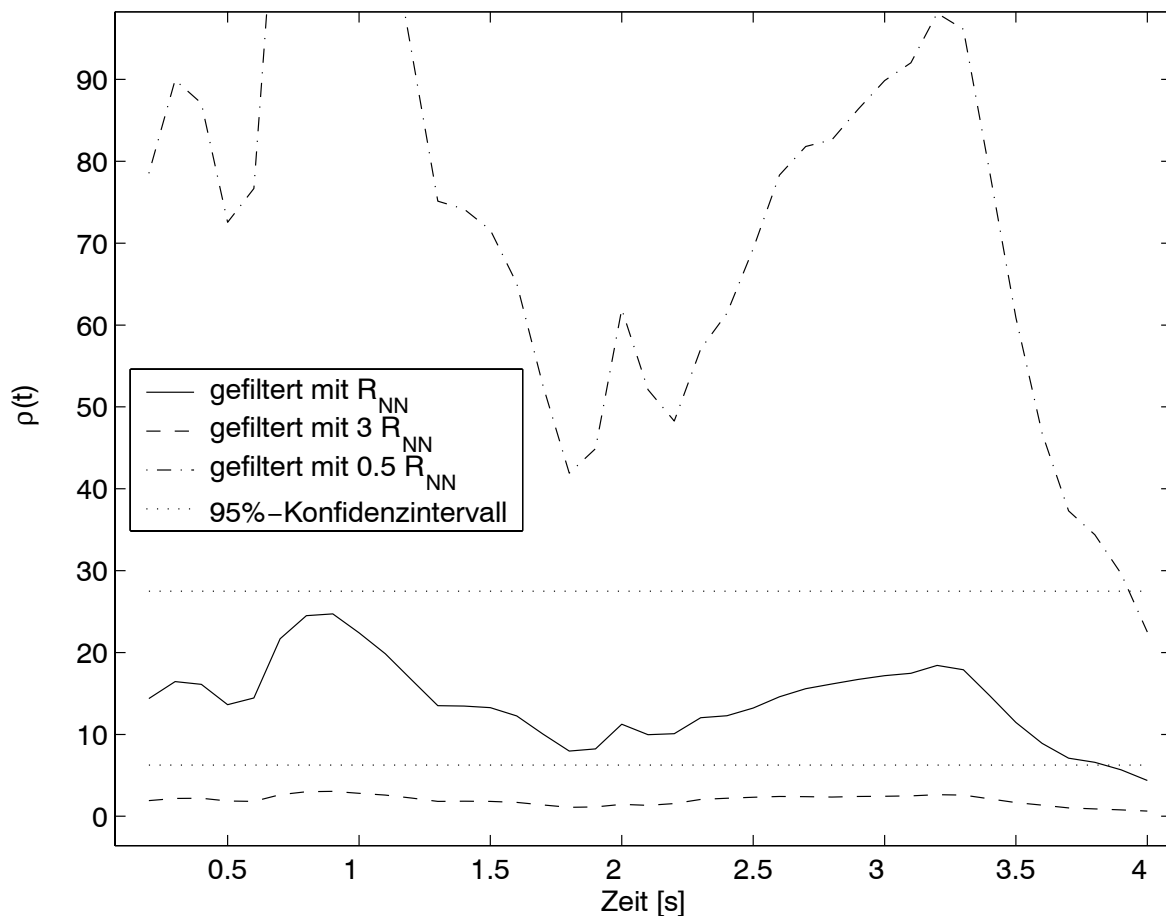


Abbildung 4.2: Konsistenztest mit unterschiedlichem Messrauschen

Abbildung 4.2 zeigt den Vergleich von $\rho(t = kT)$, wobei T das Zeitintervall zwischen zwei Messwerten bezeichnet, bei verschiedenen Annahmen für die Messfehlerkovarianz R_{NN} . Das richtig parametrisierte Filter befindet sich innerhalb der 95%-Grenzen, während die Kurven für die fehlparametrisierten Filter darunter oder darüber liegen. Falls das Messrauschen beim Filterentwurf zu klein angenommen wird ($0.5 \cdot R_{NN}$), so zeigt sich dies in der Funktion ρ daran, dass die Kurve überhalb des 95%-Konfidenzintervalls liegt. Wurde das Messrauschen wie in der Kurve für $3 \cdot R_{NN}$ zu groß angesetzt, so findet man diese unterhalb des 95%-Konfidenzintervalls wieder.

Insgesamt ist durch diesen Konsistenztest zwar eine Aussage über die Parameterwahl zu machen, aber es ist nicht direkt ableitbar, auf welche Werte sie verändert werden müssen, um ein Optimalergebnis zu erhalten.

4.3 Modellwahrscheinlichkeit

Bisher wurden die drei Modellansätze beim Kalman-Filter unabhängig von einander betrachtet. Um nun zu einer Aussage über die Güte der Modellbildung zu gelangen, soll hier der Begriff der *Modellwahrscheinlichkeit* eingeführt werden. Dazu müssen zunächst die Modelle α_i mit $1 \leq i \leq L$ nummeriert werden. Dann ist die bedingte Wahrscheinlichkeit $f(\alpha_i|\mathbf{y}(k))$ die Wahrscheinlichkeit dafür, dass das i -te Modell auf der Grundlage der Messwerte $\mathbf{y}(1), \dots, \mathbf{y}(k)$ das richtige ist. Wendet man darauf die Bayes-Regel an, so ergibt sich

$$f(\alpha_i|\mathbf{y}(k)) = \frac{f(\mathbf{y}(k)|\alpha_i) f(\alpha_i)}{f(\mathbf{y}(k))}. \quad (4.6)$$

Für die praktische Implementierung wird der Nenner durch die totale Wahrscheinlichkeit

$$f(\mathbf{y}(k)) = \sum_{j=1}^L f(\mathbf{y}(k)|\alpha_j) f(\alpha_j) \quad (4.7)$$

ersetzt. Berücksichtigt man, dass

$$\begin{aligned} f(\mathbf{y}(k)|\alpha_i) &= f(\mathbf{y}(k), \mathbf{y}(k-1)|\alpha_i) \\ &= f(\mathbf{y}(k)|\mathbf{y}(k-1), \alpha_i) f(\mathbf{y}(k-1)|\alpha_i) \end{aligned} \quad (4.8)$$

$$= f(\mathbf{y}(k)|\mathbf{y}(k-1), \alpha_i) \frac{f(\alpha_i|\mathbf{y}(k-1)) f(\mathbf{y}(k-1))}{f(\alpha_i)} \quad (4.9)$$

gilt, so lässt sich eine rekursive Formel für Gleichung (4.6) angeben:

$$f(\alpha_i|\mathbf{y}(k)) = \frac{f(\mathbf{y}(k)|\mathbf{y}(k-1), \alpha_i) f(\alpha_i|\mathbf{y}(k-1))}{\sum_{j=1}^L f(\mathbf{y}(k)|\mathbf{y}(k-1), \alpha_j) f(\alpha_j|\mathbf{y}(k-1))}. \quad (4.10)$$

Die Terme $f(\mathbf{y}(k)|\mathbf{y}(k-1), \alpha)$ können mit jedem Rekursionsschritt des Kalman-Filters mit (4.4) nach der Gleichung

$$f(\mathbf{y}(k)|\mathbf{y}(k-1), \alpha_i) = \frac{1}{(2\pi)^{\frac{3}{2}} \sqrt{|\mathbf{R}_{\nu\nu,i}(k)|}} \exp \left\{ -\frac{1}{2} \nu_i^T(k) \mathbf{R}_{\nu\nu,i}^{-1}(k) \nu_i(k) \right\} \quad (4.11)$$

berechnet werden [Spo99]. Als Anfangswerte wählt man, sofern man kein Vorwissen besitzt,

$$f(\alpha_i|\mathbf{y}(0)) = \frac{1}{L}, \quad (4.12)$$

also für jedes Modell die gleiche Wahrscheinlichkeit. Die Modellwahrscheinlichkeit für das richtige Modell wird mit fortschreitendem Zeitindex k gegen Eins konvergieren, während die Modellwahrscheinlichkeiten der anderen Modelle sich dem Wert Null nähern. Ist bei der numerischen Berechnung allerdings einmal die Null erreicht, so ändert sich auf Grund der Berechnungsvorschrift (4.10) daran auch nichts mehr. Um diesen *dead lock* zu vermeiden, kann man für

4 Beurteilung des Filters

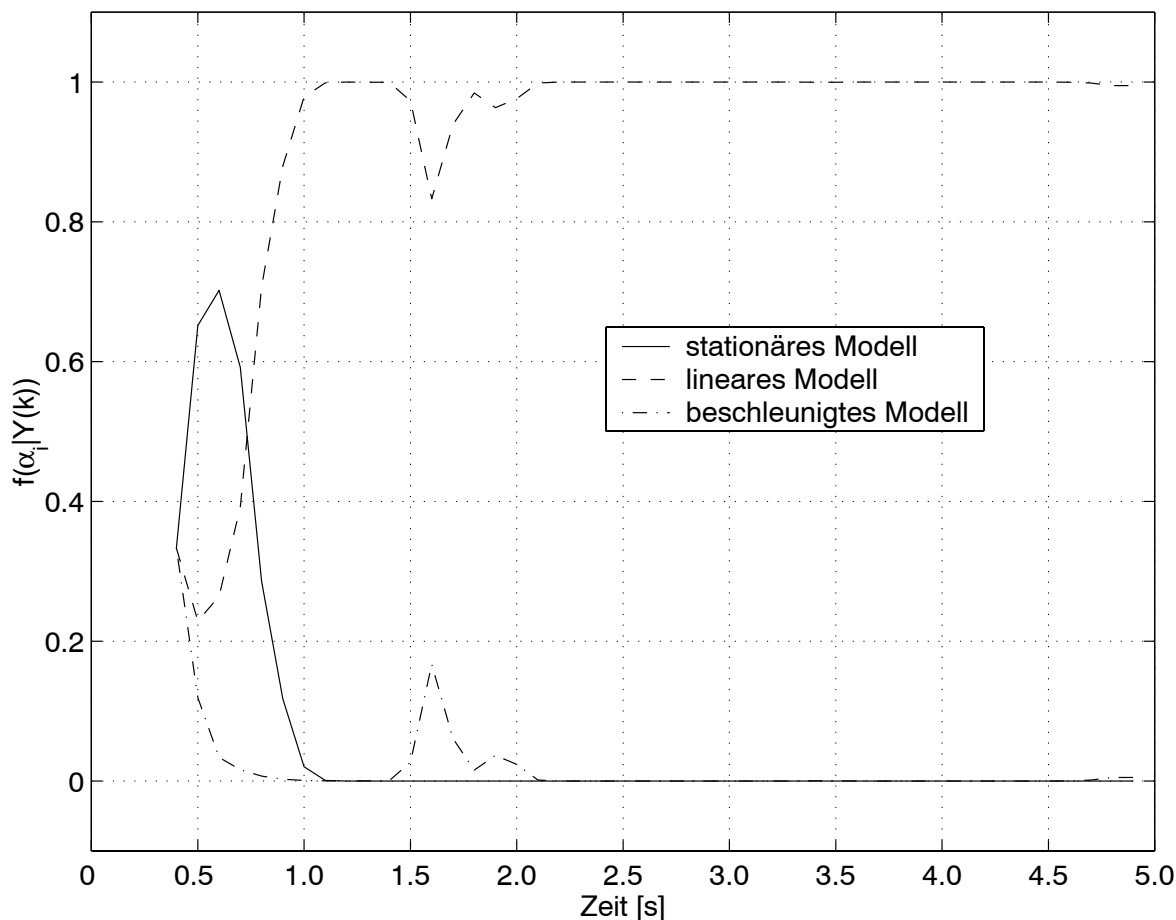


Abbildung 4.3: Modellwahrscheinlichkeiten

jedes Modell eine untere Schranke ϵ angeben, unter dessen Wert die Modellwahrscheinlichkeit nicht sinken darf:

$$\tilde{f}(\alpha_i | \mathbf{y}(k)) = \frac{\max\{\epsilon, f(\alpha_i | \mathbf{y}(k))\}}{\sum_{j=1}^L \max\{\epsilon, f(\alpha_j | \mathbf{y}(k))\}} \quad (4.13)$$

Abbildung 4.3 zeigt beispielhaft den Verlauf der Modellwahrscheinlichkeiten für eine lineare Bewegung. Es werden $L = 3$ Modelle verwendet. Ausgehend von der Wahrscheinlichkeit $\frac{1}{3}$ steigt der Wert für das lineare Modell schnell an, während die Werte für die beiden nichtzutreffenden Systemmodelle auf $\epsilon = 10^{-4}$ sinken.

