

**LEHRSTUHL FÜR MULTIMEDIAKOMMUNIKATION  
UND SIGNALVERARBEITUNG**

UNIVERSITÄT ERLANGEN-NÜRNBERG

Prof. Dr.-Ing. W. Kellermann

Studienarbeit

**Robuste Schätzung von Laufzeitdifferenzen zur  
Quellenlokalisierung mit mehr als zwei Mikrofonen**

von

**Matthias Reichart**

Betreuer: Dipl.-Ing. Heinz Teutsch

Erlangen, Februar 2005



# Erklärung

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

Matthias Reichart  
Gebbertstrae 8  
91052 Erlangen

Erlangen, den 25. Februar 2005



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
<b>2</b>	<b>Schätzung von Laufzeitunterschieden mit Korrelationsmethoden</b>	<b>4</b>
2.1	Schallausbreitungsmodelle . . . . .	4
2.2	Die GCC-Methode . . . . .	5
2.3	Die MCC-Methode — Verallgemeinerung von GCC auf mehrere Kanäle . . . . .	6
2.4	Veranschaulichung des MCC-Verfahrens . . . . .	11
<b>3</b>	<b>Laufzeitdifferenzschätzung mit adaptiven Algorithmen</b>	<b>15</b>
3.1	Signalmodell und grundlegende Annahmen . . . . .	15
3.2	Der Mehrkanal-AED Algorithmus . . . . .	18
3.2.1	Implementierung im Zeitbereich . . . . .	18
3.2.2	Implementierung im Frequenzbereich . . . . .	21
3.3	Schätzung des Laufzeitunterschieds . . . . .	23
<b>4</b>	<b>Untersuchungen und Vergleichstests zur Leistungsfähigkeit der Algorithmen</b>	<b>26</b>
4.1	Testsetup . . . . .	27
4.2	Vergleich von zwei- und mehrkanaligen Korrelationsalgorithmen .	31
4.2.1	Untersuchungen mit generierten Impulsantworten . . . . .	31
4.2.2	Untersuchungen mit gemessenen Impulsantworten . . . . .	37

4.3	Vergleich von zwei- und mehrkanaligen Adaptionalgorithmen . .	42
4.4	Abschließende Bemerkungen . . . . .	51
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>53</b>
<b>A</b>	<b>Matlabfunktionen</b>	<b>55</b>

# Abbildungsverzeichnis

2.1	Laufzeitunterschiedsschätzung mit PHAT . . . . .	7
2.2	Fernfeld . . . . .	9
2.3	Nahfeld . . . . .	10
2.4	Laufzeitunterschiedsschätzung mit MCC . . . . .	11
2.5	Parallelepipet/Parallelogramm . . . . .	12
3.1	Schaltbild MAED . . . . .	20
3.2	Initialisierung der Kanalfilter . . . . .	24
3.3	Schätzung des Laufzeitunterschieds mit dem MAED-Algorithmus	25
4.1	Raumgeometrie . . . . .	27
4.2	Impulsantworten . . . . .	28
4.3	Sprachsignal . . . . .	29
4.4	Geometrie, Multimediarraum . . . . .	30
4.5	GCC vs. MCC, Zunahme der Nachhallzeit, feste Quellenposition .	33
4.6	GCC vs. MCC, Zunahme der Nachhallzeit, gesamter Raum . . . .	35
4.7	GCC vs. MCC, Verschlechterung des SNR, Bild 1 . . . . .	37
4.8	GCC vs. MCC, Verschlechterung des SNR, Bild 2 . . . . .	38
4.9	GCC vs. MCC, N517, Rauschen . . . . .	39
4.10	GCC vs. MCC, N517, Sprache . . . . .	40
4.11	Fernfeldproblematik mit gemessenen Impulsantworten, Bild 1 . . .	42
4.12	Fernfeldproblematik mit gemessenen Impulsantworten, Bild 2 . . .	43
4.13	Fernfeldproblematik mit gemessenen Impulsantworten, Bild 3 . . .	44

4.14 Fernfeldproblematik mit gemessenen Impulsantworten, Bild 4 . . .	45
4.15 Vergleich von AED mit MAED, Multimediarraum mit zugezogenen Vorhängen . . . . .	47
4.16 Vergleich von AED mit MAED, Multimediarraum mit aufgezogenen Vorhängen . . . . .	48
4.17 Vergleich von AED mit MAED, Raum N5.17 . . . . .	48
4.18 Vergleich von AED mit MAED, Garage . . . . .	49



# Tabellenverzeichnis

4.1	Schrittweiten . . . . .	46
4.2	mit (M)AED-Algorithmen geschätzte Laufzeitdifferenzen . . . . .	47

# Notationen

## Allgemeine Notationen

$x$	skalare Größe
$\mathbf{x}$	Vektor, allgemein
$\vec{x}$	Vektor, geometrische Interpretation
$\mathbf{X}$	Matrix
$\mathbf{X}_{L \times P}$	Matrix der Dimension $L \times P$
$x^*$	$x$ konjugiert komplex
$\mathbf{x}^T$	transponierter Vektor
$\mathbf{x}^H$	adjungierter Vektor
$\mathbf{X}^{-1}$	Inverse der Matrix $\mathbf{X}$
$x(n)$	Funktion der kontinuierlichen Variable $n$
$x[k]$	Funktion der diskreten Variable $k$
$\mathcal{FFT}_N\{\cdot\}$	N-Punkte Fast Fourier Transformation
$X(z)$	Z-Transformierte von $x[k]$
$\underline{x}$	Größe im Frequenzbereich
$E\{\cdot\}$	Erwartungwertoperator
$\odot$	Schur-Produkt (elementweise Multiplikation)
$ \cdot $	Betrag
$\ \cdot\ $	Norm
$\nabla$	Gradientenoperator
$\frac{\partial}{\partial x}$	Partielle Ableitung nach $x$

## Symbole und Variablen

$\alpha$	Dämpfungsfaktor
$\beta$	Reflexionsfaktor
$\delta$	Regularisierungsfaktor
dp	Betrag des Direktschallpeaks
$e_{ij}$	Fehlersignal zwischen i-tem und j-tem Kanal
$f_i(\tau)$	relative Verzögerung zwischen 0-tem und l-tem Mikrofon
$f_s$	Abtastfrequenz
$\Phi(f)$	Gewichtungsfunktion
$g[n]$	Impulsantwort
$h$	Kanalimpulsantwort
$h_t$	wahre Kanalimpulsantwort
$H_t$	wahre Kanalparamtermatrix
ir	Impulsantwort (generiert)
$J, J_f$	Kostenfunktion
$\lambda$	Glättungs-/Vergessensfaktor
$L$	Filterlänge, Mikrofonanzahl -1
$m$	Blockzeitindex, Zeitverschiebung in Samples
$\mu, \mu_f$	Schrittweite
$M$	Anzahl der Mikrofone
$N$	Filterlänge
$n$	Zeitindex, Blockzeitindex
$p$	Leistung
$\Psi_{x_1x_2}(\tau)$	Kreuzkorrelationsfunktion zwischen $x_1$ und $x_2$
$\rho$	Korrelationskoeffizient
$\rho_{m,kl}$	Kreuzkorrelationskoeffizient zwischen k-tem und l-tem Mikrofonsignal zum Zeitpunkt m

## Symbole und Variablen

$\tilde{\mathbf{R}}$	Korrelationsmatrix
$\tilde{\mathbf{R}}_m$	Kreuzkorrelationskoeffizientenmatrix
$\mathbf{R}_{ss}$	Autokorrelationsmatrix
$s[n]$	Quellensignal
$t$	Grundlaufzeit
$\tau$	Laufzeitunterschied
$\hat{\tau}$	geschätzter Laufzeitunterschied
$T_{60}$	Nachhallzeit
$w[n]$	Rauschsignal
$x[n]$	Mikrofonsignal

## Abkürzungen

AED	Adaptive Eigenvalue Decomposition (adaptive Eigenwertzerlegung)
ETD	Estimated Time Delay (geschätzter Zeitunterschied)
GCC	Generalized Cross Correlation (allgemeine Kreuzkorrelation)
KKF	Kreuzkorrelationsfunktion
MAED	Multichannel Adaptive Eigenvalue Decomposition (Mehrkanal AED)
MCC	Multichannel Cross Correlation (Mehrkanal Kreuzkorrelation)
MMR	Multimediarraum
PHAT	Phasentransformation
SNR	Signal to Noise Ratio (Signal- zu Geräuschleistungs-verhältnis)

## **Zusammenfassung**

Bis heute werden Laufzeitunterschiede akustischer Signale zwischen zwei Mikrofonen mit der GCC-Methode und dem AED-Algorithmus ermittelt. Diese Arbeit stellt zwei Verfahren aus jüngerer Literatur vor, die Mehrkanalkorrelationsmethode und den Mehrkanal-AED-Algorithmus, die mit Hilfe von mehr als zwei Mikrofonen den Laufzeitunterschied eines akustischen Signals zwischen zwei Kanälen schätzen. Man verpricht sich davon eine größere Robustheit der Algorithmen bei Nachhall. Es werden Vergleichstest durchgeführt, um die Leistungsfähigkeit der mehrkanaligen Verfahren gegenüber der Leistungsfähigkeit der zweikanaligen Verfahren beurteilen zu können.

# Kapitel 1

## Einführung

Bei der Objektlokalisierung mit Hilfe akustischer Signale bedient man sich einem einfachen Prinzip. Durch Ermittlung des Laufzeitunterschieds des von dem Objekt ausgehenden akustischen Signals an einem Sensorenpaar, kann auf die Richtung geschlossen werden, aus der das akustische Signal (zum Beispiel von einem Sprecher) abgestrahlt wurde [3].

Die robuste Schätzung von Laufzeitdifferenzen ist seit einigen Jahrzehnten eine immer noch herausfordernde Aufgabe in der Sprach- und Audiosignalverarbeitung. Im Wesentlichen wurden in der Vergangenheit zwei unterschiedliche Verfahren, die GCC-Methode (engl.: *Generalized Cross Correlation*) [3] und die blinde Quellentrennung mittels AED (engl.: *Adaptive Eigenvalue Decomposition*) [3], entwickelt, die je nach Szenario eine gute oder minder gute Leistungsfähigkeit aufzeigen. Gerade die GCC-Methode in all ihren Varianten stellt eine einfache und effiziente Möglichkeit dar, den Laufzeitunterschied eines an einem Mikrofonpaar ankommenden akustischen Signals zu schätzen. Leider ist dieses Verfahren in Umgebungen mit erhöhtem Nachhall und unter Zunahme von Störgeräuschen fehleranfällig und bietet deshalb in anwendungsnahen Fällen, wie zum Beispiel automatischer Kameraverfolgung eines Sprechers bei einer Videokonferenz, nur eine unzureichende Genauigkeit.

Verfahren zur blinden Quellentrennung beruhen auf adaptiven Algorithmen, die

wesentlich besser mit Nachhall und Störgeräuschen zurecht kommen. Bisher wurden die GCC-Methode und der AED-Algorithmus vorwiegend mit Mikrofonpaaren eingesetzt. Jacob Benesty, Yiteng (Arden) Huang und Jingdong Chen schlagen Methoden vor, die mit Hilfe von mehr als zwei Mikrofonen eine allgemeine Zunahme der Robustheit gegenüber Nachhall und somit eine Verminderung der Störanfälligkeit bei der Laufzeitdifferenzschätzung gewähren sollen [1, 2]. Die vorgestellten Verfahren stellen eine Verallgemeinerung und Erweiterung der Korrelationsmethode und des AED-Algorithmuses von zwei auf  $M$  Kanäle dar.

In dieser Arbeit werden das Mehrkanal-Korrelations-Verfahren (MCC) und das Mehrkanal-AED-Verfahren (MAED) vorgestellt. In Kapitel 2 werden die Grundlagen der GCC-Methode und des MCC-Verfahrens erläutert. Dabei wird unter anderem versucht die Verallgemeinerung der zweikanaligen GCC-Methode auf das MCC-Verfahren anhand einer mathematischen Veranschaulichung zu verdeutlichen.

In Kapitel 3 wird der Mehrkanal-AED-Algorithmus vorgestellt. Ausgehend von der zu Grunde liegenden Signaltheorie werden die Implementierung im Zeit- sowie im Frequenzbereich behandelt.

Im sich darauf anschließenden Kapitel 4 werden in einem Unterabschnitt die zweikanalige GCC-Methode mit der MCC-Methode verglichen und die Leistungsfähigkeit unter ideal simulierten Bedingungen und realistischen Bedingungen untersucht. In einem weiteren Abschnitt wird der zweikanalige AED-Algorithmus mit dem MAED-Algorithmus verglichen. Unter Verwendung der erhaltenen Testergebnisse wird versucht Aussagen über die Leistungsfähigkeit der mehrkanaligen Verfahren zu treffen.

Im letzten Kapitel werden die Ergebnisse zusammengefasst und kritisch beleuchtet. Außerdem wird ein Ausblick auf mögliche weiterführende Untersuchungen gewährt.