Es ist zu bemerken, dass für die Berechnung der Modellwahrscheinlichkeiten stets alle L Kalman-Filter parallel arbeiten müssen. Dies bedeutet einen höheren Zeitaufwand mit jedem Rekursionschritt. Deshalb ist es nicht sinnvoll, L sehr groß zu wählen. Außerdem erlaubt die Berechnung der Modellwahrscheinlichkeit nur den Vergleich mehrerer fest vorgegebener Modelle. Falls das *richtige* Modell nicht unter den L Modellen ist, wird trotzdem eine der Modellwahrscheinlichkeiten gegen Eins konvergieren.

4.3 Modellwahrscheinlichkeit

In Kapitel 5.4 wird dieses Wissen über die Güte der Modellbildung verwendet, um ein adaptives Kalman-Filter aufzubauen.

4 Beurteilung des Filters

5 Adaption an reale Szenarien

Die bisherigen Betrachtungen reichen für das Konzept eines Kalman-Filters unter simulativen Bedingungen aus. Bei unserer Anwendung, in Küchen- oder Büroumgebung eine bewegte Geräuschquelle robust tracken zu können, müssen jedoch weitere Faktoren berücksichtigt werden. So kann vor allem nicht davon ausgegangen werden, dass die Geräuschquelle während des gesamten Trackingvorganges aktiv bleibt. Außerdem wird gerade im Falle eines Sprechers die Dynamik in der Bewegung variabel sein, sodass die Entscheidung für ein Bewegungsmodell bei dem Kalman-Filter kontinuierlich geregelt werden muss. Dies soll in diesem Abschnitt näher betrachtet werden.

5.1 Testumgebung

Die Grundlage für den Praxisbezug bilden 43 aufgenommene Szenarien, in denen jeweils eine oder zwei Personen Testsätze sprechen. In einigen Fällen wird auch konstant ein energiereicher Vokal artikuliert, um ununterbrochen Messwerte zur Verfügung zu haben. Die genaue Auflistung befindet sich in Anhang B. Dabei werden neben stationären Positionen im Sitzen und Stehen (26%) auch Bewegungen nach vorgegebenen Trajektorien mit nahezu konstanter Geschwindigkeit untersucht. Da dies der interessanteste Fall für das passive Audio-Tracken ist, wurden dafür die meisten Aufnahmen gemacht (60%). Für den Test des beschleunigten Modells werden Übergänge von der ruhenden Position in schnelles Gehen betrachtet (9%). Außerdem wurden Szenarien aufgenommen, in denen ein Sprecher ruht und ein anderer sich linear bewegt. Dabei zeigte sich jedoch kein anderes Verhalten als bei der getrennten Betrachtung der beiden Fälle. Die Testumgebung bildet ein Büroraum der Größe 4m x 5m x 3m mit gewöhnlicher Büroaustattung. Wegen glatten Wänden sind die Audiosignale relativ stark mit Echo behaftet. Als Störquelle ist nur das niederfrequente Umgebungsrauschen, z.B. durch Lüfter, anzusehen.

Die Audiosignale wurden mit Hilfe von fünf Mikrofonen in gleichseitiger Doppel-Tetraederanordnung aufgezeichnet. Diese Randbedingung ist durch das Einsatzziel eines humanoiden Roboters festgelegt. Der Abstand zwischen den einzelnen Mikrofonen beträgt jeweils 28 cm. Die zeitlichen Abstände der einzelnen Messwerte, die dem Kalman-Filter zur Verfügung stehen, erhält man für unseren Algorithmus durch folgende Betrachtung. Jeweils 512 Abtastwerte des mit 16 kHz abgetasteten Audiosignals werden zu Frames der Länge 32 ms zusammengefasst bei einer Überlappungslänge von 16 ms. Für den Lokalisationsalgorithmus wird aus jeweils 12 Frames ein Block gebildet, wobei auch die einzelnen Blöcke sich um 50% überlappen (siehe Abbildung

5 Adaption an reale Szenarien

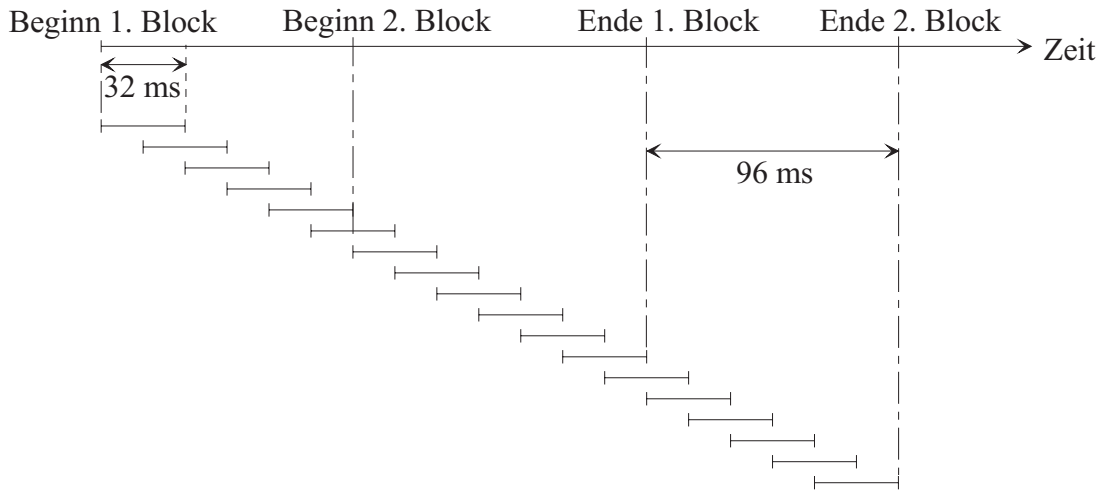


Abbildung 5.1: Zeitlicher Abstand zweier Messwerte

5.1). Da aus jedem Block ein Messwert für die Position der Geräuschquelle folgt, ist damit der zeitliche Abstand zwischen zwei Messwerten 96 ms. Für das robuste Tracken von Menschen werden mindestens vier Messwerte pro Sekunde benötigt [SK92]; damit ist sichergestellt, dass bei uns das Tracken in jedem Fall möglich ist. Die Aufnahmen besitzen eine Länge von jeweils drei bis acht Sekunden, womit pro Szenario dem Kalman-Filter zwischen 32 und 87 Messwerte zur Verfügung stehen.

5.2 Reine Prädiktion

In realen Szenarien wird es vorkommen, dass vereinzelt der neue Messwert $y(k+1)$ nicht verfügbar ist. Dies ist beispielsweise bei Sprechpausen zwischen einzelnen Sätzen der Fall oder wenn die TDEs in einzelnen Mikrofonpaaren vom Algorithmus als nicht zuverlässig genug eingestuft werden. Dann ist die Berechnung des neuen Schätzwerts nach Gleichung (2.19) nicht möglich. Stattdessen verwendet man den *A priori* Schätzwert $\hat{x}(k+1|k)$ selbst. Dies hat zur Folge, dass im stationären Fall für die Dauer der reinen Prädiktion einfach der letzte Schätzwert beibehalten wird. Beim linearen Modell folgen die rein prädizierten Schätzwerte der durch die konstante Geschwindigkeit bestimmten Geraden bzw. beim beschleunigten Modell der durch die konstante Beschleunigung gegebenen Parabel. Die Modellwahl und die Historie aus den bisher gemachten Schätzungen bestimmen also entscheidend den Verlauf, solange keine neuen Messwert zur Verfügung stehen. Dies wird auch aus Abbildung 5.2 ersichtlich, die beispielhaft am Schätzwert der z-Koordinate zeigt, wie sich reine Prädiktion auswirkt. Für dieses Bild liegt ein Bewegungsvorgang zugrunde, bei dem sich der zu trackende Sprecher aus dem aufrechten Stand auf einen Stuhl setzt.

Die Schätzfehlerkovarianz wird bei der reinen Prädiktion natürlich ansteigen, da keine neue Information in das Filter eingeht. Man kann den Fall *fehlender Messwert* auch so interpretieren,

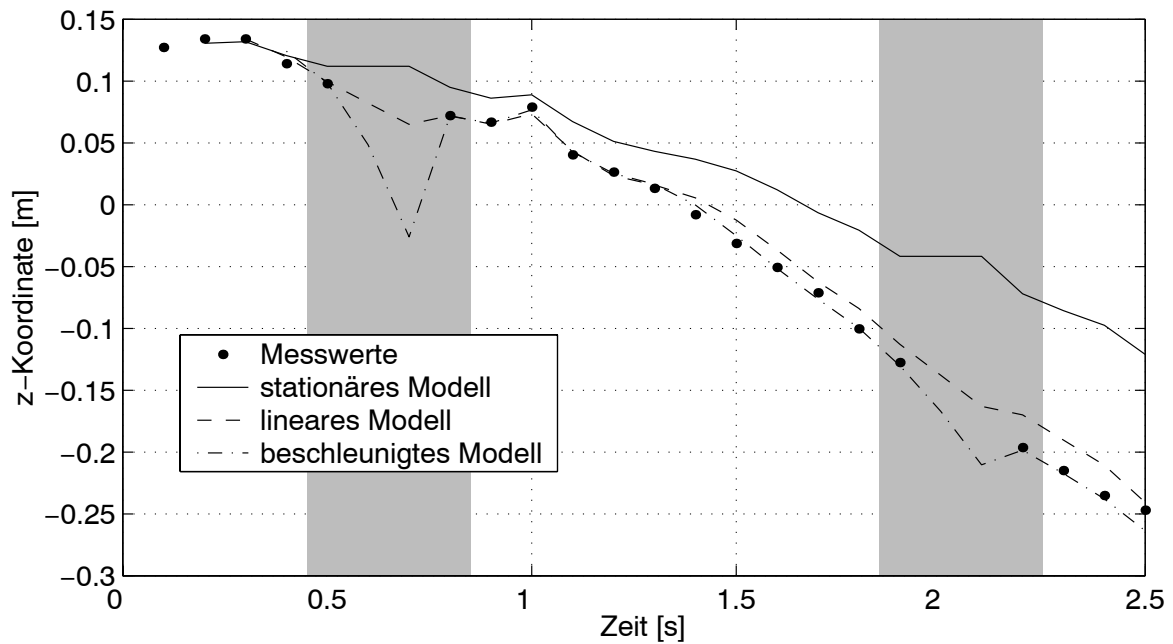


Abbildung 5.2: Verlauf der Schätzung bei Prädiktion

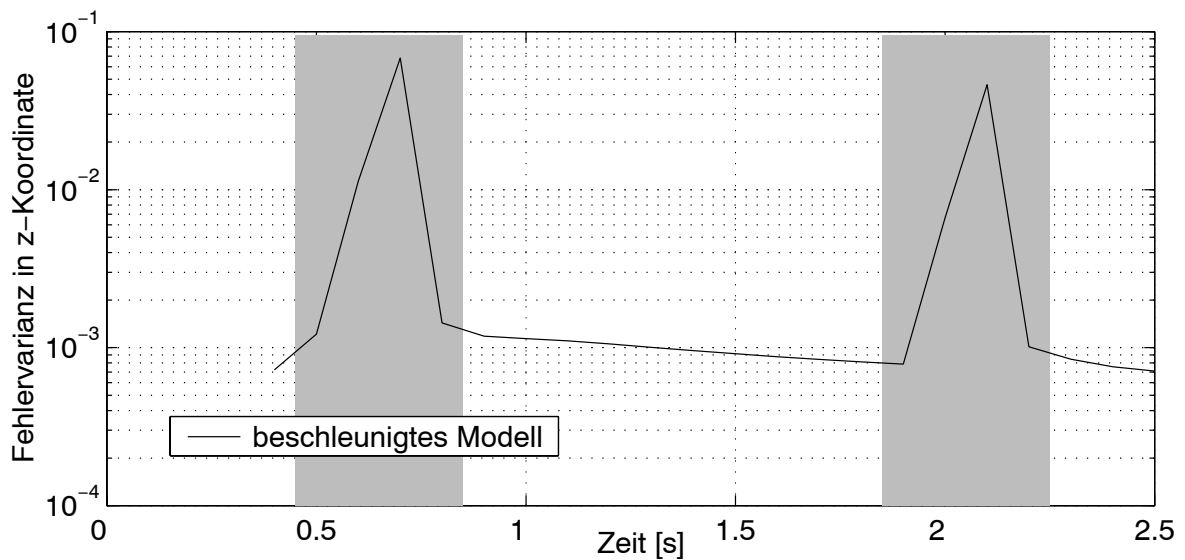


Abbildung 5.3: Erhöhte Fehlervarianz bei Prädiktion

dass die Elemente der Messfehlerkovarianzmatrix \mathbf{R}_{NN} gegen Unendlich gehen und damit der Filter Gain aus Gleichung (2.12) gegen Null strebt. Damit ergibt sich die Schätzfehlerkovarianz $\mathbf{R}_{EE}(k+1|k)$ aus Gleichung (2.13), welche im Vergleich zur Schätzfehlerkovarianz *mit* neuem

5 Adaption an reale Szenarien

Messwert $\mathbf{R}_{EE}(k+1|k+1)$ aus (2.16) um die quadratische Form $\mathbf{W}(k+1)\mathbf{R}_{\nu\nu}(k+1)\mathbf{W}^T(k+1)$ größer ist.

Für das gleiche Beispiel wie in Abbildung 5.2 zeigt nun Abbildung 5.3 das Verhalten einer Komponente in der Hauptdiagonalen der Fehlerkovarianzmatrix. Man erkennt die erhöhte Fehlervarianz zu den Zeiten, bei denen keine Messwerte für die Schätzung herangezogen werden können. Es ist auffällig, dass zu Beginn der Filterung die reine Prädiktion einen größeren Fehler mit sich zieht als im späteren Verlauf. Das ist damit zu erklären, dass auf Grund der kürzeren Historie die Unsicherheit in der Zustandsschätzung, die das Verhalten für die Prädiktion bestimmt, größer ist. Während die Prädiktion bei einem fehlenden Messwert stets unproblematisch ist, steigt der Fehler deutlich an, wenn über längere Zeit prädiziert wird. Die Fähigkeit, zumindest für eine bestimmte Zeit auch ohne Messwert eine sinnvolle Schätzung abzugeben, ist ein entscheidender Vorteil des Kalman-Filters gegenüber den reinen Messwerten. Liegen über eine längere Zeitdauer keine Messwerte vor, so sollte das Tracking allerdings besser abgebrochen werden.

5.3 Trackingpausen

Sind über längere Zeit keine Messwerte verfügbar, so steigt der Fehler bei reiner Prädiktion deutlich an. Außerdem wird mit einem einzelnen Sprecher nur ein Sonderfall der akustischen Szene berücksichtigt. Im allgemeinen Fall können mehrere Sprecher neben anderen Umgebungsgeräuschen, in der Küche zum Beispiel Klingel- oder Wassergeräusche, vorkommen. Deshalb muss das Kalman-Filter in der Lage sein, nach einer fest vorzugebenden Zeit τ das Verfolgen einer Geräuschquelle abubrechen (*track drop*), um es zu einem späteren Zeitpunkt nach einer Reinitialisierung (*track init*) wieder aufzunehmen (*track continue*).

Falls eine Geräuschquelle inaktiv wird, sollte das Tracken möglichst schnell beendet werden. Wird die Positionsbestimmung einer aktiven Quelle nicht möglich, weil kurzzeitig keine Messwerte vorliegen, so sollte bei guter Modellwahl und eingeschwungenem Filterzustand ein *track drop* möglichst lange unterbunden werden. Insofern ist für den Schwellwert τ ein Kompromiss einzugehen. Eine sinnvolle Grenze liegt bei den 43 aufgenommenen realen Szenarien bei drei Messwerten. Dies entspricht einem Zeitintervall von $\tau = 0.288$ s.

Abbildung 5.4 zeigt beispielhaft das Verhalten der Positionsbestimmung am Winkel Azimut bei *track drop* und *track continue* für ein zu klein gewähltes $\tau = 0.192$ s. Zum Zeitpunkt $t = 1.2$ s wird die Geräuschquelle dauerhaft inaktiv und es wurde nach zwei Prädiktionswerten das Tracking sinnvollerweise abgebrochen. In dem Intervall $[3 \text{ s}, 3.5 \text{ s}]$ fehlen allerdings nur zwei Messwerte, und auf Grund der kleinen Schwelle τ wurde das Tracking auch hier abgebrochen, obwohl die Messwerte nur kurzzeitig nicht verfügbar waren, z.B. durch Echoeinflüsse. Das Tracking wird in beiden Fällen erst wieder aufgenommen, wenn genügend Messwerte für eine Reinitialisierung vorhanden sind.

Der Nachteil beim Unterbrechen eines Trackingvorganges besteht vor allem darin, dass die Information aus der Historie beim Abbruch verloren geht. Das Filter muss reinitialisiert werden, womit wieder ein neues Einschwingen verbunden ist.

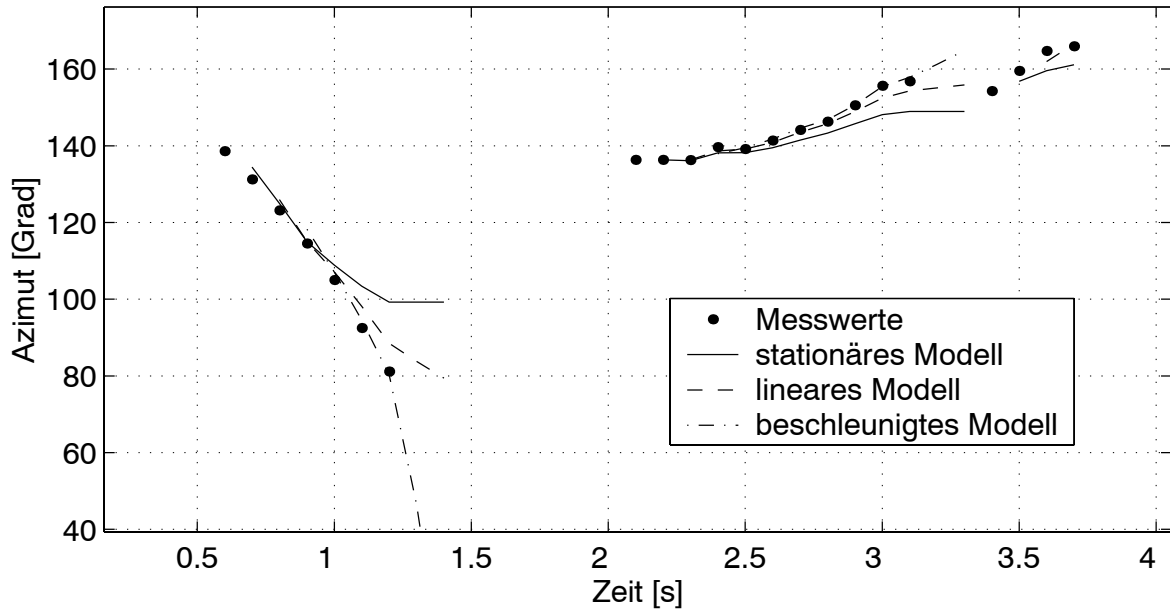


Abbildung 5.4: Verlauf der Schätzung bei Trackingpausen

5.4 Adaptives Kalman-Filter

Bei den bisherigen Überlegungen wurden die verschiedenen Modellansätze des Kalman-Filters unabhängig von einander betrachtet. Nun soll mit Hilfe der Definition der Modellwahrscheinlichkeit aus Kapitel 4.3 ein Ansatz untersucht werden, bei dem mehrere Modellbildungen gleichzeitig für ein Schätzergebnis herangezogen werden. Man nennt einen solchen Ansatz *multiple model approach* [BS88] und die zugehörige Filtereinrichtung ein *adaptive Kalman-Filter* oder *multiple model adaptive estimator* (MMAE).

Abbildung 5.5 zeigt das zugehörige Signalblockschaltbild: Der Messwert $\mathbf{y}(k)$ wird in L Kalman-Filtern parallel zu einem Schätzwert für den Systemzustand $\hat{\mathbf{x}}_1(k|k), \dots, \hat{\mathbf{x}}_L(k|k)$ verwendet. Diese Schätzwerte haben entsprechend der jeweiligen Modellbildung der einzelnen Kalman-Filter unterschiedliche Dimensionen. Durch die Beobachtungsmatrizen $\mathbf{C}_1, \dots, \mathbf{C}_L$ erfolgt dann die Abbildung auf die Positionsschätzungen $\hat{\mathbf{y}}_1(k|k), \dots, \hat{\mathbf{y}}_L(k|k)$ gleicher Dimension. An dieser Stelle werden nun die Schätzergebnisse mit den Modellwahrscheinlichkeiten $f(\alpha_i|\mathbf{y}(k))$, mit $1 \leq i \leq L$, gewichtet und summiert. Somit ergibt sich für den Schätzwert des adaptiven Kalman-Filters die Berechnungsvorschrift

$$\hat{\mathbf{y}}_{MMAE}(k|k) = \sum_{i=1}^L f(\alpha_i|\mathbf{y}(k)) \mathbf{C}_i \hat{\mathbf{x}}_i(k|k). \quad (5.1)$$

Setzt man das adaptive Kalman-Filter mit den drei bisher betrachteten Modellen für stationäre, lineare und beschleunigte Bewegung an, so sind die Modellwahrscheinlichkeiten zu Beginn gleich groß. Die der richtigen Modellbildung steigt dann aber schnell auf Werte nahe der Eins an. Damit

5 Adaption an reale Szenarien

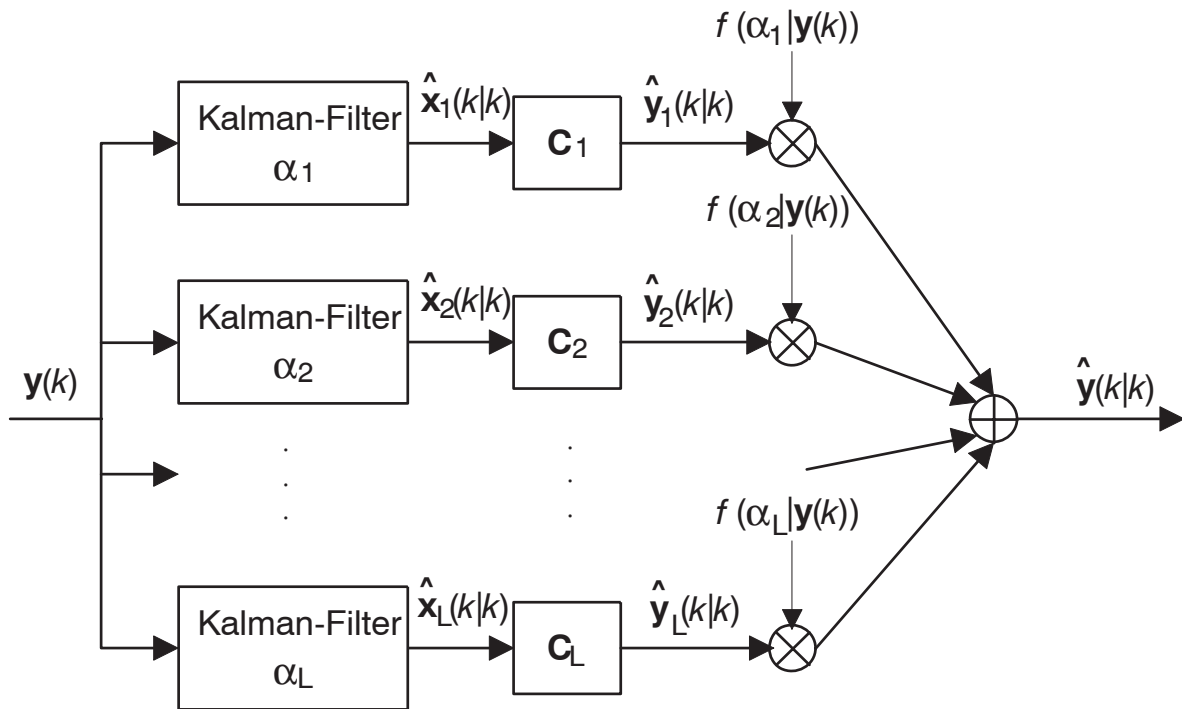


Abbildung 5.5: Struktur des adaptiven Kalman-Filters

besteht auch das MMAE-Schätzergebnis zunächst anteilig aus den verschiedenen Schätzergebnissen der einzelnen Modellbildungen. Wenn nun aber die Modellwahrscheinlichkeit des *richtigen* Modells gegen Eins konvergiert, besteht das MMAE-Ergebnis in gleichem Maße nur noch aus der Schätzung dieses einen Kalman-Filters mit der richtigen Modellbildung. Somit wählt das adaptive Kalman-Filter zwischen mehreren möglichen Ergebnissen das beste im Sinne der Modellwahrscheinlichkeit aus.