An dieser Stelle möchte ich mich bei meinem Betreuer Heinz Teutsch, der mir stets mit Rat und Tat zur Seite stand, bedanken. Vielen Dank auch an Herbert Buchner, der mir wertvolle Tips und Hinweise zur mathematischen Erläuterung der Korrelationsmethoden gegeben hat und ein zuverlässiger Ansprechpartner bei Fragen zu den AED-Algorithmen war.



# Kapitel 2

## Schätzung von Laufzeitunterschieden mit Korrelationsmethoden

### 2.1 Schallausbreitungsmodelle

Beim idealen Schallausbreitungsmodell wird nur der direkte Pfad zwischen Quelle und Senke berücksichtigt [1, 3]:

$$x[n] = \alpha s[n - t] + w[n] \quad (2.1)$$

$x[n]$  ist das Mikrofonsignal,  $\alpha$  ein für den Weg zwischen Quelle und Senke charakteristischer Dämpfungsfaktor,  $s[n]$  ist das Quellensignal, das um die Zeit  $t$  verzögert am Mikrofon ankommt und  $w[n]$  steht für ein sich addierendes zu  $s[n]$  unkorreliertes Rauschen. Dieses Modell berücksichtigt keine Mehrwegeausbreitung und ist daher zur Modellierung realistischer Umgebungen wie z.B. Büroräumen ungeeignet.

Möchte man Schallausbreitung in realistischen Umgebungen modellieren, so

bedient man sich einer allgemeineren Beschreibung, die den Weg des Schalls von Quelle zu Senke in Form einer Impulsantwort  $g[n]$  darstellt [1, 3]:

$$x[n] = s[n] * g[n] + w[n] \quad (2.2)$$

Das Mikrofonsignal setzt sich nun aus dem Faltungsprodukt der Impulsantwort  $g[n]$  und dem Quellensignal  $s[n]$  zusammen, addiert mit einem zum Quellensignal unkorreliertem Rauschen.

## 2.2 Die GCC-Methode

Die GCC-Methode (engl.: Generalized Crosscorrelation Method) ist ein einfaches und effizientes Verfahren um Laufzeitunterschiedsschätzungen zwischen Signalen an einem Mikrofonpaar durchzuführen. Die Kreuzkorrelationsfunktion (KKF) zweier Signale  $x_1(t)$  und  $x_2(t)$  ist definiert durch die Fourierrücktransformation ihres Kreuzleistungsdichtespektrums [3]:

$$\Psi_{x_1x_2}(\tau) = \int_{-\infty}^{+\infty} \Phi(f) S_{x_1x_2}(f) e^{j2\pi f\tau} df \quad (2.3)$$

Dabei ist  $S_{x_1x_2}(f) = E\{X_1(f)X_2(f)\}$  das Kreuzleistungsdichtespektrum der Zeitsignale  $x_1(t)$  und  $x_2(t)$ ,  $X_1(f)$  und  $X_2(f)$  bezeichnen deren Spektren.  $\Phi(f)$  ist eine beliebige Gewichtungsfunktion, oder auch Vorfilter. Im klassischen und auch einfachsten Fall ist  $\Phi(f) = 1$ . Je nach Anwendung sind eine Vielzahl von Gewichtungsfaktoren denkbar. In dieser Arbeit wird ausschließlich die Phasentransformation<sup>1</sup> (PHAT) verwendet. Dazu setzt man  $\Phi(f) = \frac{1}{|S_{x_1x_2}(f)|}$ , man teilt also durch den Betrag des Kreuzleistungsdichtespektrums. Der Vorteil der Phasentransformation besteht darin, dass die KKF einen schärferen Peak aufweist.

---

<sup>1</sup>Im weiteren Verlauf der Arbeit wird zwischen PHAT und GCC-Methode, als dessen allgemeine Form, nicht unterschieden

Dies rührt daher, dass die Phase mehr Information über den Laufzeitunterschied enthält, als der Betrag [3].

Das Prinzip der Zeitunterschiedsschätzung mit Hilfe der GCC-Methode beruht darauf, das Maximum der Kreuzkorrelationsfunktion des ankommenden Signals an Mikrofon 1 ( $x_1$ ) und Mikrofon 2 ( $x_2$ ) zu suchen [3]:

$$\hat{\tau}_{x_1x_2} = \underset{\tau}{\operatorname{argmax}} \Psi_{x_1x_2}(\tau) \quad (2.4)$$

Der Zeitunterschied zwischen  $x_1$  gegenüber  $x_2$  ermittelt sich aus dem Abstand des Peaks zur Mitte der Kreuzkorrelationsfunktion (Abb. 2.1). Da das Verfahren auf dem idealen Schallausbreitungsmodell beruht, welches nur den direkten Pfad von Quelle zu Senke berücksichtigt, ist es bei realen Anwendungen, vor allem in Räumen mit nicht zu vernachlässigenden Nachhallzeiten, sehr fehleranfällig, denn es können weitere Peaks auftreten, die keine eindeutige Bestimmung des Laufzeitunterschieds zulassen.

## 2.3 Die MCC-Methode — Verallgemeinerung von GCC auf mehrere Kanäle

Die Multikanal-Korrelationsmethode (engl.: Multichannel Crosscorrelation Method, MCC-Method) bedient sich zur Berechnung des Laufzeitunterschieds eines ankommenden Signals zwischen zwei Mikrofonen nicht nur der Information aus dem Mikrofonpaar selbst, sondern aus den Signalen eines gesamten Mikrofonarrays mit  $L + 1 \geq 2$  Mikrofonen [1].

Ausgangspunkt für die Erläuterung des MCC-Algorithmus ist das ideale Schallausbreitungsmodell (vgl. Gl. (2.1)), hier dargestellt für  $L + 1$  Kanäle [1]:

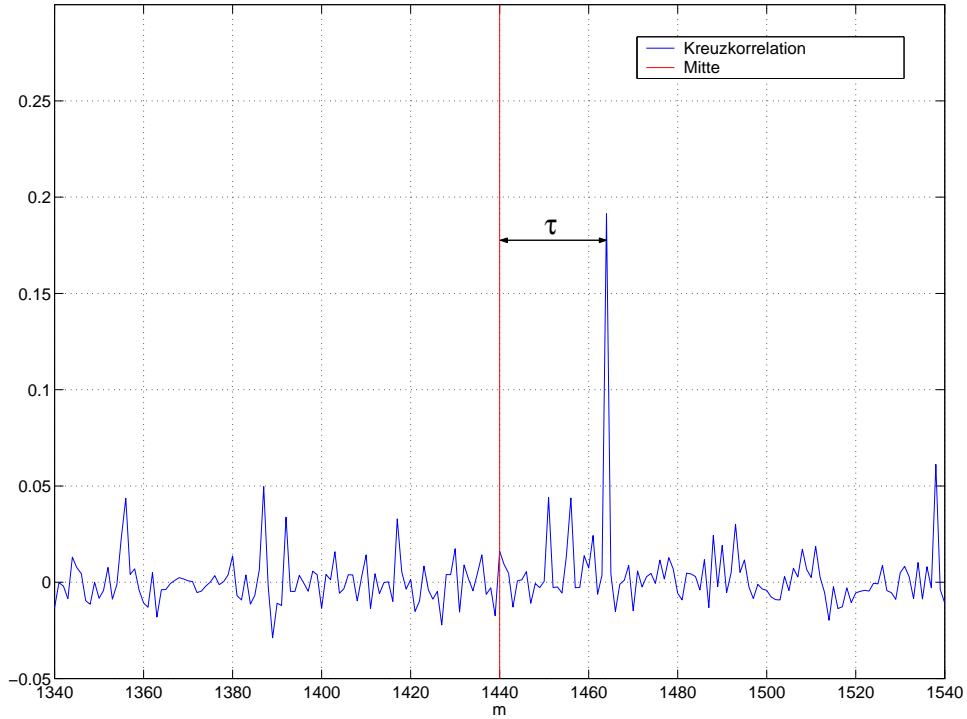


Abbildung 2.1: Auf die Mitte der KKF vergrößerter Ausschnitt einer KKF zwischen zwei Mikrofonen aufgetragen über dem Verschiebungszeitindex  $m$  (in Samples). Der Laufzeitunterschied zwischen den Signalen von Mikrofon 1 und Mikrofon 2 ergibt sich aus dem Abstand des Peaks der KKF von der Mitte der KKF.

$$\begin{bmatrix} x_0[n] \\ x_1[n] \\ x_2[n] \\ \vdots \\ x_L[n] \end{bmatrix} = \begin{bmatrix} \alpha_0 & 0 & 0 & \cdots & 0 \\ 0 & \alpha_1 & 0 & \cdots & 0 \\ 0 & 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \alpha_L \end{bmatrix} \begin{bmatrix} s[n+t] \\ s[n+t+\tau] \\ s[n+t+f_2(\tau)] \\ \vdots \\ s[n+t+f_L(\tau)] \end{bmatrix} + \begin{bmatrix} w_0[n] \\ w_1[n] \\ w_2[n] \\ \vdots \\ w_L[n] \end{bmatrix} \quad (2.5)$$

Hierbei stehen  $x_0$  bis  $x_L$  für die  $L+1$  Mikrofon-signale.  $\alpha_l$  ist ein für jeden Kanal individueller Dämpfungsfaktor,  $t$  steht für die Grundlaufzeit des Schalls zwischen Quelle  $s[n]$  und Mikrofonarray und ist für jeden Kanal gleich.  $\tau$  ist die relative

Verzögerung zwischen Mikrofon 0 und 1,  $f_l(\tau)$  beschreibt die relative Verzögerung zwischen Mikrofon 0 und  $l$ .  $w_l[n]$  ist ein sich am  $l$ -ten Mikrofon addierendes Rauschen, das als unkorreliert zum Quellensignal und zum Rauschen an anderen Mikrofonen angenommen wird. Für ein lineares Mikrofonarray, bei dem der Abstand aufeinanderfolgender Mikrofone konstant ist, ergibt sich für die relative Verzögerung zwischen den Mikrofonensignalen  $x_0$  und  $x_l$  unter Fernfeldannahme (siehe Abb. 2.2) ein einfacher Zusammenhang [1]:

$$f_l(\tau) = l\tau \quad (2.6)$$

Im Nahfeld (siehe Abb. 2.3) hängt  $f_l$  auch von der Position der Quelle ab. Da in dieser Arbeit der Algorithmus ausschließlich zur Schätzung des Laufzeitunterschieds ankommender Signale am Mikrofonarray verwendet wird, um daraus weiterführend auf die Position der Quelle schließen zu können, muss von Fernfeldbetrachtungen ausgegangen werden. Bei Fernfeldannahme geht man davon aus, dass das Quellensignal in Form einer ebenen Welle am Mikrofonarray ankommt. Somit ist der Winkel unter dem die Quelle an den einzelnen Mikrofonen erscheint gleich. Durch  $f_l$  fließt die Information über die Geometrie des Arrays in die Signalbeschreibung mit ein.

Die MCC-Methode kann als Erweiterung der GCC-Methode verstanden werden. Unter Berücksichtigung der relativen Verzögerung zwischen den  $L + 1$  Mikrofonen berechnet man eine sogenannte räumliche Kreuzkorrelationskoeffizientenmatrix (KKK-Matrix) [1]:

$$\tilde{\mathbf{R}}_m = \begin{bmatrix} 1 & \rho_{m,01} & \rho_{m,02} & \cdots & \rho_{m,0L} \\ \rho_{m,01} & 1 & \rho_{m,12} & \cdots & \rho_{m,1L} \\ \rho_{m,02} & \rho_{m,12} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_{m,L-1L} \\ \rho_{m,0L} & \rho_{m,1L} & \cdots & \rho_{m,L-1L} & 1 \end{bmatrix}, \quad (2.7)$$

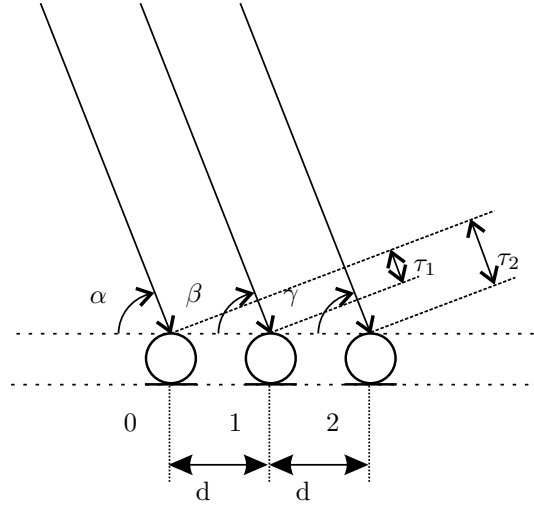


Abbildung 2.2: Fernfeldannahme. Der Abstand der Quelle vom Mikrofonarray ist im Vergleich zum Abstand  $d$  der Mikrofone untereinander so groß, dass die Quelle für jedes Mikrofon unter dem selben Winkel erscheint ( $\alpha = \beta = \gamma$ ). Somit gilt für die Laufzeitdifferenzen:  $\tau_2 = 2 \cdot \tau_1$

mit

$$\rho_{m,kl} = \frac{E\{x_k[n - f_l(m)]x_l[n - f_k(m)]\}}{\sqrt{E\{x_k^2[n]\}E\{x_k^2[n]\}}} \quad (2.8)$$