Durch diese Auswahl ergibt sich allerdings auch ein Nachteil, wenn die zu verfolgende Geräuschquelle im Laufe eines Trackingvorganges die Dynamik in der Bewegung ändert. Die Schätzergebnisse für die verschiedenen Modelle können um einige Zentimeter von einander abweichen. Wenn sich nun die Wahrscheinlichkeitsverhältnisse für die einzelnen Modelle schnell umkehren, so wechselt das Ergebnis des MMAE-Schätzers auch schnell zwischen den einzelnen Schätzergebnissen. Dadurch „springt“ das adaptive Ergebnis zwischen den einzelnen Schätzwerten. Um diesen Effekt zu vermeiden, sollten die verschiedenen Modelle nicht zu ähnlich gewählt werden. Aus diesem Grund wurden auch nicht mehrere Modelle gleicher Dynamik mit unterschiedlichen Annahmen für das Messrauschen implementiert.

Für den Sonderfall der reinen Prädiktion ist es sinnvoll, die Modellwahrscheinlichkeiten konstant zu halten, bis der nächste Messwert zur Verfügung steht. Somit verwendet das adaptive Kalman-Filter während dieser Zeit auch konstant die gleichen Gewichtungsfaktoren für die einzelnen Schätzergebnisse.

Es sei an dieser Stelle noch einmal bemerkt, dass das adaptive Kalman-Filter sich sehr gut eig-

5.4 Adaptive Kalman-Filter

net, um zwischen den Filterergebnissen verschiedener Bewegungsmodelle auszuwählen. Eine Mittelung der einzelnen Ergebnisse ist auf Grund der sprunghaften Änderungen der Modellwahrscheinlichkeiten jedoch nicht möglich.

5 Adaption an reale Szenarien

6 Ergebnisse

6.1 Zusammenfassung

Es hat sich bei der Analyse gezeigt, dass für robuste Eingangsdaten das Kalman-Filter gut funktioniert und es Verbesserungen bringt, die den Aufwand rechtfertigen. Ein Sprecher kann am besten verfolgt werden, wenn ein Kalman-Filter mit einem multiplen Modellansatz verwendet wird. Für die Dynamik von bewegten Geräuschquellen in Büro- oder Küchenumgebung reicht es aus, die drei Modellansätze *Stationarität*, *Bewegung mit konstanter Geschwindigkeit* und *Bewegung mit konstanter Beschleunigung* zu implementieren. Durch die Filterung kann dann eine Glättung der Messdaten erreicht werden, d.h. die Standardabweichung der einzelnen Positionsschätzungen ist geringer als die der Eingangsdaten aus dem Lokalisationsalgorithmus. Außerdem erhält man durch die Prädiktion eine kontinuierliche Trajektorie für den Positionsverlauf der verfolgten Geräuschquelle, auch wenn einzelne Messwerte nicht verfügbar sind.

Die Verbesserungen durch das Kalman-Filter sind allerdings stark von der Güte der Messdaten abhängig. Weisen diese geringe systematische Fehler auf, die zeitlich weitgehend unkorreliert sind, so ist eine sehr gute Glättung zu erreichen und das Tracken einer bewegten Geräuschquelle gut möglich. Beim Test mit *real data* hat sich aber gezeigt, dass auf Grund schlechter Zeitversatzschätzungen mit der modifizierten GCC-Methode in nur *einem* von vier Mikrofonpaaren die Ergebnisse des Lokalisationsalgorithmus so schlecht werden können, dass auch ein Kalman-Filter keine brauchbaren Ergebnisse liefern kann (Beispiel: Anhang B.1, Session 23). Vereinzelt gibt es auch Fälle, in denen die Messwerte so gut sind, dass das Kalman-Filter die Positionsschätzung sogar geringfügig verschlechtert. Dies ist bedingt durch eine Überschätzung der Messfehlerkovarianz (siehe Kapitel 4.2) und tritt vor allem bei einer Quelle nahe am Array auf, wenn die Bewegung in einer Koordinatenrichtung stationär ist, während in den anderen Koordinatenrichtungen eine lineare oder beschleunigte Bewegung vorliegt (Beispiel: Anhang B.1, Session 19).

Eine allgemeingültige quantitative Aussage über die Verbesserung durch das Kalman-Filter bei der Verwendung in realen Szenarien ist nicht möglich. Zum einen ist bei dem Tracken von Sprechern die Anzahl der fehlenden Messwerte in Sprechpausen und damit die Länge der reinen Prädiktion und die Anzahl der Trackingpausen sehr unterschiedlich. Diese Parameter haben jedoch durch die erhöhte Fehlervarianz einen großen Einfluss auf den Filtergewinn. Zum anderen wird das Kalman-Filter auch mit den *mittleren* Werten für die Standardabweichungen der Messdaten parametrisiert. Da die tatsächlichen Werte davon abweichen, schwankt auch die Leistungsfähigkeit des Kalman-Filters in der Anwendung.

Im Folgenden werden die zeitlichen Verläufe mehrerer Szenarien dargestellt. Zum besseren Ver-

6 Ergebnisse

gleich der Standardabweichungen vor und nach der Filterung sind die sphärischen Koordinaten der Position im Raum aufgetragen.

6.2 Tracking einer ruhenden Geräuschquelle

Die folgende Abbildung 6.1 zeigt den Verlauf der Positionsschätzung beim Tracken eines stationären Sprechers als Beispiel für eine ruhende Geräuschquelle. In Anhang B.1 ist diese Aufnahme als *Session 19* zu finden. Der Sprecher befindet sich in der in Abschnitt 5.1 beschriebenen Büroumgebung bei der Position

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -0.14 \text{ m} \\ 1.5 \text{ m} \\ -0.21 \text{ m} \end{pmatrix}, \quad (6.1)$$

dies entspricht den sphärischen Koordinaten

$$\begin{pmatrix} \theta \\ \phi \\ r \end{pmatrix} = \begin{pmatrix} 95^\circ \\ -7.9^\circ \\ 1.5 \text{ m} \end{pmatrix}. \quad (6.2)$$

Die Abweichung vom Mittelwert gibt in diesem stationären Fall die Standardabweichung des Messrauschens an:

$$\begin{pmatrix} \sigma_\Theta \\ \sigma_\Phi \\ \sigma_R \end{pmatrix}^{(mess)} = \begin{pmatrix} 0.044 \text{ rad} \\ 0.017 \text{ rad} \\ 19 \text{ cm} \end{pmatrix} = \begin{pmatrix} 2.5^\circ \\ 0.99^\circ \\ 19 \text{ cm} \end{pmatrix}. \quad (6.3)$$

Der Lokalisationsalgorithmus liefert offensichtlich mit Ausnahme des Abstandes recht gute Messwerte, vor allem die berechneten Elevationswinkel stimmen schon sehr gut und weisen geringe Varianz auf. Auch die Standardabweichung für den Winkel Azimut ist schon sehr gut, wenn man den Wert mit dem akustischen Auflösungsvermögen des Menschen vergleicht, das bei $\pm 8^\circ$ liegt [CaH99]. Das adaptive Kalman-Filter folgt erwartungsgemäß dem stationären Modell und weist entsprechend wenig Dynamik auf. Dadurch wird die Standardabweichung gesenkt auf

$$\begin{pmatrix} \sigma_\Theta \\ \sigma_\Phi \\ \sigma_R \end{pmatrix}^{(filt)} = \begin{pmatrix} 0.016 \text{ rad} \\ 0.013 \text{ rad} \\ 9 \text{ cm} \end{pmatrix} = \begin{pmatrix} 0.9^\circ \\ 0.8^\circ \\ 9 \text{ cm} \end{pmatrix}, \quad (6.4)$$

es hat also eine sichtbare Glättung stattgefunden. Es ist zu bemerken, dass für den Abstand damit zwar die Varianz reduziert wird, insgesamt der Fehler aber steigt, da grundsätzlich der Abstand schon durch den Lokalisationsalgorithmus um ca. 40 cm unterschätzt wird und dieser systematische Fehler durch das Kalman-Filter natürlich nicht ausgeglichen werden kann.

6.2 Tracking einer ruhenden Geräuschquelle

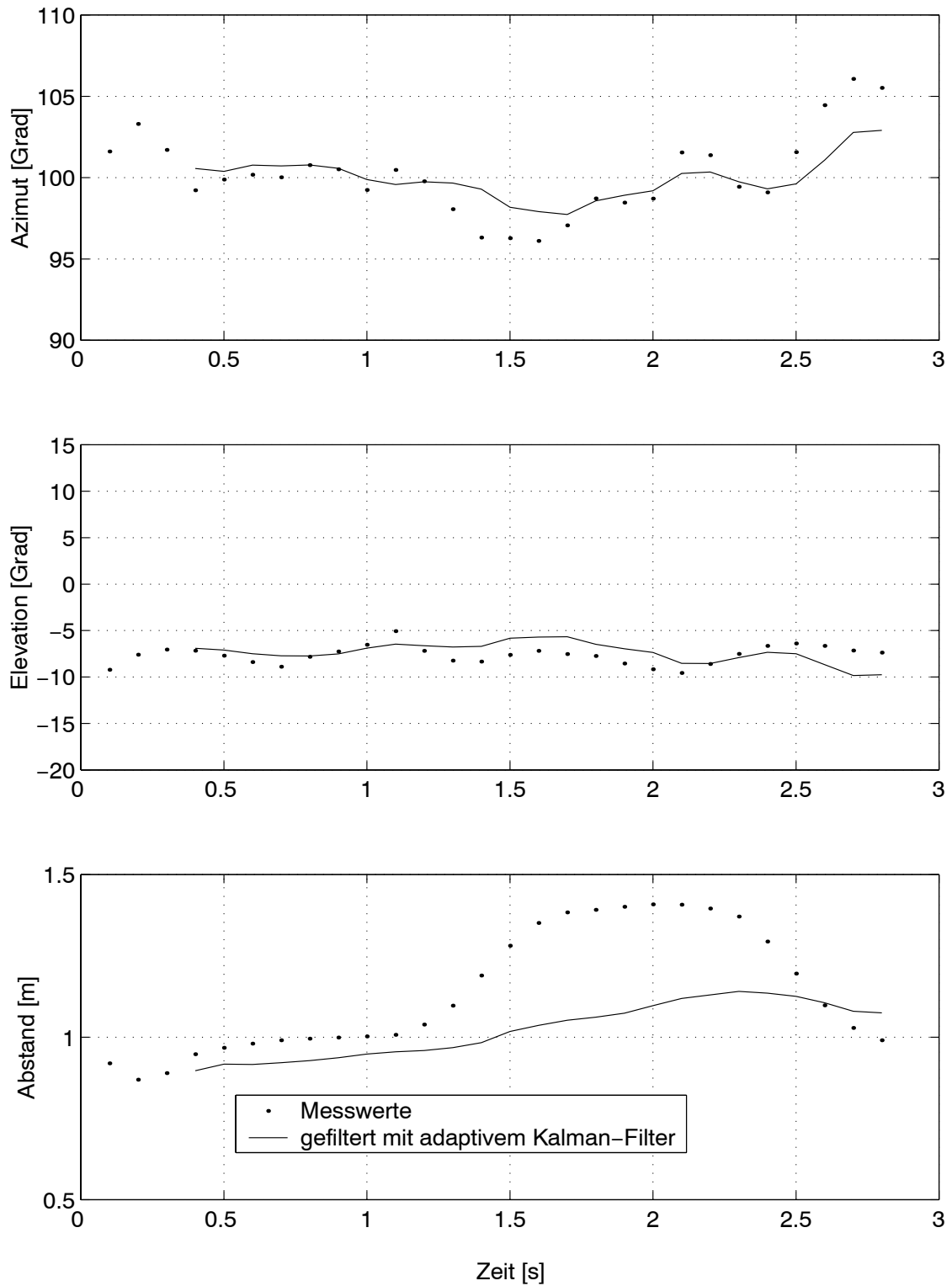


Abbildung 6.1: Sphärische Koordinaten beim Tracken einer ruhenden Geräuschquelle

6.3 Tracking einer linear bewegten Geräuschquelle

Als zweites soll das Tracken eines gehenden Sprechers als Beispiel einer linear bewegten Geräuschquelle betrachtet werden. Diese Aufnahme ist in Anhang B.1 als *Session 43* zu finden. Der Sprecher bewegt sich vom Mikrofonarray aus gesehen von rechts nach links:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0.86 \text{ m} \\ 1 \text{ m} \\ 0.13 \text{ m} \end{pmatrix} \mapsto \begin{pmatrix} -1.14 \text{ m} \\ 1 \text{ m} \\ 0.13 \text{ m} \end{pmatrix}. \quad (6.5)$$

Dies entspricht in sphärischen Koordinaten der Trajektorie

$$\begin{pmatrix} \theta \\ \phi \\ r \end{pmatrix} = \begin{pmatrix} 49^\circ \\ 5.6^\circ \\ 1.3 \text{ m} \end{pmatrix} \mapsto \begin{pmatrix} 139^\circ \\ 4.9^\circ \\ 1.5 \text{ m} \end{pmatrix}. \quad (6.6)$$

Der Positionsverlauf ist in x-Richtung linear, während er in y- und z-Richtung konstant bleibt. Entsprechend steigt (nach Umrechnung in sphärische Koordinaten) der Winkel Azimut stark an, während die Elevation nahezu konstant bleibt und sich der Abstand in der Mitte des Verlaufes auf $r = 1 \text{ m}$ verringert. Abbildung 6.2 veranschaulicht noch einmal, dass der Verlauf der sphärischen Koordinaten nicht linear ist im Gegensatz zu den kartesischen Koordinaten (siehe auch Abschnitt 3.1).

Den Vergleich der Messwerte mit dem Ergebnis des adaptiven Kalman-Filters zeigt Abbildung 6.3 für die einzelnen Koordinaten und Abbildungen 6.4 und 6.5 in einer räumlichen Darstellung. Die Berechnungen des Lokalisationsalgorithmus sind in diesem Fall schon recht gut, vor allem bezüglich θ und ϕ . Am Winkel Azimut ist zu sehen, dass das Kalman-Filter gut dem linearen Verlauf folgt. Deshalb sind aber auch die Schätzungen für Elevation und Abstand einer größeren Dynamik unterworfen. Trotzdem kann gerade für den Abstand eine gute Glättung erreicht werden.

Eine quantitative Aussage über die Veränderung der Varianz ist bei diesem linearen Fall schwieriger als im vorigen Abschnitt. Bestimmt man für die Messwerte und die Schätzergebnisse die Standardabweichung bezogen auf eine Regressionsgerade, so ist nur bei dem Abstand ein Gewinn zu erzielen. Die Standardabweichung σ_R verringert sich in diesem Fall um 6 cm. Der Einsatz des Kalman-Filters bringt neben der Glättung bei dieser Aufnahme aber noch einen weiteren Vorteil durch die reine Prädiktion. In dem hervorgehobenen Bereich [3.6 s, 4.1 s] wird ohne Messwerte die Geräuschquelle trotzdem kontinuierlich weiter verfolgt. Dies könnte wichtig werden, wenn man beispielsweise an eine Datenfusion mit visuellen Daten denkt [Goo97].

6.3 Tracking einer linear bewegten Geräuschquelle

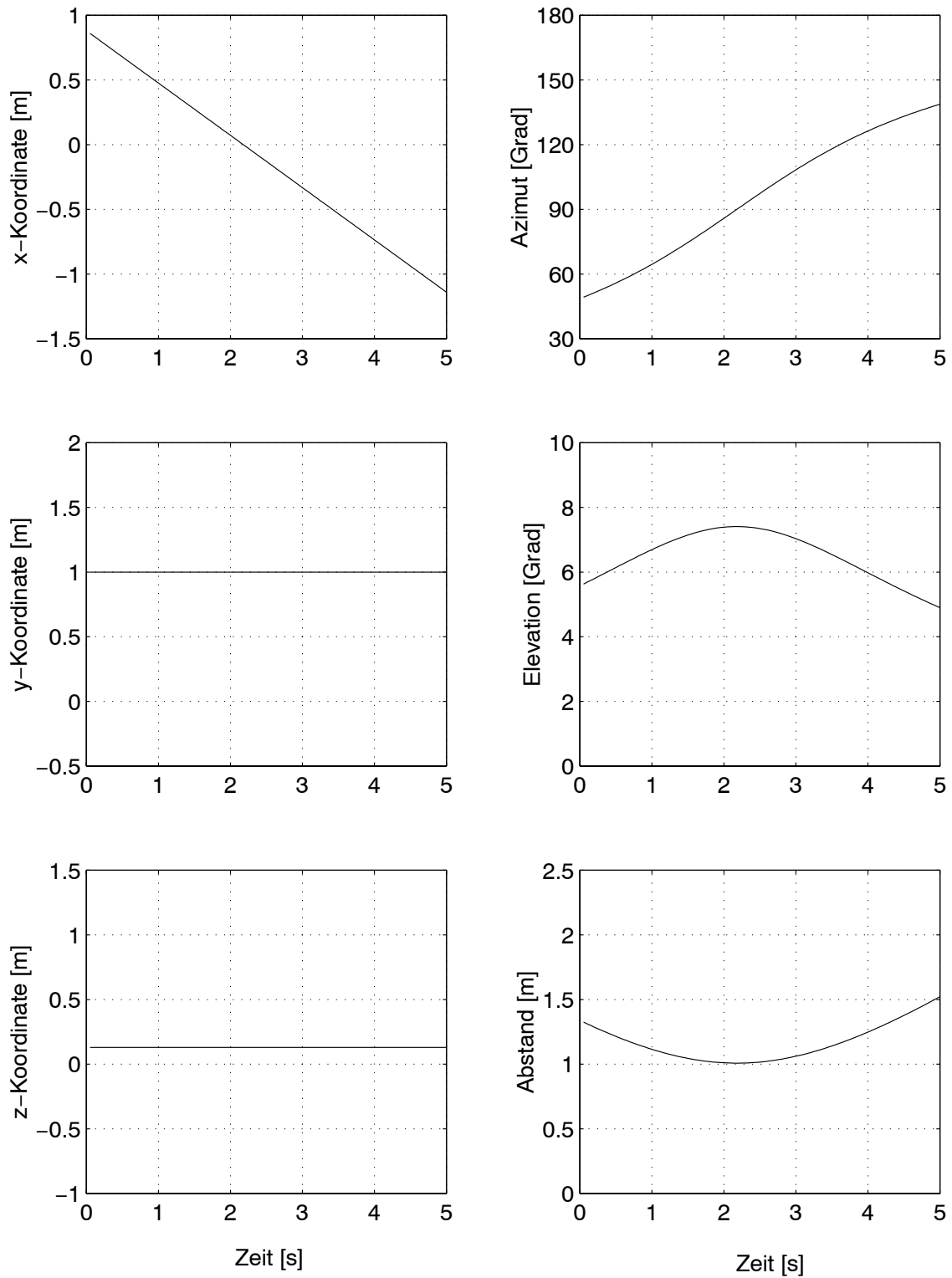


Abbildung 6.2: Koordinatenverläufe der vorgegebenen Trajektorie

6 Ergebnisse

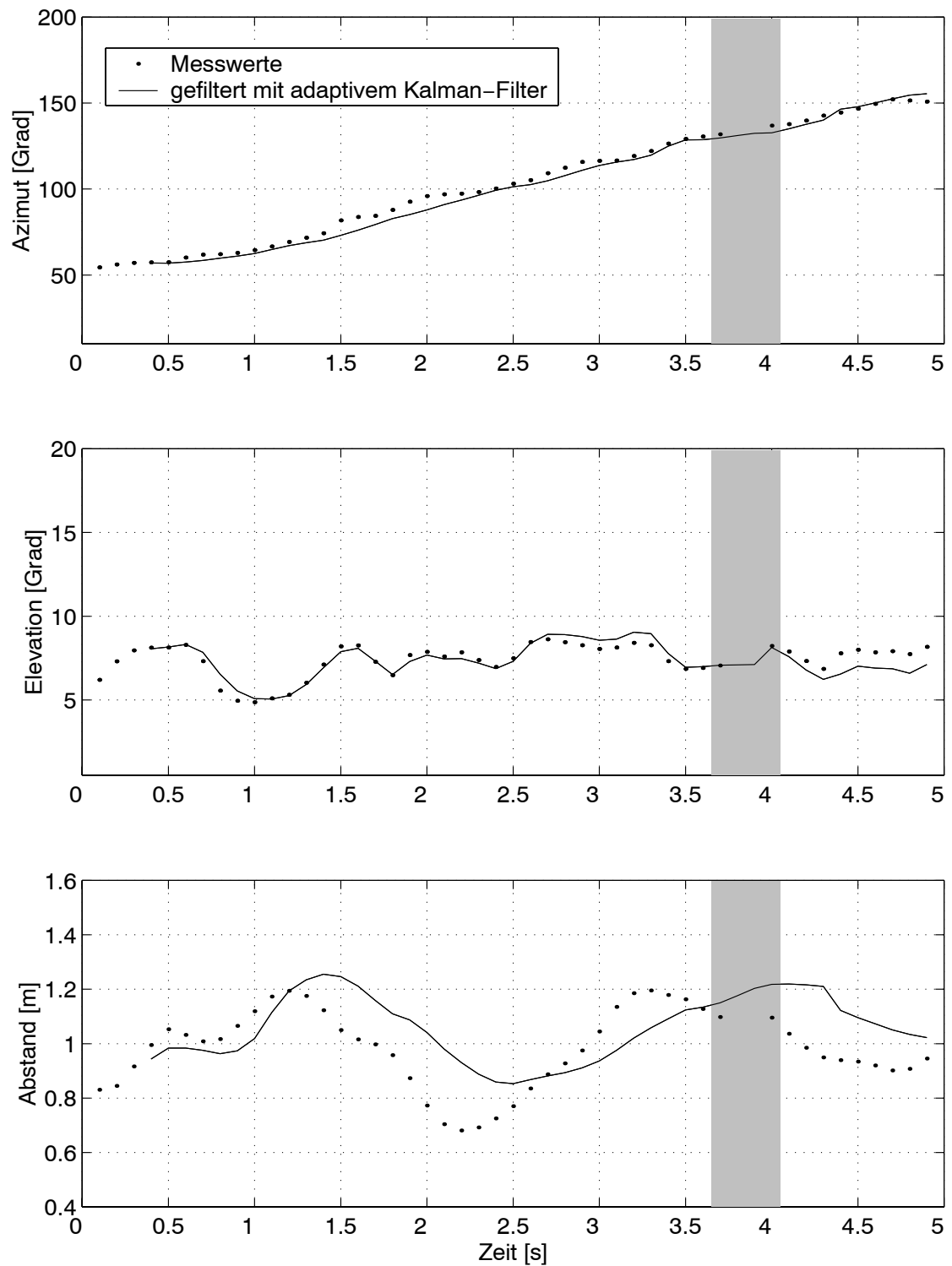


Abbildung 6.3: Sphärische Koordinaten beim Tracken einer linear bewegten Geräuschquelle