Die Matrix ist symmetrisch, da  $\rho_{m,kl} = \rho_{m,lk} \forall l, k = 0, 1, \dots, L$ . Durch einsetzen von  $l = k$  in Gleichung (2.8) ist ebenso sofort ersichtlich, dass  $\rho_{m,kl} = 1 \forall l = k$ . Desweiteren gilt [1]:

$$0 < \det(\tilde{\mathbf{R}}_m) \leq 1 \quad (2.9)$$

Wenn die Korrelationen zwischen den unterschiedlichen Mikrofonen gering sind, haben die Nebendiagonalelemente ( $\rho_{m,kl}$  mit  $l, k = 0, 1, \dots, L$  und  $l \neq k$ ) einen kleinen Betrag nahe 0. Somit ergibt sich für die Determinante ein großer Wert nahe 1. Entsprechend gilt diese Überlegung für den umgekehrten Fall. Die Determinan-

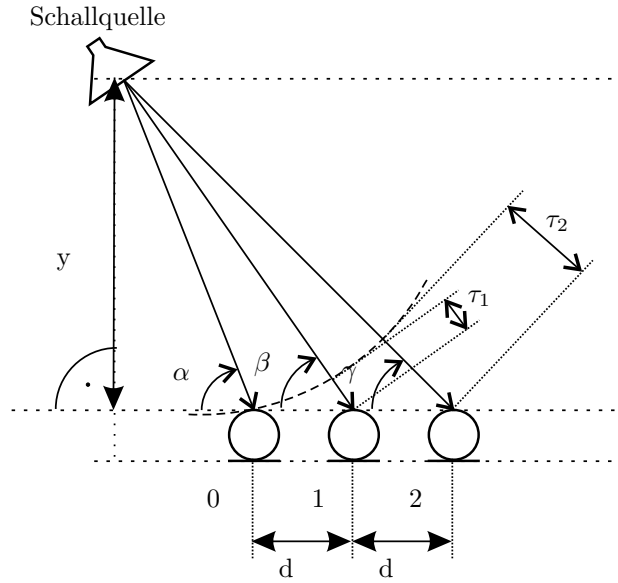


Abbildung 2.3: Nahfeldsituation. Die Position der Quelle ist für die Ermittlung der Laufzeitverzögerungen  $\tau_1$  und  $\tau_2$  notwendig. Die Winkel  $\alpha$ ,  $\beta$  und  $\gamma$  unter denen die Quelle an den Mikrofonen 0 bis 2 erscheint sind nicht gleich.

te der KKK-Matrix wird klein, wenn die Nebendiagonalelemente großen Betrag haben. Somit ist klar, dass als Maß für den Laufzeitunterschied zwischen den Signalen  $x_0$  und  $x_1$  das Minimum der Determinante der KKK-Matrix bezüglich  $m$  (= angenommene zeitliche Verschiebung) dient [1]:

$$\hat{\tau} = \underset{m}{\operatorname{argmin}} [\det(\tilde{\mathbf{R}}_m)] \quad (2.10)$$

Die Bestimmung des Laufzeitunterschieds mit dem das Quellensignal an den Mikrofonen 0 und 1 (Signale  $x_0$  und  $x_1$ ) ankommt, ist analog zur GCC-Methode (vgl. Abb 2.4).

Die Laufzeitunterschiedsschätzung mit der MCC-Methode soll nach [1] eine größere Robustheit in Räumen mit höherer Nachhallzeit haben als die GCC-Methode. Desweiteren soll sie eine geringere Fehleranfälligkeit bei schlechterem Signal- zu

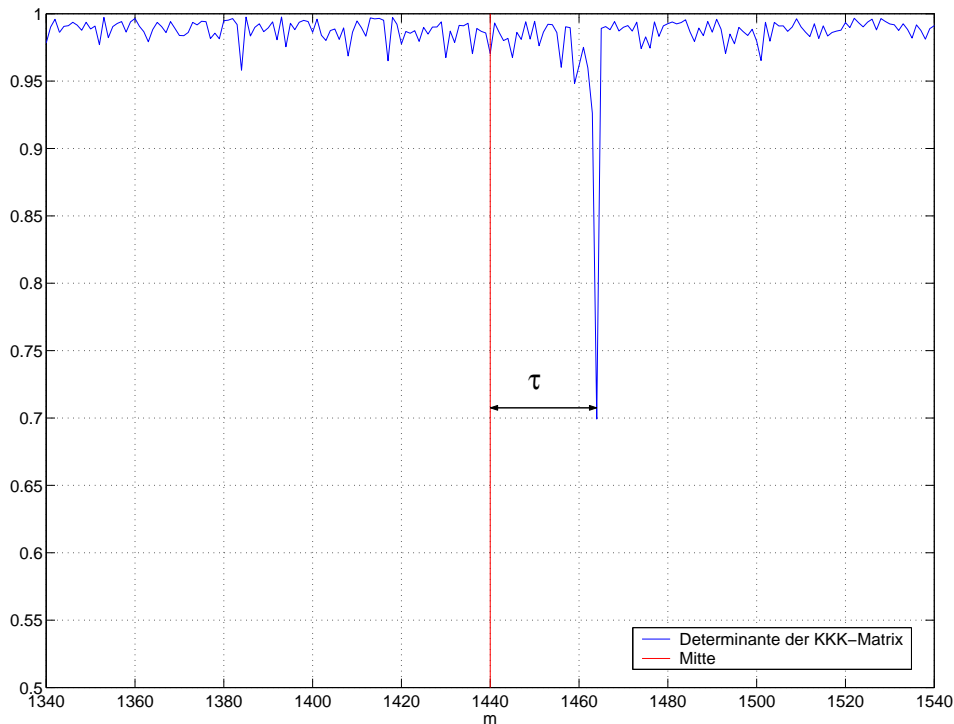


Abbildung 2.4: *Determinante einer  $5 \times 5$  KKK-Matrix aufgetragen über der zeitlichen Verschiebung  $m$  (in Samples). Die Bestimmung des Zeitunterschieds verläuft analog zur GCC-Methode*

Geräuschleistungsverhältnis aufweisen. Begründet wird dies damit, dass durch die Verwendung weiterer Mikrofone zusätzliche Information in die Berechnung miteinfließt. In Kapitel 4 wird die Leistungsfähigkeit der MCC-Methode mit mehr als zwei Mikrofonen untersucht und mit der GCC-Methode (PHAT) verglichen. Es wird an dieser Stelle darauf hingewiesen, dass das MCC-Verfahren mit zwei Mikrofonen nicht der PHAT entspricht.

## 2.4 Veranschaulichung des MCC-Verfahrens

Dieser Abschnitt dient der Erklärung des mathematischen Hintergrundes der MCC-Methode und soll die Analogie von GCC und MCC verdeutlichen. Basis dieser Erläuterung ist [4].



Die Kovarianz zweier Signale kann analog zum Skalarprodukt [6] zweier Vektoren  $\vec{x}_1, \vec{x}_2$  gesehen werden:

$$\overbrace{\vec{x}_1 \cdot \vec{x}_2}^{\text{analog zu Kovarianz}} = \|\vec{x}_1\| \cdot \|\vec{x}_2\| \cdot \underbrace{\cos \phi}_{-1 \leq \rho \leq 1} \quad (2.11)$$

Zwei Vektoren sind zueinander parallel, wenn der Winkel  $\phi$  den sie zusammen einschließen gleich 0 ist. Betrachtet man nun in Gleichung (2.11) die Vektoren  $\vec{x}_1$  und  $\vec{x}_2$  als Signalvektoren, so entspräche  $\cos \phi (= \rho)$  deren Korrelationskoeffizient. Sind die Signale  $\vec{x}_1$  und  $\vec{x}_2$  nun parallel bzw. stark korreliert, so ist  $\rho$  maximal. Damit ist erklärt warum das Maximum der Kreuzkorrelationsfunktion zweier Mikrofonsignale ein Maß für ihren zeitlichen Versatz ist. Man kann sich diesen Zusammenhang weiter veranschaulichen. Die Fläche, die zwei Vektoren miteinander aufspannen ist minimal, wenn der Winkel zwischen ihnen 0 ist (vgl. Abb. 2.5), sie also parallel sind.

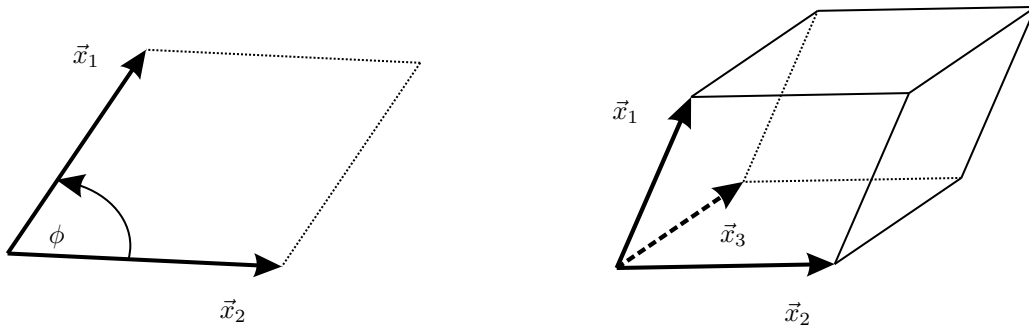


Abbildung 2.5: *links: Parallelogramm, aufgespannt durch zwei Vektoren; rechts: Parallelepiped, aufgespannt durch drei Vektoren*

Es werden nun eine Reihe von Umformungen vorgenommen [4, 6]:

$$(\vec{x}_1 \cdot \vec{x}_2)^2 = \|\vec{x}_1\|^2 \|\vec{x}_2\|^2 \rho^2 \quad (2.12)$$

$$\rho^2 = \frac{(\vec{x}_1 \cdot \vec{x}_2)^2}{\|\vec{x}_1\|^2 \|\vec{x}_2\|^2} = \frac{\|\vec{x}_1\|^2 \|\vec{x}_2\|^2 - (\|\vec{x}_1\|^2 \|\vec{x}_2\|^2 - (\vec{x}_1 \cdot \vec{x}_2)^2)}{\|\vec{x}_1\|^2 \|\vec{x}_2\|^2} \quad (2.13)$$

Mit  $\|\vec{x}_1\|^2 = \vec{x}_1^T \vec{x}_1$  und  $(\vec{x}_1 \cdot \vec{x}_2) = \vec{x}_1^T \vec{x}_2$  kann der Zähler umformuliert werden:

$$(\vec{x}_1^T \vec{x}_1)(\vec{x}_2^T \vec{x}_2) - (\vec{x}_1^T \vec{x}_2)(\vec{x}_2^T \vec{x}_1) = \det \begin{bmatrix} \vec{x}_1^T \vec{x}_1 & \vec{x}_2^T \vec{x}_1 \\ \vec{x}_1^T \vec{x}_2 & \vec{x}_2^T \vec{x}_2 \end{bmatrix} =: \det(\mathbf{R}),$$

so dass man schließlich erhält:

$$\rho^2 = 1 - \frac{\det(\mathbf{R})}{\|\vec{x}_1\|^2 \|\vec{x}_2\|^2} \quad (2.14)$$

Gleichung (2.14) kann nun für n Dimensionen mit Vektoren  $\vec{x}_i$  ( $i = 1, 2, \dots, n$ ) verallgemeinert werden:

$$\rho^2 = 1 - \frac{\det(\mathbf{R})}{\prod_i \|\vec{x}_i\|^2} \quad (2.15)$$

mit  $\mathbf{R} = \begin{bmatrix} \vec{x}_1^T \vec{x}_1 & \vec{x}_1^T \vec{x}_2 & \cdots & \vec{x}_1^T \vec{x}_n \\ \vec{x}_2^T \vec{x}_1 & \vec{x}_2^T \vec{x}_2 & \cdots & \vec{x}_2^T \vec{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{x}_n^T \vec{x}_1 & \vec{x}_n^T \vec{x}_2 & \cdots & \vec{x}_n^T \vec{x}_n \end{bmatrix}$

Die Elemente der Matrix  $\mathbf{R}$  sind Skalarprodukte. Die Analogie von Skalarprodukt und Kovarianz wurde gezeigt und so kann Matrix  $\mathbf{R}$  als eine Korrelationsmatrix (ähnlich zu  $\tilde{\mathbf{R}}_m$  in Gl. (2.7)) gesehen werden. Jeder Spaltenvektor enthält als Elemente sämtliche Korrelationen bzw. Skalarprodukte aller n Vektoren  $\vec{x}_i$ , während  $\rho^2$  ein Maß für die Gesamtkorrelation darstellt.  $\rho^2$  ist nun maximal, wenn  $\det(\mathbf{R})$  minimal ist. An dieser Stelle lässt sich die Analogie zu Gleichung (2.10) feststellen. Im dreidimensionalen Fall lässt sich diese Tatsache dahingehend veranschaulichen, dass man die drei Spalten der  $3 \times 3$  Matrix als Vektoren ansieht, die ein Parallelepipid aufspannen. Das Volumen eines Parallelepipeds wird durch die

Determinante der Vektoren die es aufspannen beschrieben. So wie im zweidimensionalen Fall die Fläche des Parallelogramms, das von zwei maximal korrelierten Vektoren aufgespannt wird, minimal ist, so ist im dreidimensionalen Fall das Volumen des Parallelepipedes minimal, wenn die Vektoren maximal korreliert sind. Auf das Korrelationsproblem übertragen bedeutet dies, wie im Abschnitt 2.3 bereits erläutert, dass die Gesamtkorrelation groß ist, wenn die Determinante klein ist. Bei vollständiger Dekorrelirtheit aller einzelner Mikrofonsignale wären die Spaltenvektoren linear unabhängig. Bis auf ein Element, nämlich  $\vec{x}_i^T \vec{x}_i$ , was aus der Sicht der Signalverarbeitung der Autokorrelation (Selbstähnlichkeit) des Signals  $\vec{x}_i$  entspricht, wären alle anderen Elemente der Spaltenvektoren gleich 0. In diesem Fall hat die Matrix vollen Rang, die Determinante ist maximal und im Dreidimensionalen betrachtet, ist das Volumen des Parallelepipedes (in diesem Fall ein Quader) ebenso maximal. Um so mehr die Vektoren in der mathematischen Anschauung aufeinander senkrecht stehen, um so weniger ähnlich sind sich die Mikrofonsignale.