6.3 Tracking einer linear bewegten Geräuschquelle

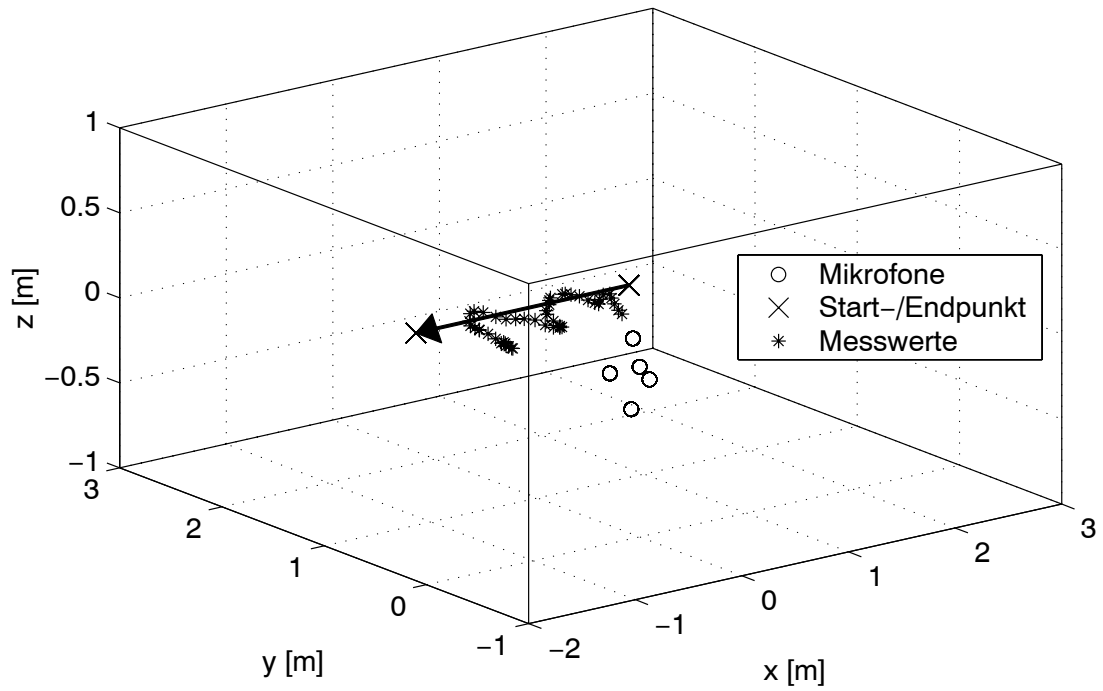


Abbildung 6.4: Lokalisation einer linear bewegten Geräuschquelle (vor dem Filtervorgang)

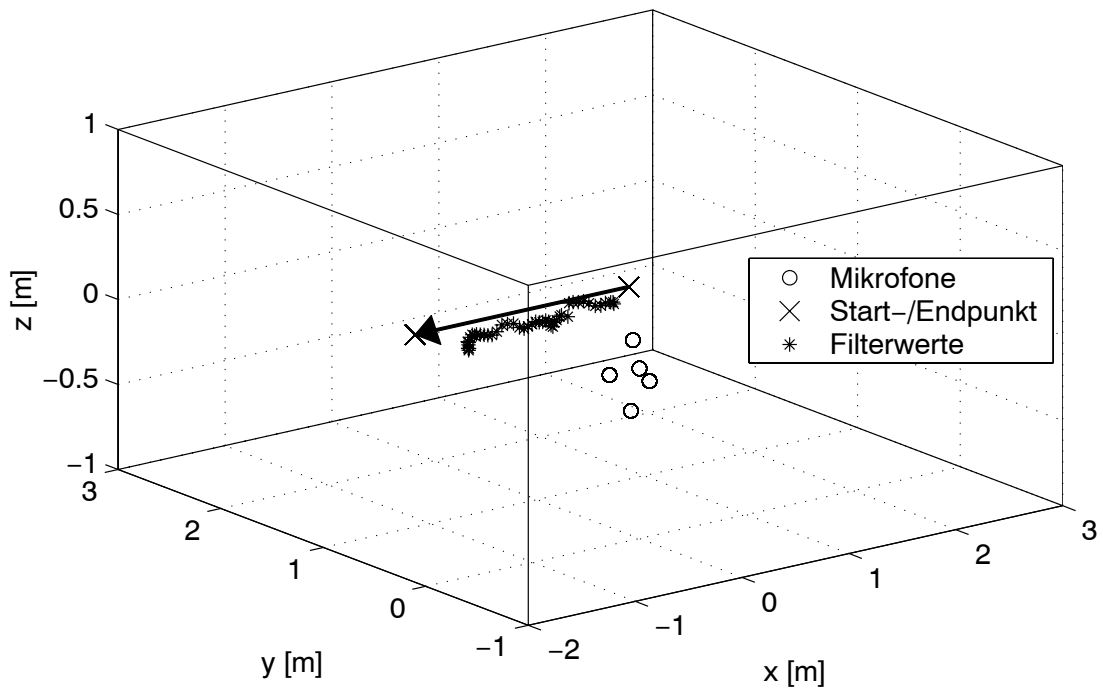


Abbildung 6.5: Tracking einer linear bewegten Geräuschquelle (nach dem Filtervorgang)

6.4 Tracking einer beschleunigten Geräuschquelle

Vor der Bestimmung der Zeitversätze wird eine kurzzeitige Mittelung der Mikrofondifferenzsignale durchgeführt, um eine große Verlässlichkeit der TDEs zu gewährleisten (siehe Kapitel 1). Dadurch reduziert sich jedoch die Dynamik in dem Positionsverlauf der Geräuschquelle. In Küchen- oder Büroumgebung sind Bewegungsvorgänge mit konstanter Beschleunigung ohnehin nur für eine kurze Zeitdauer denkbar. Deshalb überwiegt im Ergebnis des adaptiven Kalman-Filters nur in wenigen Fällen die Schätzung des Modellansatzes mit konstanter Beschleunigung. Trotzdem findet das beschleunigte Modell Verwendung in der Schätzung mit dem multiplen Modellansatz. Insbesondere wurde es bei Übergängen zwischen stationärem und linearem Verhalten eingesetzt (Beispiel: Anhang B.1, Session 24). Gerade für solche kurze Dynamikänderungen folgt das beschleunigte Modell der Bewegung der Geräuschquelle am besten. Aus diesem Grund ist der Aufwand eines beschleunigten Modellansatzes im adaptiven Kalman-Filter berechtigt, auch wenn die beiden anderen Modelle weitaus häufiger das Gesamtergebnis bestimmen.

7 Ausblick

Bisher wurden sämtliche Aufnahmen zum Test des Kalman-Filters in derselben Büroumgebung gemacht. Da der humanoide Roboter sein Einsatzgebiet neben dem Büro auch in der Küche haben soll, wäre es sinnvoll, einige Tests auch in Küchenumgebung durchzuführen. Wegen anderen akustischen Bedingungen, vor allem mehr Echoeinflüsse, kann eine andere Parametrisierung des Filters notwendig werden.

Um eine größere Leistungsfähigkeit des Kalman-Filters beim passiven Audio-Tracken zu erreichen, könnte man die Statistik der Messfehler weiter verbessern. Je mehr Information über die funktionale Abhängigkeit des Mess- und Systemfehlers von der Versuchsanordnung und der Position der Geräuschquelle verwendet werden kann, um die Matrizen zu parametrisieren, desto besser können diese Störgrößen durch das Filter ausgegletzt werden. Gerade die Kenntnis über die bekannten systematischen Fehler wie die Unterschätzung des Abstandes auf Grund der Verwendung eines suboptimalen closed-form Lokalisationsalgorithmus könnte für eine Verbesserung des Filtergewinns herangezogen werden. Insbesondere die Robustheit der TDOA Schätzungen ist für das Kalman-Filter entscheidend, deshalb könnten für deren Bestimmung zusätzliche Konfidenzkriterien herangezogen werden [BK02].

In Kapitel 3.1 wurde bereits angesprochen, dass die numerische Berechnung der Kovarianzmatrizen recht aufwändig ist. In diesem Zusammenhang könnte man durch schnellere Algorithmen die Echtzeitfähigkeit noch verbessern. Der Rechenaufwand könnte auch gesenkt werden, wenn die Berechnung nicht für jeden Messwert durchgeführt wird, sondern für eine gewisse Dauer Quasistationarität angenommen wird (siehe Kapitel 3.1).

Um letztlich zu einer vollständigen akustischen Szenenanalyse zu gelangen, kann diese Arbeit aufbauend auf die Lokalisationsaufgabe mit der Quellentrennung und der Geräuschquellenklassifikation kombiniert werden. Einzelne Aufnahmen mit zwei sich abwechselnden Sprechern zeigen bereits, dass das Kalman-Filter auch dafür verwendet werden kann. Für das gleichzeitige Tracken mehrerer Geräuschquellen werden dann allerdings auch entsprechend viele parallel arbeitende Kalman-Filter benötigt. Inwieweit dadurch die einzelnen Schätzungen beeinflusst werden, bleibt noch zu untersuchen.

Eine große Verbesserung bei dem Tracken von Geräuschquellen könnte auch durch die Datenfusion von akustischen und visuellen Daten erreicht werden. In einem multimodalen System könnten dann die Vorteile jedes Sensortyps genutzt werden, um insgesamt zu einem besseren Lokalisationsergebnis zu kommen, als es mit einem passiven Mikrofonarray schon möglich ist.

7 Ausblick

A Koordinatentransformation

An dieser Stelle soll eine detaillierte Betrachtung der in Kapitel 3.1 benötigten Koordinatentransformation von sphärischen in kartesische Koordinaten durchgeführt werden. Das Ziel ist, eine Kovarianzmatrix geschlossen zu transformieren:

$$\mathbf{R}_{NN}^{(sph)} \rightarrow \mathbf{R}_{NN}^{(kart)}.$$

Gegeben sind die Zufallsvariablen Azimut Θ , Elevation Φ und Abstand R mit den Wahrscheinlichkeitsdichtefunktionen

$$f_{\Theta}(\theta) = \frac{1}{\sqrt{2\pi}\sigma_{\Theta}} \exp\left(-\frac{\theta^2}{2\sigma_{\Theta}^2}\right) \quad (\text{A.1})$$

$$f_{\Phi}(\phi) = \frac{1}{\sqrt{2\pi}\sigma_{\Phi}} \exp\left(-\frac{\phi^2}{2\sigma_{\Phi}^2}\right) \quad (\text{A.2})$$

$$f_R(r) = \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left(-\frac{r^2}{2\sigma_R^2}\right) \quad (\text{A.3})$$

Da diese Zufallsvariablen statistisch unabhängig sind, besitzt der dreidimensionale Zufallsvariablenvektor $(\Theta, \Phi, R)^T$ den Mittelwert $(0, 0, 0)^T$, die Kovarianzmatrix

$$\mathbf{R}_{NN}^{(sph)} = \begin{pmatrix} \sigma_{\Theta}^2 & 0 & 0 \\ 0 & \sigma_{\Phi}^2 & 0 \\ 0 & 0 & \sigma_R^2 \end{pmatrix} \quad (\text{A.4})$$

und die Verbunddichtefunktion

$$f_{\Theta, \Phi, R}(\theta, \phi, r) = f_{\Theta}(\theta) \cdot f_{\Phi}(\phi) \cdot f_R(r). \quad (\text{A.5})$$

Gesucht ist nun die Wahrscheinlichkeitsdichtefunktion des kartesischen Zufallsvariablenvektors $(X, Y, Z)^T$. Dazu benötigt man zunächst die funktionale Abhängigkeit der beiden Koordinatensysteme. Für die kartesischen Koordinaten gilt:

$$X = R \cos \Theta \cos \Phi \quad (\text{A.6})$$

$$Y = R \cos \Theta \sin \Phi \quad (\text{A.7})$$

$$Z = R \sin \Phi. \quad (\text{A.8})$$

A Koordinatentransformation

Die Umrechnung in sphärische Koordinaten berechnet man nach

$$\Theta = \arctan \frac{Y}{X} = \arccos \frac{X}{\sqrt{X^2 + Y^2}} \quad (\text{A.9})$$

$$\Phi = \arctan \frac{Z}{\sqrt{X^2 + Y^2}} \quad (\text{A.10})$$

$$R = \sqrt{X^2 + Y^2 + Z^2}. \quad (\text{A.11})$$

Es soll nun die Verbunddichte $f_{\Theta, \Phi, R}(\theta, \phi, r)$ in die Verbunddichte $f_{X, Y, Z}(x, y, z)$ transformiert werden. Dazu berechnet man die *Jacobi-Determinante* $|\mathbf{J}|$ aus der funktionalen Abhängigkeit. Mit

$$|\mathbf{J}| = \begin{vmatrix} \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial r} \\ \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial r} \\ \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial r} \end{vmatrix} = |r^2 \cos \theta| \quad (\text{A.12})$$

lässt sich diese gemeinsame Dichte nun nach [Jon00] berechnen:

$$f_{X, Y, Z}(x, y, z) = f_{\Theta, \Phi, R}(\theta, \phi, r) \cdot |\mathbf{J}|^{-1} \quad (\text{A.13})$$

$$\begin{cases} \theta = \theta(x, y, z) \\ \phi = \phi(x, y, z) \\ r = r(x, y, z) \end{cases}$$

und schließlich

$$f_{X, Y, Z}(x, y, z) = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_{\Theta} \sigma_{\Phi} \sigma_R} \exp \left[-\frac{1}{2} \left(\left(\frac{\arctan \frac{y}{x}}{\sigma_{\Theta}} \right)^2 + \left(\frac{\arctan \frac{z}{\sqrt{x^2 + y^2}}}{\sigma_{\Phi}} \right)^2 + \left(\frac{\sqrt{x^2 + y^2 + z^2}}{\sigma_R} \right)^2 \right) \right] \cdot \frac{\sqrt{x^2 + y^2}}{(x^2 + y^2 + z^2) |x|}. \quad (\text{A.14})$$