# Kapitel 3

## Laufzeitdifferenzschätzung mit adaptiven Algorithmen

### 3.1 Signalmodell und grundlegende Annahmen

Im Zuge der Untersuchungen von Methoden zur Laufzeitdifferenzschätzung akustischer Signale hat sich ein weiteres Verfahren etabliert, das das Problem aus einer anderen Perspektive beleuchtet, der so genannte AED-Algorithmus (engl.: *adaptive eigenvalue decomposition*) [1, 2, 7]. Motiviert durch die unbefriedigende Leistungsfähigkeit der GCC-Methode in nicht nachhallarmer Umgebung, was letzten Endes auf die Verwendung des idealen Schallausbreitungsmodells zurück zu führen ist, versucht das AED-Verfahren, basierend auf dem allgemeinen Schallausbreitungsmodell, blind (d.h. ohne Kenntnis über das Quellensignal selbst) die Kanäle zwischen Quelle und den einzelnen Mikrofonen zu identifizieren. Dabei wird jeder Kanal durch eine FIR-Impulsantwort beschrieben. Die Unterschiede in der Position ausgeprägter Extrema in den geschätzten Impulsantworten lassen Rückschlüsse auf die Laufzeitdifferenz zwischen den Mikrofonen zu. Der AED-Algorithmus verwendet zur Kanalschätzung zwei Kanäle bzw. zwei Mikrofonsignale. Die Fehleranfälligkeit des Algorithmus ist vor allem dann sehr hoch, wenn

die Nullstellen der Kanalübertragungsfunktionen nahe beieinander liegen [1, 2]. Im Falle dessen, dass die Übertragungsfunktionen sogar gemeinsame Nullstellen haben, sind die Kanäle nicht identifizierbar. Da in der Realität die Impulsantworten sehr lang sein können (mehrere tausend Koeffizienten), ist die Wahrscheinlichkeit von gemeinsamen Nullstellen nicht gering, und die Problematik beinhaltet ein nicht zu vernachlässigendes Fehlerrisiko. Daher wird nach [1, 2] ein Mehrkanal-AED-Algorithmus vorgeschlagen, der eine Verallgemeinerung des AED-Algorithmus darstellt. Verwendet man mehrere Kanäle, so ist es weniger wahrscheinlich, dass die Kanalübertragungsfunktionen gemeinsame Nullstellen haben.

Im Folgenden werden grundlegende Annahmen und Prinzipien erläutert, die der Funktionsweise des Mehrkanal-AED-Algorithmus zu Grunde liegen.

Durch Erweiterung des allgemeinen Schallausbreitungsmodells (Gl. (2.2)) auf  $M$  Kanäle bzw. Mikrofone und Umformulierung in Matrixschreibweise, erhält man folgendes lineares SIMO (single-input multiple-output) FIR System [1, 2]:

$$\mathbf{x}_i(n) = \mathbf{H}_{t,i} \cdot \mathbf{s}(n) + \mathbf{w}_i, \quad (3.1)$$

$$\text{mit } \mathbf{x}_i(n) = [x_i(n) \ x_i(n-1) \ \cdots \ x_i(n-N+1)]^T,$$

$$\mathbf{H}_{t,i} = \begin{bmatrix} h_{t,i,0} & h_{t,i,1} & \cdots & h_{t,i,N-1} & 0 & \cdots & 0 \\ 0 & h_{t,i,0} & \cdots & h_{t,i,N-2} & h_{t,i,N-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{t,i,0} & h_{t,i,1} & \cdots & h_{t,i,N-1} \end{bmatrix},$$

$$\mathbf{s}(n) = [s(n) \ s(n-1) \ \cdots \ s(n-L+1) \ \cdots \ s(n-2N+2)]^T,$$

$$\mathbf{w}_i(n) = [w_i(n) \ w_i(n-1) \ \cdots \ w_i(n-N+1)]^T$$

Dabei steht  $N$  für die Länge der längsten Impulsantwort, der Index  $t$  steht für “true“ und soll verdeutlichen, dass es sich hier um die wahre Impulsantwort handelt,  $s(n)$  ist das Quellensignal und  $w(n)$  steht wieder für unkorreliertes Rauschen,

was sich an jedem Mikrofon hinzuaddiert. Die Kanalparametermatrix  $\mathbf{H}_{\mathbf{t},i}$  hat die Dimension  $L \times (2L - 1)$  und kann aus der Kanalimpulsantwort zusammengesetzt werden:

$$\mathbf{h}_{\mathbf{t},i} = [h_{t,i,0} \ h_{t,i,1} \ \cdots \ h_{t,i,L-1}]^T \quad (3.2)$$

Desweiteren werden nach [2] zwei Annahmen getroffen.

1. Die Kanalübertragungsfunktionen  $\mathbf{H}_{\mathbf{t},i}(z)$  ( $i = 1, 2, \dots, M$ ) haben keine gemeinsamen Nullstellen.
2. Die Autokorrelationsmatrix des Quellensignals  $\mathbf{R}_{\mathbf{ss}} = E\{s(n)s^T(n)\}$  hat vollen Rang. Das bedeutet die einzelnen Spaltenvektoren sind linear unabhängig.

Außerdem geht man davon aus, dass der Peak in der Impulsantwort der durch den Direktschall (direkter Weg des Schalls von Quelle zu Mikrofon) hervorgerufen wird, dominant ist. Allerdings kann es in stark halligen Umgebungen konstruktive Überlagerungen von Reflexionen, die von der Mehrwegeausbreitung des Schalls herrühren, geben, die weitere Peaks in der Kanalimpulsantwort erzeugen, die stärker gewichtet sind, als der Direktschallpeak. Für die Schätzung der Laufzeitdifferenz in Samples gilt daher [2]:

$$\tau_i = \min\{\arg\max_l^q |h_{t,i,l}|\}, \text{ mit } i = 1, 2, \dots, M, \quad q = 1, 2, \dots, Q \quad (3.3)$$

Hier bedeutet  $\max^q$  das  $q$ -te Maximum von insgesamt  $Q$  Maxima von  $|h_{t,i,l}|$ . Somit ist  $\tau_i$  die kleinste Verzögerungszeit der  $Q$  größten Elemente. Für den Laufzeitunterschied zwischen dem  $i$ -ten und dem  $j$ -ten Kanal ergibt sich demnach [2]:

$$\tau_{ij} = \tau_i - \tau_j \quad i = 1, 2, \dots, M \quad (3.4)$$

Es ist klar, dass für die Aufgabe der Laufzeitdifferenzschätzung eine genaue Identifikation der Kanalimpulsantworten nicht nötig ist, da nur die Spitzenwerte interessieren, die den Direktschall repräsentieren.

## 3.2 Der Mehrkanal-AED Algorithmus

In den nun folgenden Abschnitten wird darauf verzichtet den AED-Algorithmus für zwei Kanäle zu erläutern. Die Erklärung und Herleitung des Mehrkanal-AED-Algorithmus beinhaltet den zweikanaligen AED-Algorithmus vollständig, wenn man sämtliche Überlegungen für nur zwei Kanäle nachvollzieht.

### 3.2.1 Implementierung im Zeitbereich

Zuerst wird der Mehrkanal-AED-Algorithmus im Zeitbereich erläutert. Ausgangspunkt ist der Zusammenhang zwischen den  $i$ -ten und  $j$ -ten Kanalausgängen [1, 2, 7]:

$$x_i(n) * h_{t,j} = s(n) * h_{t,i} * h_{t,j} = x_j(n) * h_{t,i}, \quad i, j = 1, 2, \dots, M \quad (3.5)$$

und in Matrixschreibweise:

$$\mathbf{x}_i^T(n) \mathbf{h}_{t,j} = \mathbf{x}_j^T(n) \mathbf{h}_{t,i}, \quad i, j = 1, 2, \dots, M, \quad i \neq j \quad (3.6)$$

Gleichung (3.6) hat dann Gültigkeit, wenn die Impulsantworten ideal geschätzt werden und keine Störung durch Rauschen vorliegt. In der Realität tritt dieser Fall allerdings nie ein. Deswegen definiert man ein *a priori* Fehlersignal, das die Abweichung von der wahren, zur geschätzten Impulsantwort beschreibt [1, 2]:

$$e_{ij}(n+1) = \frac{\mathbf{x}_i^T(n+1)\mathbf{h}_j(n) - \mathbf{x}_j^T(n+1)\mathbf{h}_i(n)}{\|\mathbf{h}(n)\|}, \quad i, j = 1, 2, \dots, M \quad (3.7)$$

$\mathbf{h}(n)$  setzt sich aus den einzelnen Kanalimpulsantworten zusammen:

$$\mathbf{h}(n) = [\mathbf{h}_1^T(n) \mathbf{h}_2^T(n), \dots, \mathbf{h}_M^T(n)]^T \quad (3.8)$$

Die Summe über der Leistung aller einzelner Fehlersignale dient als Kostenfunktion  $J(n+1)$ :

$$J(n+1) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M e_{ij}^2(n+1) \quad (3.9)$$

Ziel des adaptiven Algorithmus muss es nun sein, die Impulsantworten  $\mathbf{h}_i$  ( $i = 1, 2, \dots, M$ ) so gut zu schätzen, dass die Kostenfunktion  $J(n+1)$  und damit die Fehlersignale  $e_{ij}(n+1)$  minimal werden. Abbildung 3.1 veranschaulicht das Prinzip des Algorithmus.

Der hier vorgestellte Mehrkanal-AED-Algorithmus bedient sich dem LMS-Ansatz (engl.: *Least Mean Square*), das heißt der Algorithmus berechnet den neuen Schätzwert in dem er vom vorhergehenden Schätzwert den mit einem positiven Schrittweitefaktor  $\mu$  gewichteten Gradienten der Kostenfunktion subtrahiert:

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \mu \nabla J(n+1) \quad (3.10)$$

Der Gradient von  $J(n+1)$  berechnet sich nach [2] zu:

$$\nabla J(n+1) = \frac{\partial J(n+1)}{\partial \mathbf{h}(n)} = \frac{2[\tilde{\mathbf{R}}(n+1)\mathbf{h}(n) - J(n+1)\mathbf{h}(n)]}{\|\mathbf{h}(n)\|^2} \quad (3.11)$$



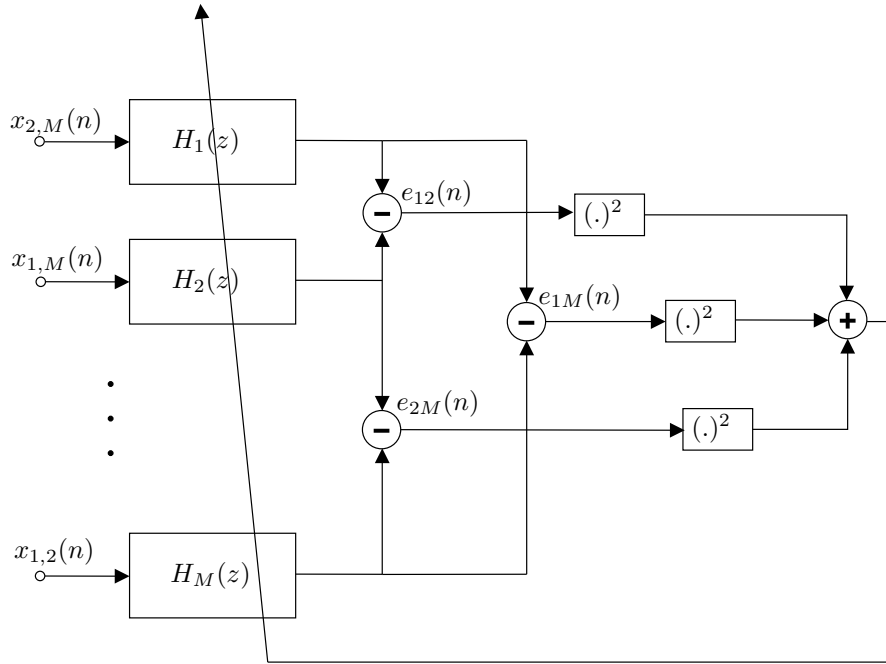


Abbildung 3.1: Schaltbild des Mehrkanal-Adaptions-Prinzips. Die Summe aller Fehlerleistungen  $e_{ij}^2$  ( $i = 1, \dots, M - 1; j = 2, \dots, M; j > i$ ) ist ausschlaggebend für die Adaption

wobei

$$\tilde{\mathbf{R}}(n+1) = \begin{bmatrix} \sum_{i \neq 1} \tilde{\mathbf{R}}_{x_i x_i}(n+1) & -\tilde{\mathbf{R}}_{x_2 x_1}(n+1) & \cdots & -\tilde{\mathbf{R}}_{x_M x_1}(n+1) \\ -\tilde{\mathbf{R}}_{x_1 x_2}(n+1) & \sum_{i \neq 2} \tilde{\mathbf{R}}(n+1) & \cdots & -\tilde{\mathbf{R}}_{x_M x_2}(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{\mathbf{R}}_{x_1 x_M}(n+1) & \tilde{\mathbf{R}}_{x_2 x_M}(n+1) & \cdots & \sum_{i \neq M} \tilde{\mathbf{R}}_{x_i x_i}(n+1) \end{bmatrix} \quad (3.12)$$

und

$$\tilde{\mathbf{R}}_{x_i x_j}(n+1) = \mathbf{x}_i(n+1)\mathbf{x}_j^T(n+1), \quad i, j, \dots, M \quad (3.13)$$

Durch Normalisierung des Modellfilters nach jedem Durchlauf gelangt man schließlich zum Mehrkanal-LMS Algorithmus [1, 2]:

$$\mathbf{h}(n+1) = \frac{\mathbf{h}(n) - 2\mu[\tilde{\mathbf{R}}(n+1)\mathbf{h}(n) - J(n+1)\mathbf{h}(n)]}{\|\mathbf{h}(n) - 2\mu[\tilde{\mathbf{R}}(n+1)\mathbf{h}(n) - J(n+1)\mathbf{h}(n)]\|} \quad (3.14)$$

### 3.2.2 Implementierung im Frequenzbereich

Um eine größere Effizienz zu erzielen kann der Algorithmus vollständig im Frequenzbereich implementiert werden. In diesem Abschnitt werden die hierzu nötigen Schritte vorgestellt.

Analog zur Implementierung im Zeitbereich, definiert man ein Fehlersignal im Frequenzbereich [2]:

$$\underline{\mathbf{e}}_{ij}(m+1) = \underline{\mathbf{y}}_{ij}(m+1) - \underline{\mathbf{y}}_{ij}(m+1), \quad (3.15)$$

wobei

$$\begin{aligned} \underline{\mathbf{y}}_{ij}(m+1) &= \underline{\mathbf{x}}_i(m+1) \odot \underline{\mathbf{h}}_j(m+1) \\ &= \mathcal{FFT}\{\mathbf{x}_i\} \odot \mathcal{FFT}_{2L}\{\mathbf{h}_j\} \end{aligned} \quad (3.16)$$

$$\text{und } \mathbf{x}_i(m+1)_{2L \times 1} = [x_i(mL) x_i(mL+1) \cdots x_i(mL+2L-1)]^T \quad (3.17)$$

Ein Unterstrich kennzeichnet, dass die Größe im Frequenzbereich zu sehen ist.  $m$  ist der Blockzeitindex und so indiziert  $(m+1)$ , dass es sich um den aktuellen Zeitpunkt handelt und  $m$  steht für den Zeitschritt davor.  $\mathcal{FFT}_N\{\cdot\}$  zeigt an, dass es sich um eine  $N$ -Punkt Fouriertransformation handelt und das Symbol  $\odot$

bedeutet elementweise Multiplikation (Schur-Produkt). Entsprechend ergibt sich für die Kostenfunktion  $J_f(m+1)$  im Frequenzbereich [2]:

$$J_f(m+1) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M \mathbf{e}_{ij}^H(m+1) \mathbf{e}_{ij}(m+1) \quad (3.18)$$

Das Filterupdate berechnet sich analog zum Zeitbereich:

$$\underline{\mathbf{h}}_k(m+1) = \underline{\mathbf{h}}_k(m) - \mu_f \frac{\partial J_f(m+1)}{\partial \underline{\mathbf{h}}_k^*(\mathbf{m})}, \quad k = 1, 2, \dots, M \quad (3.19)$$

Dabei gilt nach [2]:

$$\frac{\partial J_f(m+1)}{\partial \underline{\mathbf{h}}_k^*(\mathbf{m})} = \sum_{i=1}^M \mathbf{x}_i^*(m+1) \odot \mathbf{e}_{ik}(m+1), \quad (3.20)$$

so dass sich schließlich für den Mehrkanal-Frequenzbereichs-LMS-Algorithmus (MCFLMS, engl.: *Multichannel Frequency Domain LMS*) ergibt:

$$\underline{\mathbf{h}}_k(m+1) = \underline{\mathbf{h}}_k(m) - \mu_f \sum_{i=1}^M [\mathbf{x}_i^*(m+1) \odot \mathbf{e}_{ik}(m+1)], \quad k = 1, 2, \dots, M \quad (3.21)$$

Um eine schnellere Konvergenz des Algorithmus zu erzielen, wird mit Hilfe der Newtonschen Methode die normalisierte Form des MCFLMS - Algorithmus (NMCFLMS) erzeugt [2]:

$$\underline{\mathbf{h}}_k(m+1) = \underline{\mathbf{h}}_k(m) - \mu_f \sum_{i=1}^M [\mathbf{x}_i^*(m+1) \odot \mathbf{e}_{ik}(m+1) \odot \underline{\mathbf{p}}_k^{-1}(m+1)], \quad k = 1, 2, \dots, M \quad (3.22)$$

$\underline{\mathbf{p}}_k(m+1)$  entspricht dem Leistungsdichtespektrum. Sprachsignale sind weder weiß noch stationär. Somit ändert sich die Signalleistung über der Zeit mehr oder

weniger stark. Um die Schwankungen etwas auszugleichen wird das Leistungsdichtespektrum im Algorithmus mit einem Vergessensfaktor  $\lambda$  geglättet [2]:

$$\underline{\mathbf{p}}_{\mathbf{k}}(m+1) = \lambda \underline{\mathbf{p}}_{\mathbf{k}}(m) + (1-\lambda) \sum_{i=1, i \neq k}^M [\underline{\mathbf{x}}_{\mathbf{i}}^{\mathbf{H}}(m+1) \underline{\mathbf{x}}_{\mathbf{i}}(m+1)] \quad (3.23)$$

Für  $\lambda$  wählt man einen Wert zwischen 0 und 1. Um zu vermeiden, dass für sehr kleine Leistungen (z.B. bei Sprachpausen) in Gl. (3.22) der Ausdruck  $\underline{\mathbf{p}}_{\mathbf{k}}^{-1}(m+1)$  nicht gegen Unendlich strebt wird ein kleiner Regularisierungsfaktor  $\delta$  eingeführt:

$$\underline{\mathbf{p}}_{\mathbf{k}}^{-1}(m+1) = [\underline{\mathbf{p}}_{\mathbf{k}}(m+1) + \delta]^{-1} \quad (3.24)$$

### 3.3 Schätzung des Laufzeitunterschieds

An dieser Stelle soll kurz erläutert werden wie die Laufzeitunterschiede zwischen einzelnen Kanälen mit dem adaptiven Algorithmus geschätzt werden. Zunächst müssen die  $M$  Kanalfilter der Länge  $N$  vorinitialisiert werden:

$$\mathbf{h}_1 = [\underbrace{0 \cdots 0}_{L/2-1} 1 \underbrace{0 \cdots 0}_{L/2}], \quad \mathbf{h}_i = \mathbf{0}, \quad i = 2, 3, \dots, M \quad (3.25)$$

Mit dieser Initialisierung bleibt der Peak im ersten Kanalfilter während der Adaption dominant. Nach einigen Adaptionsschritten bilden sich in den anderen Kanälen ebenso deutliche Extremalstellen ab, die den Direktschall der jeweiligen zu identifizierenden Kanalimpulsantwort repräsentieren. Dadurch, dass der Peak in der Mitte plaziert wurde, können positive wie negative Laufzeitunterschiede ermittelt werden [2]. Abbildung 3.2 zeigt beispielhaft eine Initalisierung für ein dreikanaliges System.

Aus der Differenz der Lage der Extremalstellen der geschätzten Kanalfilter, kann

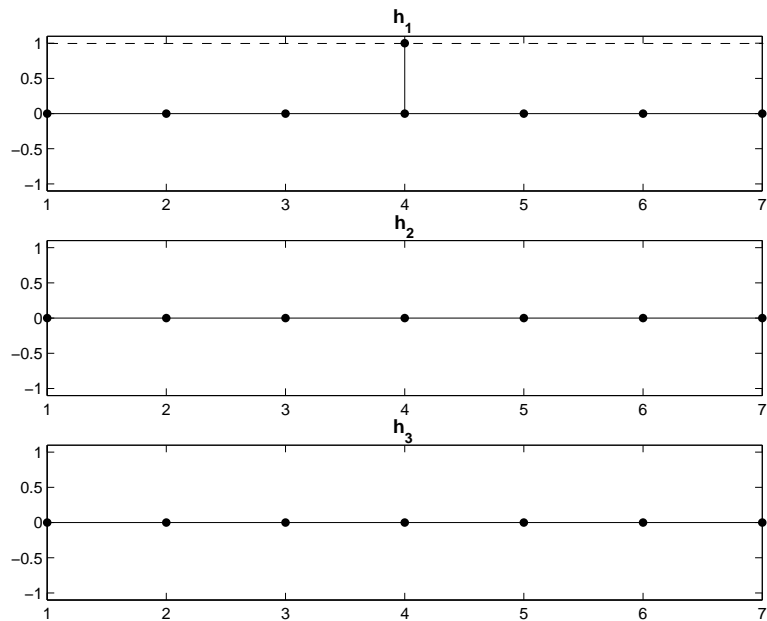


Abbildung 3.2: *Initialisierung der Kanalfilter für ein dreikanaliges System ( $N = 7$ )*

auf den Laufzeitunterschied zwischen den Kanälen geschlossen werden (vgl. Gleichung (3.3)). Beispielhaft ist dies in Abbildung 3.3 für ein System mit drei Kanälen dargestellt. Die Abbildung zeigt keine konkrete Anwendung und dient nur der Veranschaulichung. Es wurde künstlich ein Fall generiert, in dem der Laufzeitunterschied der Kanäle 30 Samples (Filtertaps) beträgt. Die verwendete Filterlänge ist hier 512, als Testsignal diente eine Rauschsequenz von einer Sekunde. Die Abtastrate ist 48kHz und das Bild zeigt den Stand der Adaption nach 92 Adaptionsschritten.