Für die einzelnen Elemente der Kovarianzmatrix sind die Erwartungswerte zweiter Ordnung zu berechnen; für die Varianz der x-Koordinate ist dies

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2 f_X(x) dx. \quad (\text{A.15})$$

Deshalb benötigt man explizit eine Darstellung der Randdichten $f_X(x)$, $f_Y(y)$ und $f_Z(z)$. Darin liegt jedoch die Problematik bei der Suche nach einer geschlossenen Lösung. Die Integrale der Form

$$f_X(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X, Y, Z}(x, y, z) dy dz \quad (\text{A.16})$$

sind, wenn man (A.14) einsetzt, nicht mehr geschlossen lösbar. Aus diesem Grund wird das in Kapitel 3.1 beschriebene numerische Verfahren angewendet.

B Aufnahmen

B.1 Übersicht

Tabelle B.1 zeigt eine Übersicht über die Szenarien, die für den Test des Kalman-Filters herangezogen wurden. Es ist jeweils der Sprecher, der Positionsverlauf und die Dynamik der Bewegung angegeben.

- **Sprecher:** Es sind die Initialen der Sprecher angegeben. Dabei steht DB für Dirk Bechler und MG für Michael Grimm.
- **Satz:** Die Angaben zum gesprochenen Text beziehen sich auf die Nummerierung in Anhang B.2. *A* bedeutet, dass konstant der Vokal *A* artikuliert wurde, um Szenarien ohne fehlende Messwerten zu erhalten.
- **Start-/Endpunkt:** Für diese Angaben wurde ein Inertialsystem mit einem der Mikrofone im Koordinatenursprung verwendet. Die x-Koordinate des Ursprungs $x_0 = -0.14$ m ist damit negativ gleich dem halben Abstand zweier Mikrofone. Die Höhe z_0 dieses Nullpunktes über dem Boden beträgt $z_0 = 1.47$ m. Daraus werden folgende Größen abgeleitet:
 - $hd = 1.76$ m: Mundhöhe aufrechter Stand Dirk Bechler. Dies entspricht im Koordinatensystem dem Wert $z = +0.29$.
 - $hm = 1.60$ m: Mundhöhe aufrechter Stand Michael Grimm. Dies entspricht im Koordinatensystem dem Wert $z = +0.13$.
 - $sd = 1.26$ m: Mundhöhe aufrechtes Sitzen Dirk Bechler. Dies entspricht im Koordinatensystem dem Wert $z = -0.21$.
 - $sm = 1.17$ m: Mundhöhe aufrechtes Sitzen Michael Grimm. Dies entspricht im Koordinatensystem dem Wert $z = -0.30$.

Session Nr.	Sprecher	Satz	Startpunkt	Endpunkt	Bewegung
1	DB	1a	(-1,2,hd)	(1,2,hd)	linear
2	DB	2b	(1,2,hd)	(-1,1,hd)	linear
3	MG	2b	(1,2,hm)	(-1,1,hm)	linear

B Aufnahmen

Session Nr.	Sprecher	Satz	Startpunkt	Endpunkt	Bewegung
4	DB	20b	(1.5,2,hd)	(-1,2,hd)	linear
5	MG	20b	(1.5,2,hm)	(-1,2,hm)	linear
6	DB	3a	(-1,2.5,hd)	(-1,0.1,hd)	linear
7	MG	3a	(-1,2.5,hm)	(-1,0.15,hm)	linear
8	DB	6a	(0.5,0,hd)	(-1,2.25,hd)	linear
9	MG	6a	(0.5,0,hm)	(-1,2.25,hm)	linear
10	DB	8a	(0.5,2,hd)	(0.5,2,sd)	linear
11	MG	8a	(0.5,2,hm)	(0.5,2,sm)	linear
12	DB	20a	(0,1,hd)	(0,1,sd)	linear
13	MG	20a	(0,1,hm)	(0,1,sm)	linear
14	DB	14b	(-0.5,1,hd)		stationär
15	MG	14b	(-0.5,1,hm)		stationär
16	DB	14a	(0.5,2,hd)		stationär
17	MG	14a	(0.5,2,hm)		stationär
18	DB	A	(0,1.5,sd)		stationär
19	DB	9a	(0,1.5,sd)		stationär
20	MG	A	(0,1.5,sm)		stationär
21	MG	9a	(0,1.5,sm)		stationär
22	DB	18a	(1.5,2,hd)	(-1,1,hd)	beschleunigt
23	MG	18a	(1.5,2,hm)	(-1,1,hm)	beschleunigt
24	DB	19a	(0.5,2.5,hd)	(0.5,0.1,hd)	beschleunigt
25	MG	19a	(0.5,2.5,hm)	(0.5,0.1,hm)	beschleunigt
26	DB	A	(-1,2.5,hd)	(-1,0.1,hd)	linear
27	MG	A	(-1,2.5,hm)	(-1,0.1,hm)	linear
28	DB	A	(1.5,1.5,hd)	(-1,1.5,hd)	linear
29	MG	A	(1.5,1.5,hm)	(-1,1.5,hm)	linear
30	DB	A	(1,2,hd)	(-1,1,hd)	linear
31	MG	A	(1,2,hm)	(-1,1,hm)	linear
32	DB	A	(-1,1.5,hd)		stationär
	MG	A	(1.5,1.5,hm)		stationär
33	DB	3a	(-1,1.5,hd)		stationär
	MG	3a	(1.5,1.5,hm)		stationär
34	DB	A	(-1,2.5,hd)	(-1,0,hd)	linear
	MG	A	(1.5,1.5,hm)		stationär
35	DB	Hallo	(-1,2.5,hd)	(-1,0,hd)	linear
	MG	Hallo	(1.5,1.5,hm)		stationär
36	DB	A	(0,0,hd)	(0,2,hd)	linear
37	DB	A	(0,0,hd)	(0,2,hd)	linear
38	DB	A	(0,2,hd)		stationär
39	DB	A	(1,1,hd)	(-1,1,hd)	linear
40	DB	1a	(1,1.5,hd)	(-1,1.5,hd)	linear

Session Nr.	Sprecher	Satz	Startpunkt	Endpunkt	Bewegung
41	DB	1a	(1,1,hd)	(-1,1,hd)	linear
42	DB	1a	(1,1,hd)	(-1,1,hd)	linear
43	MG	1a	(1,1,hm)	(-1,1,hm)	linear

Tabelle B.1: Aufgenommene Szenarien

B.2 Testsätze

Folgende Tabelle enthält die zum Test des Kalman-Filters verwendeten Testsätze zur Spracherkennung nach *Beckmann-Schilling*.

1a	Der Schnellzug in Richtung Kiel verspätet sich um eine Stunde.
2a	Witwe sucht morgens Arbeit im Haushalt.
3a	Der Anwalt fordert Freispruch wegen erwiesener Unschuld.
4a	Reinecke Fuchs belog auch den König der Tiere.
5a	Jener Kran trägt im Höchstfall 50 Tonnen.
6a	Mit einer Spritze impfte der Schularzt die Kinder gegen Scharlach.
7a	Ein hoher Gewinn fiel an einen Bremer Rentner.
8a	Die Sendung wird am Mittwoch nächster Woche wiederholt.
9a	Der Täter trug eine blaue Mütze mit Schirm.
10a	Sandboden eignet sich gut für den Anbau von Spargel.
11a	Wegen des Festes erscheint die nächste Zeitung am Dienstag.
12a	Der Wind drehte am Abend von Nord nach Südwest.
13a	Um drei Prozent stieg der Preis der Butter.
14a	In der zweiten Halbzeit spielte die Eintracht gegen den Wind.
15a	Fahrpreis für einen Kilometer: sieben Pfennig.
16a	das Skelett des Menschen besteht aus zweihundert Knochen.
17a	Die schweizer Flagge: Rot mit weißem Kreuz.
18a	Die Schonzeit des Rehbocks reicht von Oktober bis Mai.
19a	Das gesamte Schulwesen steht unter der Aufsicht des Staates.
20a	Eier enthalten soviel Eiweiß wie Fett.
1b	Zum Eilzug nach Gießen: Heute Bahnsteig vier.
2b	Händler sucht ein leeres Zimmer in der Stadtmitte.
3b	Das Urteil lautet auf drei Monate Gefängnis.
4b	Am Bache drohte der Wolf dem kleinen Lämmchen.
5b	Ein einziger Schalter bringt den Wagen zum Stehen.
6b	Mit einer Zange entfernt der Zahnarzt die kranke Wurzel.
7b	Der Gewinner bedenkt die Wohlfahrt mit einer reichen Spende.
8b	Um siebzehn Uhr hatte der Sender Köln eine Störung.

B Aufnahmen

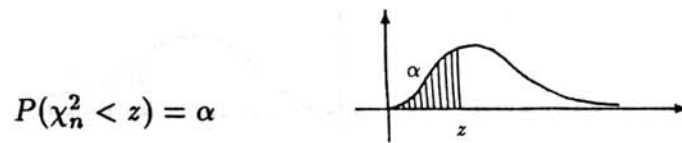
9b	Beide Räuber wurden nach kurzem Gefecht getötet.
10b	In diesem Jahr fiel die Obsternte gering aus.
11b	In der nächsten Nummer brachte die Zeitschrift einen Bericht.
12b	Das Wetter: im Flachland Regen, in den Bergen Schnee.
13b	Nach geraumer Zeit geriet so das Geschäft in Konkurs.
14b	Spannend bis zum letzten Augenblick war der Wettlauf der Damen.
15b	Zuschlag für Schnellzüge beträgt zwei Mark.
16b	Das menschliche Herz wiegt dreihundert Gramm.
17b	Italiens Farben: Grün - Weiß - Rot.
18b	Schildkröten erreichen ein Alter von dreihundert Jahren.
19b	Nach Artikel drei sind Männer und Frauen gleichberechtigt.
20b	Pfifferlinge sind essbar, Fliegenpilze gefährlich giftig.

Tabelle B.2: Testsätze zur Spracherkennung nach Beckmann-Schilling

C Tabellen zur χ^2 -Verteilung

Folgende Tabellen enthalten die Quantile $\chi_{n;p}^2$ von χ^2 -verteilten Zufallsvariablen mit n Freiheitsgraden [MW96].