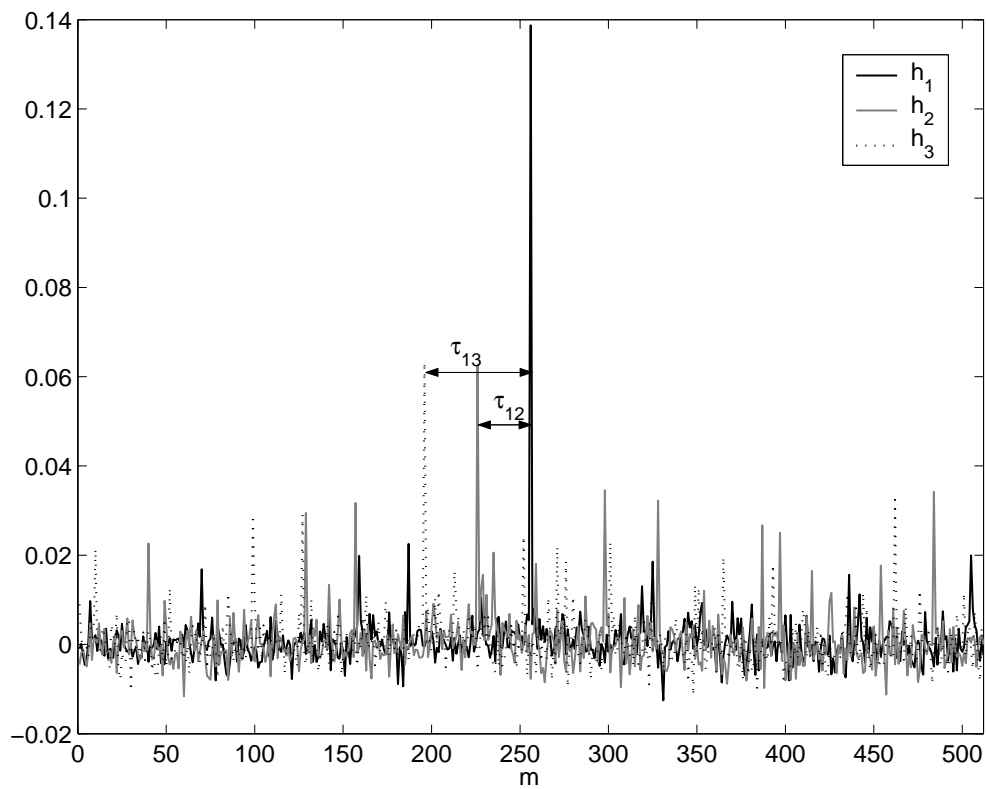


Abbildung 3.3: Schätzung des Laufzeitunterschieds mit dem MAED-Algorithmus. Laufzeitunterschied zwischen Kanal 1 und 2:  $\tau_{12}$ ; Laufzeitunterschied zwischen Kanal 1 und 3:  $\tau_{13}$

# Kapitel 4

## Untersuchungen und Vergleichstests zur Leistungsfähigkeit der Algorithmen

Dieses Kapitel stellt die Untersuchungen vor, die an den in dieser Arbeit vorgestellten Algorithmen, gemacht wurden. Im ersten Abschnitt (4.1) wird das Testsetup erörtert, das zur Durchführung der Simulationen verwendet wurde. Anschließend wird im Abschnitt 4.2 die GCC-Methode<sup>1</sup> mit der MCC-Methode verglichen und die Unterschiede herausgestellt. Abschnitt 4.3 beschäftigt sich mit den adaptiven Algorithmen und zeigt Stärken und Schwachstellen des zwei- und mehrkanaligen Verfahrens auf.

---

<sup>1</sup>es wird stets die PHAT betrachtet

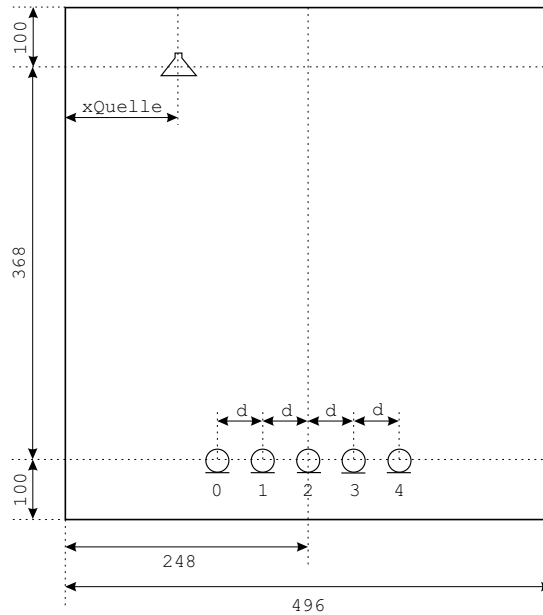


Abbildung 4.1: Für die Simulationen verwendetes Szenario. Alle Maße sind in cm angegeben. Mikrofonarray und Quelle befinden sich in einer Höhe von 106cm über dem Boden.

## 4.1 Testsetup

Die Algorithmen sollen in erster Linie bezüglich ihrer Robustheit bei zunehmendem Nachhall, sowie sich verschlechterndem Signal- zu Geräuschleistungsverhältnis (SNR, engl.: *Signal to Noise Ratio*) untersucht werden. Hierzu werden mit Hilfe der Spiegelquellenmethode [5] Impulsantworten erzeugt, die die Kanäle von der Quelle zu den einzelnen Mikrofonen eines Mikrofonarrays nachbilden sollen. Der Vorteil dieses Verfahrens besteht darin, dass die Nachhallzeit, indirekt über den Reflexionsfaktor  $\beta$  der Wände, des Bodens und der Decke, sehr leicht erhöht oder erniedrigt werden kann. Ein zu- oder abnehmender Reflexionsfaktor bedeutet eine zu- oder abnehmende Nachhallzeit. Desweiteren erlaubt die Spiegelquellenmethode eine beliebige Anordnung von Schallquelle und Mikrofonen, in einem kleinen quaderförmigen Raum. Für die Simulationen wird ein Szenario gemäß Abbildung 4.1 entworfen. In einem Raum der Abmessung 495 x 567 x 233 cm



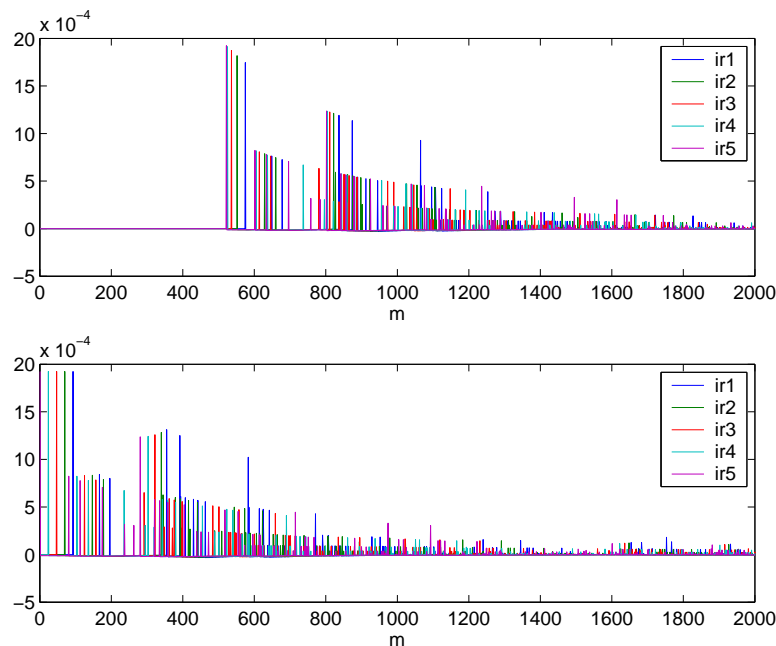


Abbildung 4.2: *Oben: Fünf, mit der Spiegelquellenmethode erzeugte, Impulsantworten. Unten: Die Impulsantworten wurden auf Fernfeldannahme getrimmt. Parameter:  $N = 16384$ ,  $f_s = 48kHz$ ,  $\beta = 0.5$*

befindet sich symmetrisch zur Mitte in 106cm Höhe ein Mikrofonarray mit fünf Mikrofonen. Der Mikrofonabstand zwischen den einzelnen Mikrofonen ist 42cm. 368cm entfernt befindet sich auf gleicher Höhe die Schallquelle, dessen Position relativ zur Symmetrieachse der Anordnung variabel ist (xQuelle ist variabel). Die erzeugten Impulsantworten haben eine Länge  $N$  von 16384 Samples, die Abtastrate  $f_s$  ist 48kHz.

Da die Mehrkanal-Korrelationsmethode von Fernfeldbedingungen ausgeht, werden die Impulsantworten nachträglich auf Fernfeldsituation getrimmt. Der Abstand der Direktschallpeaks in den Kanälen 0 und 1 wird ermittelt und anschließend werden alle anderen Impulsantworten so verschoben, dass die Direktschallpeaks gleichen Abstand voneinander haben. Außerdem wird durch folgende Normalisierung Sorge getragen, dass alle Impulsantworten bezüglich des Direktschalls gleiche Dämpfung aufweisen:

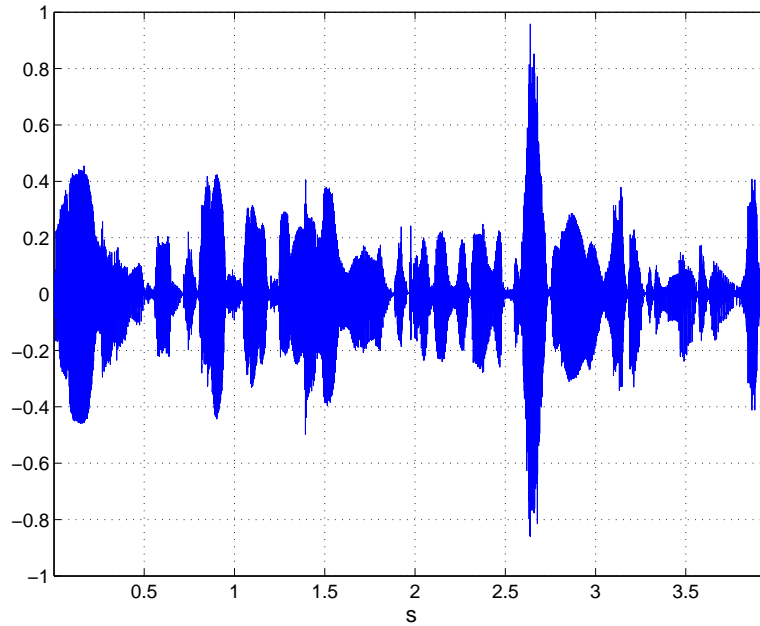


Abbildung 4.3: Sprachsignal einer weiblichen Sprecherin. Die Abtastrate wurde von 44.1kHz auf 48kHz erhöht.

$$ir_i = ir_i \cdot \frac{\max(\mathbf{dp})}{dp_i}, \quad i = 0, 1, \dots, L - 1 \quad (4.1)$$

$$\mathbf{dp} = [dp_0 \ dp_1 \ \dots \ dp_i] \quad (4.2)$$

$ir_i$  ist die Impulsantwort des  $i$ -ten Kanals und  $dp_i$  bezeichnet den Betrag des Direktschallpeaks im  $i$ -ten Kanal. Zur Veranschaulichung zeigt Abbildung 4.2 eine mit der Spiegelquellenmethode erzeugte Impulsantwort und deren Trimmung auf Fernfeldbedingungen.

Für die Vergleichstests unter realistischen Bedingungen werden gemessene Impulsantworten aus den Räumen des Laboratoriums für Nachrichtentechnik an der Universität Erlangen-Nürnberg verwendet. Es werden Impulsantworten aus dem Multimediaraum (geringer bis mittlerer Nachhall,  $T_{60} \approx 250 - 350\text{ms}$ ), aus dem Seminarraum N5.17 (erhöhter Nachhall,  $T_{60} \approx 1000\text{ms}$ ) und aus der Tiefga-

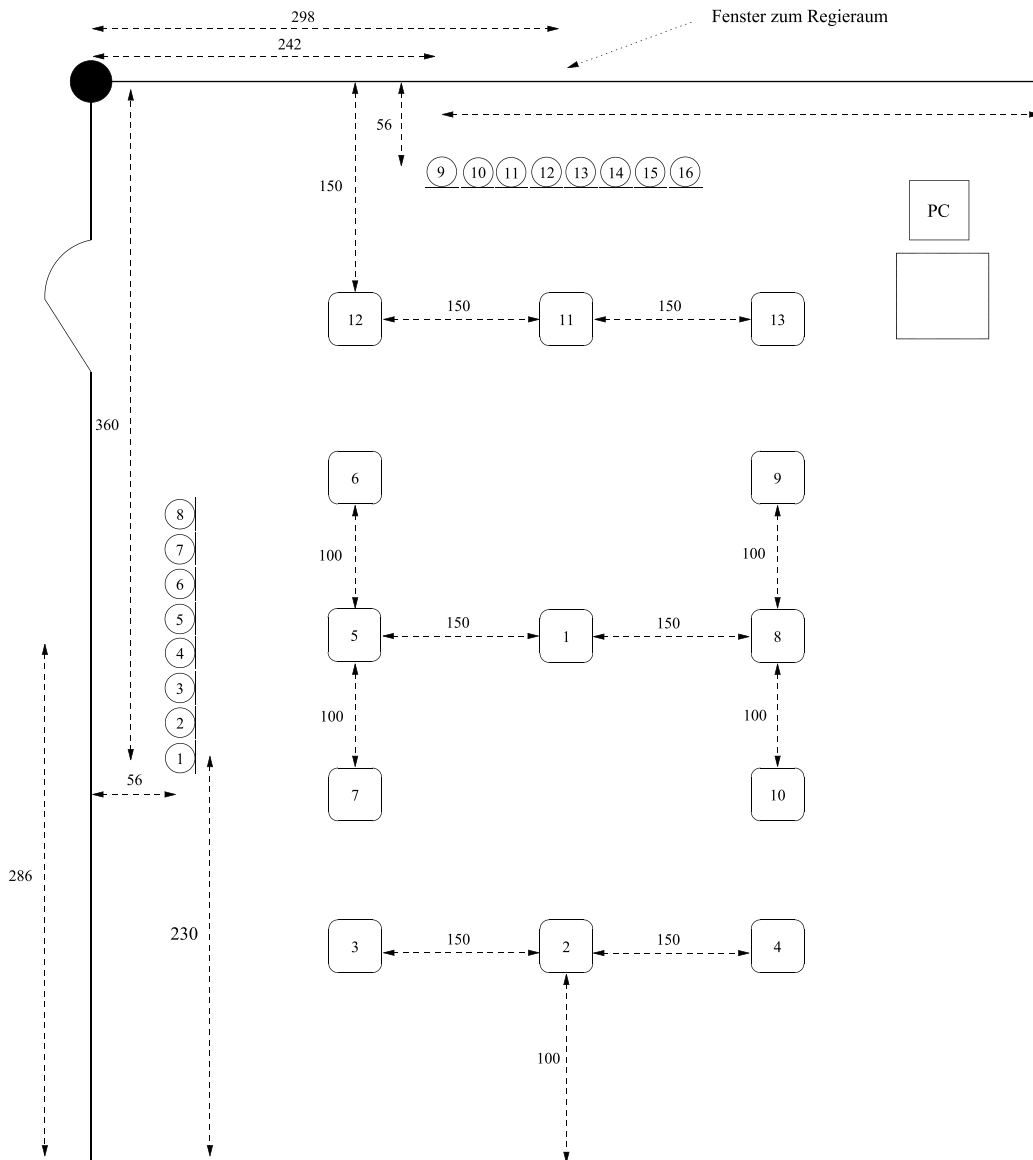


Abbildung 4.4: Grundrisskizze des Multimediaräumes. Die eingerahmten Zahlen, stehen für die Quellenpositionen. Die Mikrofone 1 bis 16 und die jeweilige Quelle befanden sich auf einer Höhe von 151cm.

rage (sehr hoher Nachhall,  $T_{60} \approx 2000\text{ms}$ ) der Technischen Fakultät verwendet. Die Messungen in den einzelnen Räumen liegen der Geometrie zu Grunde, wie sie in Abbildung 4.4 dargestellt ist. Die Abtastrate ist hier ebenfalls 48kHz.

Als Testsignale werden sowohl Rausch- als auch Sprachsignale verwendet. Für

die Simulationen an den Korrelationsalgorithmen bietet sich weißes Rauschen besonders gut an, da dessen Autokorrelationsfunktion einen scharfen Peak aufweist. Das tendentielle Verhalten und die richtige Funktion der Korrelationsalgorithmen kann damit besonders gut gezeigt werden. Als Sprachsignal dient eine ca. 4 Sekunden lange, mit 44.1kHz abgetastete Aufnahme einer weiblichen Sprecherin. Das Sprachsignal wurde auf 48kHz hochgesampled. Wenn der Sprecher gerade nicht spricht, kann es vor allem bei den Korrelations-Algorithmen, durch Eigenrauschen der Mikrofone oder durch Hintergrundgeräusche zu Fehlschätzungen kommen. Darum werden längere Sprachpausen grob per Hand herausgeschnitten.

## **4.2 Vergleich von zwei- und mehrkanaligen Korrelationsalgorithmen**

Die prinzipielle Funktionsweise der MCC-Methode stützt sich auf die Annahme, dass Fernfeldbedingungen vorliegen. Um die richtige Funktion und das bessere Verhalten von MCC gegenüber GCC zumindest unter idealisierten Bedingungen zu zeigen, werden im ersten Unterpunkt dieses Kapitels ausschließlich auf Fernfeldannahme getrimmte Impulsantworten und weißes Rauschen als Testsignal verwendet. Im darauf folgenden Abschnitt wird versucht die unter idealisierten Bedingungen erhaltenen Ergebnisse ebenso mit realen Impulsantworten und Sprachsignalen zu erzielen.

### **4.2.1 Untersuchungen mit generierten Impulsantworten**

#### **Zunahme der Nachhallzeit**

Als erstes wird gezeigt, wie sich die MCC-Methode mit drei, vier und fünf Mikrofonen gegenüber der GCC-Methode bei der Schätzung der Laufzeitdifferenz zwischen zwei Mikrofonen bei zunehmender Nachhallzeit behauptet. Als Test-

signal dient eine Rauschsequenz von vier Sekunden. Aus diesem Signal werden insgesamt elf aufeinanderfolgende Blöcke von je 30ms herausgeschnitten, die sich jeweils um 50 Prozent überlagern. Mit Hilfe der Spiegelquellenmethode werden fünf Impulsantworten ( $ir_{0,1,\dots,4}$ ) generiert, die die Kanäle von der Schallquelle zu den fünf Mikrofonen des Mikrofonarrays repräsentieren. Jeder Block wird mit den Kanalimpulsantworten gefiltert, so dass man die Mikrofonensignale  $x_0$ ,  $x_1$ ,  $x_2$ ,  $x_3$  und  $x_4$  erhält (vgl. Gleichung (4.3)).

$$x_i[n] = ir_i * s[n], \quad i = 0, 1, \dots, 4 \quad (4.3)$$

Ziel ist es, den Laufzeitunterschied zwischen den Signalen  $x_0$  und  $x_1$  zu schätzen. Die GCC-Methode braucht zur Schätzung der Laufzeitdifferenz nur  $x_0$  und  $x_1$ . Die MCC-Methode verwendet bis zu fünf Signale auf einmal ( $x_0, x_1, x_2$ ), ( $x_0, x_1, x_2, x_3$ ) und ( $x_0, x_1, x_2, x_3, x_4$ ) um die Laufzeitdifferenz zwischen  $x_0$  und  $x_1$  zu schätzen. In jedem Durchlauf wird der Reflexionskoeffizient  $\beta$  bei der Berechnung der Impulsantworten erhöht. So erhält man für eine bestimmte Quellenposition und für einen Signalblock ein Ergebnis wie in Abbildung 4.5. Entlang des zunehmenden Reflexionsfaktors wird nach oben der errechnete Schätzwert für die Laufzeitdifferenz aufgetragen. Als Referenz wird der wahre Laufzeitunterschied (engl.: *lag*) in Samples mit aufgetragen. Es ist deutlich zu erkennen, dass die MCC-Methode mehr richtige Schätzungen vornimmt, als die GCC-Methode. Die Vermutung liegt Nahe, dass die Verwendung von mehreren Mikrofonen, eine größere Robustheit bei der Schätzung von Laufzeitdifferenzen bei höheren Nachhallzeiten hat, als die GCC-Methode, die nur zwei Mikrofone verwendet.

Um diese Aussage zu überprüfen, wird nicht nur eine Quellenposition betrachtet, sondern insgesamt 60. Für jede Quellenposition werden nun wieder, wie oben beschrieben, die Impulsantworten berechnet. Dieses Mal wird nicht nur für einen 30ms Block die Laufzeitdifferenz ermittelt, sondern für alle elf. Eine

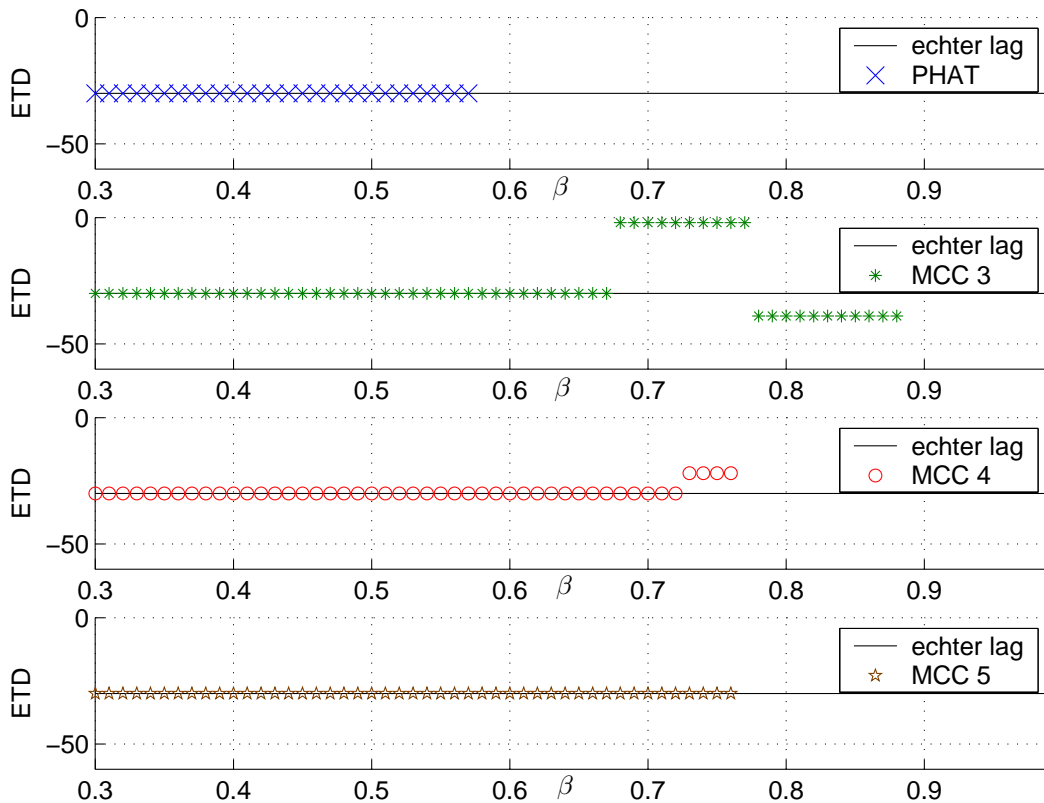


Abbildung 4.5: Über dem Reflexionsfaktor  $\beta$  aufgetragen der Schätzwert (ETD, engl.: Estimated Time Delay) für die Laufzeitdifferenz zwischen  $x_0$  und  $x_1$ . Durch die Verwendung mehrerer Mikrofone (MCC 3, MCC 4 und MCC 5) werden die Schätzungen robuster gegenüber stärkerem Nachhall

Schätzung wird als richtig eingestuft, wenn die Differenz der geschätzten Laufzeitdifferenz und der wahren Laufzeitdifferenz gleich Null ist. Die Anzahl der richtigen Schätzungen wird für jede Quellenposition, für jedes Verfahren und für jeden Signalblock gezählt und anschließend über die Gesamtzahl der Blöcke (hier 11) gemittelt. Das Ergebnis dieser Simulation zeigt Abbildung 4.6. Die Mittelwerte der Anzahl der richtigen Schätzungen jedes Verfahrens sind ebenso mitangegeben. Es ist deutlich zu erkennen, dass die GCC-Methode hier am schlechtesten abschneidet. Das MCC-Verfahren hat mit der Verwendung von drei Mikrofonen im Mittel ca. fünf richtige Schätzungen mehr, mit vier Mikrofonen im Mittel ca.

10 und mit fünf Mikrofonen im Mittel ca. 14 richtige Schätzungen mehr als die GCC-Methode.

Übertragen auf das Ergebniss in Abbildung 4.5 bedeutet dies, dass der Reflexionsfaktor  $\beta$  bei einer Hinzunahme eines weiteren Mikrofonen um 0.05 erhöht werden kann, ohne dass die Zahl der Fehlschätzungen zunimmt. Somit ist gezeigt, dass das MCC-Verfahren in Szenarien mit größerem Nachhall durchaus robuster bei der Laufzeitdifferenzschätzung ist, wie das GCC-Verfahren. Einschränkend muss allerdings betont werden, dass das Testsignal hier weißes Rauschen war und die mit der Spiegelquellenmethode erzeugten Impulsantworten die Fernfeldbedingung ideal erfüllten.

Eine weitere interessante Eigenschaft der Korrelationsalgorithmen ist in Abbildung 4.6 zu erkennen. Offenbar gibt es Positionen, an denen die Algorithmen besonders gut funktionieren. Dies zeigen die ausgeprägten Spitzen im Bereich von ca. 1,6m - 3,2m. Gerade dort befindet sich die Quelle genau gegenüber dem Mikrofonarray. Die Anzahl dieser idealen Quellenpositionen nimmt mit der Anzahl der verwendeten Mikrofone zu. Andererseits gibt es in diesem Bereich auch Quellenpositionen, die eine Schätzung unterhalb des Mittelwerts hervorbringen. Dieser beobachtete Effekt wurde in dieser Arbeit nicht näher untersucht. Dennoch lässt die Beobachtung den Schluss zu, dass die Leistungsfähigkeit der Korrelationsmethoden offensichtlich mit der Position der Quelle relativ zum Mikrofonarray in gewisser Weise korreliert ist.