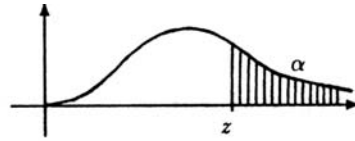
C Tabellen zur χ^2 -Verteilung



n	$\alpha=0.5\%$	$\alpha=1\%$	$\alpha=2.5\%$	$\alpha=5\%$	n	$\alpha=0.5\%$	$\alpha=1\%$	$\alpha=2.5\%$	$\alpha=5\%$
1			0.001	0.004	51	28.7	30.5	33.2	35.6
2	0.010	0.020	0.051	0.103	52	29.5	31.2	34.0	36.4
3	0.072	0.115	0.216	0.352	53	30.2	32.0	34.8	37.3
4	0.207	0.297	0.484	0.711	54	31.0	32.8	35.6	38.1
5	0.412	0.554	0.831	1.15	55	31.7	33.6	36.4	39.0
6	0.676	0.872	1.24	1.64	56	32.5	34.3	37.2	39.8
7	0.989	1.24	1.69	2.17	57	33.2	35.1	38.0	40.6
8	1.34	1.65	2.18	2.73	58	34.0	35.9	38.8	41.5
9	1.73	2.09	2.70	3.33	59	34.8	36.7	39.7	42.3
10	2.16	2.56	3.25	3.94	60	35.5	37.5	40.5	43.2
11	2.60	3.05	3.82	4.57	61	36.3	38.3	41.3	44.0
12	3.07	3.57	4.40	5.23	62	37.1	39.1	42.1	44.9
13	3.56	4.11	5.01	5.89	63	37.8	39.9	43.0	45.7
14	4.07	4.66	5.63	6.57	64	38.6	40.6	43.8	46.6
15	4.60	5.23	6.26	7.26	65	39.4	41.4	44.6	47.4
16	5.14	5.81	6.91	7.96	66	40.2	42.2	45.4	48.3
17	5.70	6.41	7.56	8.67	67	40.9	43.0	46.3	49.2
18	6.26	7.01	8.23	9.39	68	41.7	43.8	47.1	50.0
19	6.84	7.63	8.91	10.1	69	42.5	44.6	47.9	50.9
20	7.43	8.26	9.59	10.8	70	43.3	45.4	48.8	51.7
21	8.03	8.90	10.3	11.6	71	44.1	46.2	49.6	52.6
22	8.64	9.54	11.0	12.3	72	44.8	47.1	50.4	53.5
23	9.26	10.2	11.7	13.1	73	45.6	47.9	51.3	54.3
24	9.89	10.9	12.4	13.8	74	46.4	48.7	52.1	55.2
25	10.5	11.5	13.1	14.6	75	47.2	49.5	52.9	56.0
26	11.2	12.2	13.8	15.4	76	48.0	50.3	53.8	56.9
27	11.8	12.9	14.6	16.2	77	48.8	51.1	54.6	57.8
28	12.5	13.6	15.3	16.9	78	49.6	51.9	55.5	58.7
29	13.1	14.3	16.0	17.7	79	50.4	52.7	56.3	59.5
30	13.8	15.0	16.8	18.5	80	51.2	53.5	57.2	60.4
31	14.5	15.7	17.5	19.3	81	52.0	54.4	58.0	61.3
32	15.1	16.4	18.3	20.1	82	52.8	55.2	58.8	62.1
33	15.8	17.1	19.0	20.9	83	53.6	56.0	59.7	63.0
34	16.5	17.8	19.8	21.7	84	54.4	56.8	60.5	63.9
35	17.2	18.5	20.6	22.5	85	55.2	57.6	61.4	64.7
36	17.9	19.2	21.3	23.3	86	56.0	58.5	62.2	65.6
37	18.6	20.0	22.1	24.1	87	56.8	59.3	63.1	66.5
38	19.3	20.7	22.9	24.9	88	57.6	60.1	63.9	67.4
39	20.0	21.4	23.6	25.7	89	58.4	60.9	64.8	68.2
40	20.7	22.2	24.4	26.5	90	59.2	61.8	65.6	69.1
41	21.4	22.9	25.2	27.3	91	60.0	62.6	66.5	70.0
42	22.1	23.6	26.0	28.1	92	60.8	63.4	67.4	70.9
43	22.9	24.4	26.8	29.0	93	61.6	64.2	68.2	71.8
44	23.6	25.1	27.6	29.8	94	62.4	65.1	69.1	72.6
45	24.3	25.9	28.4	30.6	95	63.2	65.9	69.9	73.5
46	25.0	26.7	29.2	31.4	96	64.1	66.7	70.8	74.4
47	25.8	27.4	30.0	32.3	97	64.9	67.6	71.6	75.3
48	26.5	28.2	30.8	33.1	98	65.7	68.4	72.5	76.2
49	27.2	28.9	31.6	33.9	99	66.5	69.2	73.4	77.0
50	28.0	29.7	32.4	34.8	100	67.3	70.1	74.2	77.9

Abbildung C.1: Untere Quantile der χ^2 -Verteilung

$$P(\chi_n^2 < z) = 1 - \alpha$$



n	$\alpha = 0.5\%$	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$	n	$\alpha = 0.5\%$	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
1	7.88	6.63	5.02	3.84	51	80.7	77.4	72.6	68.7
2	10.6	9.21	7.38	5.99	52	82.0	78.6	73.8	69.8
3	12.8	11.3	9.35	7.82	53	83.3	79.8	75.0	71.0
4	14.9	13.3	11.1	9.49	54	84.5	81.1	76.2	72.1
5	16.8	15.1	12.8	11.1	55	85.7	82.3	77.4	73.3
6	18.5	16.8	14.4	12.6	56	87.0	83.5	78.6	74.5
7	20.3	18.5	16.0	14.1	57	88.2	84.7	79.8	75.6
8	22.0	20.1	17.5	15.5	58	89.5	86.0	80.9	76.8
9	23.6	21.7	19.0	16.9	59	90.7	87.2	82.1	77.9
10	25.2	23.2	20.5	18.3	60	92.0	88.4	83.3	79.1
11	26.8	24.7	21.9	19.7	61	93.2	89.6	84.5	80.2
12	28.3	26.2	23.3	21.0	62	94.4	90.8	85.7	81.4
13	29.8	27.7	24.7	22.4	63	95.6	92.0	86.8	82.5
14	31.3	29.1	26.1	23.7	64	96.9	93.2	88.0	83.7
15	32.8	30.6	27.5	25.0	65	98.1	94.4	89.2	84.8
16	34.3	32.0	28.8	26.3	66	99.3	95.6	90.3	86.0
17	35.7	33.4	30.2	27.6	67	100.6	96.8	91.5	87.1
18	37.2	34.8	31.5	28.9	68	101.8	98.0	92.7	88.3
19	38.6	36.2	32.8	30.1	69	103.0	99.2	93.9	89.4
20	40.0	37.6	34.2	31.4	70	104.2	100.4	95.0	90.5
21	41.4	38.9	35.5	32.7	71	105.4	101.6	96.2	91.7
22	42.8	40.3	36.8	33.9	72	106.6	102.8	97.4	92.8
23	44.2	41.6	38.1	35.2	73	107.9	104.0	98.5	93.9
24	45.6	43.0	39.4	36.4	74	109.1	105.2	99.7	95.1
25	46.9	44.3	40.6	37.6	75	110.3	106.4	100.8	96.2
26	48.3	45.6	41.9	38.9	76	111.5	107.6	102.0	97.4
27	49.6	47.0	43.2	40.1	77	112.7	108.8	103.2	98.5
28	51.0	48.3	44.5	41.3	78	113.9	110.0	104.3	99.6
29	52.3	49.6	45.7	42.6	79	115.1	111.1	105.5	100.7
30	53.7	50.9	47.0	43.8	80	116.3	112.3	106.6	101.9
31	55.0	52.2	48.2	45.0	81	117.5	113.5	107.8	103.0
32	56.3	53.5	49.5	46.2	82	118.7	114.7	108.9	104.1
33	56.7	54.8	50.7	47.4	83	119.9	115.9	110.1	105.3
34	59.0	56.1	52.0	48.6	84	121.1	117.1	111.2	106.4
35	60.3	57.3	53.2	49.8	85	122.3	118.2	112.4	107.5
36	61.6	58.6	54.4	51.0	86	123.5	119.4	113.5	108.6
37	62.9	59.9	55.7	52.2	87	124.7	120.6	114.7	109.8
38	64.2	61.2	56.9	53.4	88	125.9	121.8	115.8	110.9
39	65.5	62.4	58.1	54.6	89	127.1	122.9	117.0	112.0
40	66.8	63.7	59.3	55.8	90	128.3	124.1	118.1	113.1
41	68.0	65.0	60.6	56.9	91	129.5	125.3	119.3	114.3
42	69.3	66.2	61.8	58.1	92	130.7	126.5	120.4	115.4
43	70.6	67.5	63.0	59.3	93	131.9	127.6	121.6	116.5
44	71.9	68.7	64.2	60.5	94	133.1	128.8	122.7	117.6
45	73.2	70.0	65.4	61.7	95	134.2	130.0	123.9	118.8
46	74.4	71.2	66.6	62.8	96	135.4	131.1	125.0	119.9
47	75.7	72.4	67.8	64.0	97	136.6	132.3	126.1	121.0
48	77.0	73.7	69.0	65.2	98	137.8	133.5	127.3	122.1
49	78.2	74.9	70.2	66.3	99	139.0	134.6	128.4	123.2
50	79.5	76.2	71.4	67.5	100	140.2	135.8	129.6	124.3

Abbildung C.2: Obere Quantile der χ^2 -Verteilung

C Tabellen zur χ^2 -Verteilung

Literaturverzeichnis

- [BK02] BECHLER, D. und K. KROSCHEL: *Confidence scoring of time difference of arrival estimation for speaker localization with microphone arrays*. Erscheint bei der 13. Konferenz für Elektronische Sprachsignalverarbeitung (ESSV), Dresden, September 2002.
- [Bro83] BROWN, R.G.: *Introduction to Random Signal Analysis and Kalman filtering*. Wiley, 1983.
- [BS88] BAR-SHALOM, Y.: *Tracking and Data Association*. Academic Press, Inc., San Diego, 1988.
- [BS97a] BRANDSTEIN, M.S., D.E. STURIM und H.F. SILVERMAN: *Tracking Multiple Talkers Using Microphone-Array Measurements*. Proceedings of ICASSP-1997, Munich, Germany, 1997.
- [BS97b] BRANDSTEIN, M.S. und H.F. SILVERMAN: *A Practical Methodology for Speech Source Localization with Microphone Arrays*. Computer, Speech and Language, 11(2), pp. 91-126, 1997.
- [CaH99] CAVACO, S. und J. AD HALLAM: *A biologically plausible acoustic azimuth estimation system*. Proceedings of IJCAI-99 Workshop on Computational Auditory Scene Analysis (CASA '99), pp 78-87, 1999.
- [Goo97] GOODRIDGE, S.G.: *Multimedia Sensor Fusion for Intelligent Camera Control and Human-Computer Interaction*. <http://www.ie.ncsu.edu/kay/msf/abstract.htm>, 1997.
- [HE00] HUANG, Y., J. BENESTY und G.W. ELKO: *Passive Acoustic Source Localization for Video Camera Steering*. Proceedings of ICASSP-2000, Istanbul, Turkey, vol. 2, pp. 909-912, 2000.
- [HS97] HUANG, J., N. OHNISHI und N. SUGIE: *Sound Localization in Reverberant Environment Based on the Model of the Precedence Effect*. IEEE, 1997.
- [Jon00] JONDRALE, F.: *Wahrscheinlichkeitstheorie*. Teubner, 2000.

Literaturverzeichnis

- [KC76] KNAPP, C.H. und G.C. CARTER: *The generalized correlation method for estimation of time delay*. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-24, no. 4, pp. 320–327, 1976.
- [Kro96] KROSCHEL, K.: *Statistische Nachrichtentheorie: Signal- und Mustererkennung, Parameter- und Signalschätzung*. Springer, 1996.
- [LW85] LEHN, J. und H. WEGMANN: *Einführung in die Statistik*. Teubner, 1985.
- [MW96] MERZIGER, G., MÜHLBACH und T. WIRTH: *Formeln + Hilfen zur Höheren Mathematik*. Binomi, 1996.
- [OS97] OMOLOGO, M. und P. SVAIZER: *Use of the cross-power spectrum phase in acoustic event localization*, 1997.
- [SK92] SILVERMAN, H. F. und S. E. KIRTMAN: *A two-stage algorithm for determining talker location from linear microphone array data*. Computer, Speech and Language, 6(2), pp. 129-152, 1992.
- [Spo99] SPORS, S.: *Analyse des Kalman-Filters zur multimodalen Objektverfolgung*, 1999. Studienarbeit, Universität Erlangen-Nürnberg.
- [WB01] WELCH, G. und G. BISHOP: *An Introduction to the Kalman Filter*, 2001. http://www.cs.unc.edu/~welch/kalman/kalman_filter/kalman.html.