### **Verschlechterung des Signal- zu Geräuschleistungsverhältnisses**

Als nächstes soll das Verhalten des GCC- und MCC-Verfahrens bei sich verschlechterndem Signal- zu Geräuschleistungsverhältnis (SNR) untersucht werden. Als Testsignal dient wieder weißes Rauschen, das in sich überlappende 30ms Blöcke segmentiert wird. Auch die Raumgeometrie wird nicht verändert. Die Schallquelle befindet sich an einer festen Position (3,33m von der linken Wand entfernt). Für sämtliche Reflexionsfaktoren ( $\beta = 0.3, 0.01, \dots, 0.99$ ) werden die Im-

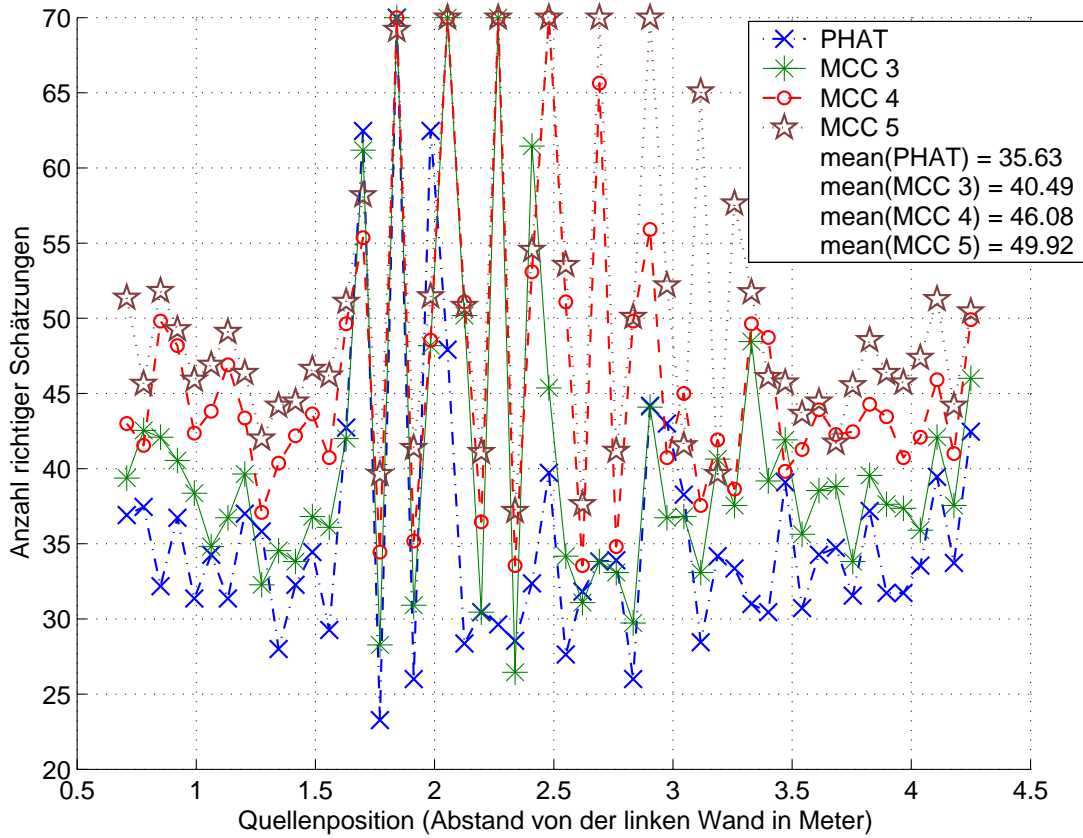


Abbildung 4.6: Anzahl der richtigen Schätzungen von MCC mit 3, 4 und 5 Mikrofonen, sowie GCC aufgetragen über sämtliche Quellenpositionen. Im Mittel schätzt die MCC-Methode mit mehreren Mikrofonen die Laufzeitdifferenz zwischen zwei Mikrofonen besser als die GCC-Methode, die nur 2 Mikrofone verwendet.

pulsantworten ( $ir_{0,1,\dots,4}$ ) von der Quelle zu den einzelnen Mikrofonen mit der Spiegelquellenmethode berechnet. Nachdem das Rauschsignal durch die Impulsantworten gefiltert wurde, wird zu jedem Signal  $x_{0,1,\dots,4}$  eine Rauschsequenz  $w_{0,1,\dots,4}$  addiert:

$$x_i[n] = ir_i * s[n] + w_i[n], \quad i = 0, 1, \dots, 4 \quad (4.4)$$

Die Rauschsequenzen sind untereinander unkorreliert. Die Leistung der Rausch-



sequenz ist variierbar und so kann über den Zusammenhang

$$\text{SNR} = \frac{\text{Signalleistung}}{\text{Rauschleistung}} = \frac{\text{Var}(x_i)}{\text{Var}(w_i)}, \quad i = 0, 1, \dots, 4 \quad (4.5)$$

das SNR direkt eingestellt werden. Das SNR wird zwischen -20dB und +23dB gewählt. Die so erzeugten Mikrofonsignale werden wieder in sich um die Hälfte überlappende 30ms Blöcke segmentiert. Jedes Verfahren (GCC, MCC mit 3 Mikrofonen, MCC mit 4 Mikrofonen, MCC mit 5 Mikrofonen) schätzt nun bei festem SNR den Laufzeitunterschied zwischen  $x_0$  und  $x_1$  bei sich erhöhendem  $\beta$ . So erhält man für ein bestimmtes SNR ein Ergebnis wie in Abbildung 4.5. Eine Laufzeitdifferenzschätzung gilt dann als richtig, wenn die Differenz zum wahren Laufzeitunterschied gleich Null ist. Die Anzahl der richtigen Schätzungen pro Signalblock wird gezählt und anschließend wird das Ergebnis über elf Blöcke gemittelt. Das Resultat dieser Berechnung zeigt Abbildung 4.7 Man kann deutlich sehen, dass unterhalb von 5dB SNR, die Leistungsfähigkeit des MCC- so wie auch des GCC-Verfahrens merklich zurückgeht. Die Algorithmen erreichen ihre volle Performanz ab ca. 15dB SNR. Vergleicht man die Abstände der Verfahren untereinander, so sind sie nahezu identisch mit denen in Abbildung 4.6 für die Quellenposition 3, 33m von der linken Wand.

Um die Vermutung zu bekräftigen, dass das mehrkanalige Korrelationsverfahren bei schlechterem SNR bessere Performanz bietet als das einfache GCC- Verfahren, wird eine weitere Simulation durchgeführt. Für eine feste Quellenposition und einen festen Reflexionsfaktor werden die Impulsantworten mit der Spiegelquellenmethode für die Pfade zwischen Schallquelle und den fünf Mikrofonen des Mikrofonarrays berechnet. Analog zur gerade gezeigten Simulation wird das SNR zwischen -20dB und +23dB variiert. Dieses Mal wird gezählt bei wie vielen Blöcken eine richtige Schätzung der Laufzeitdifferenz erfolgen kann. Das Ergebnis zeigt Abbildung 4.8.

Es ist zu erkennen, dass gerade im Bereich unterhalb von 5dB SNR das MCC-

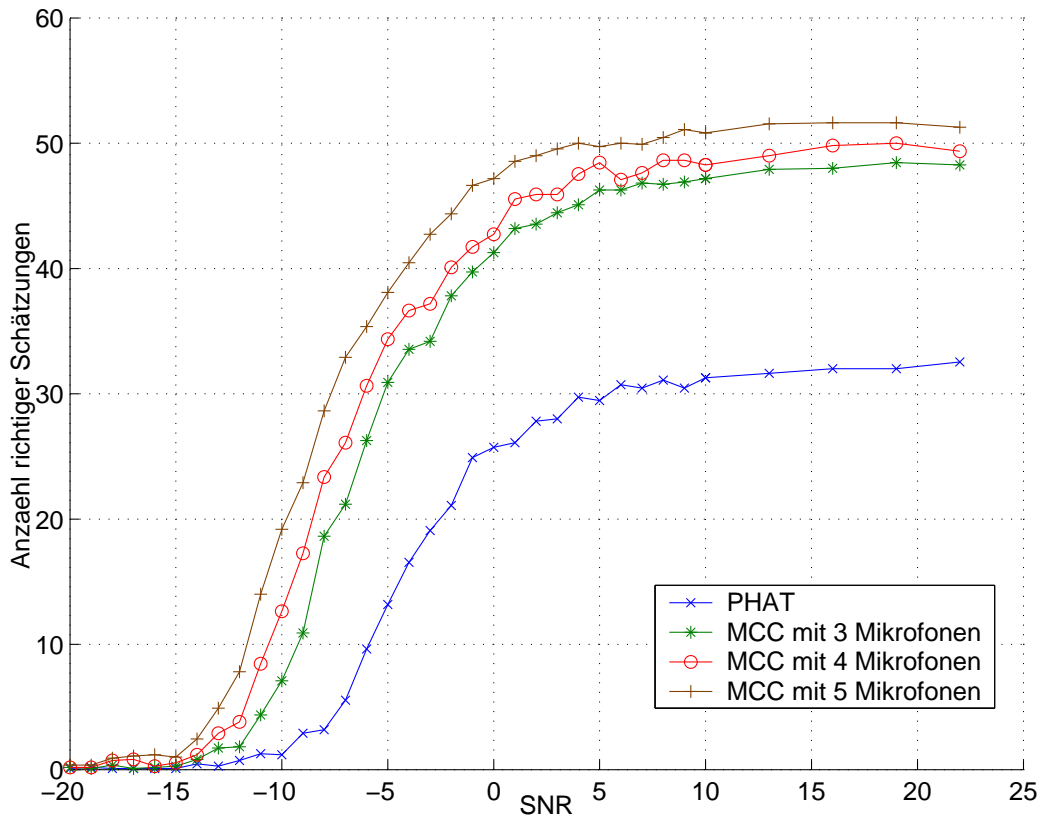


Abbildung 4.7: Anzahl der richtigen Schätzungen von MCC mit 3, 4 und 5 Mikrofonen, sowie GCC aufgetragen über dem SNR. Maximal können 70 richtige Schätzungen erfolgen.

Verfahren besser ist, als das GCC-Verfahren.

#### 4.2.2 Untersuchungen mit gemessenen Impulsantworten

In Abschnitt 4.2.1 konnte gezeigt werden, dass das MCC-Verfahren, zumindest unter idealen Rahmenbedingungen (weißes Rauschen, Fernfeldannahme), eine höhere Robustheit gegenüber Nachhall und sich verschlechterndem SNR hat, als das GCC-Verfahren. Diese idealisierte Betrachtung, hat allerdings mit realen Umständen recht wenig zu tun. Zum einen kann nicht erwartet werden, dass in Räumen mit begrenzten Abmessungen, in denen sich ein Sprecher wenige Meter vom Mikrofonarray befindet, an jeder Position ideale Fernfeldbedingungen herr-

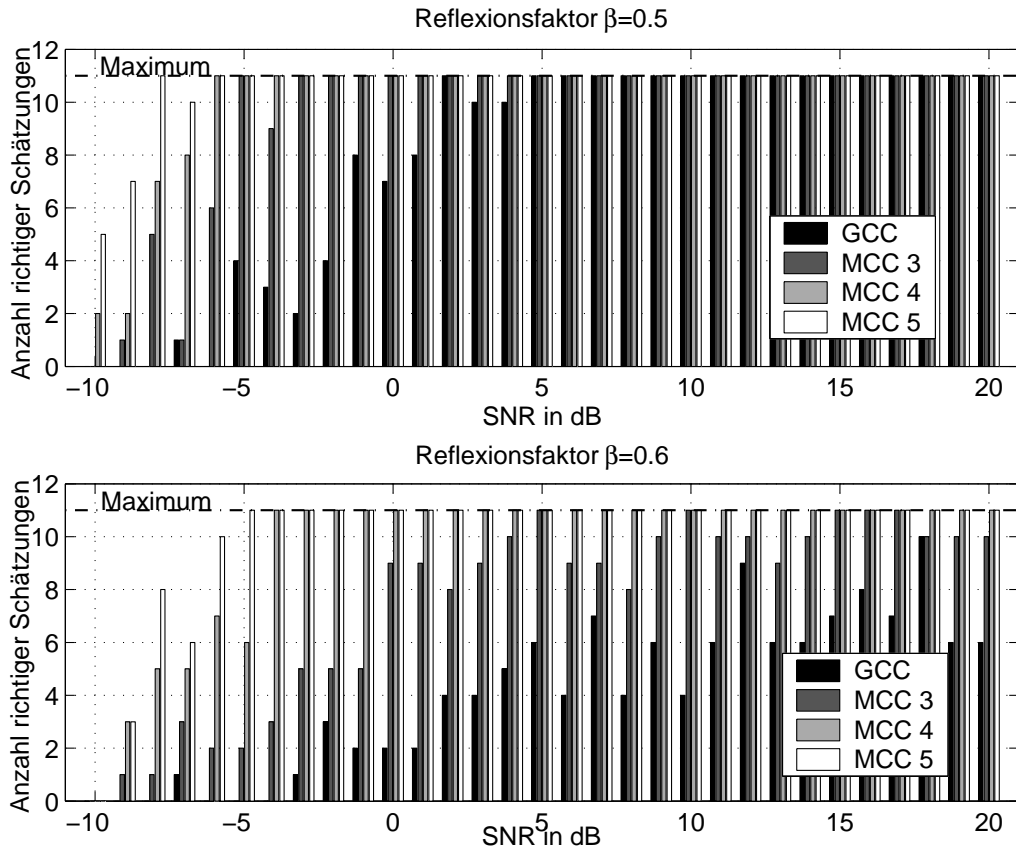


Abbildung 4.8: Anzahl der richtigen Schätzungen von MCC mit 3, 4 und 5 Mikrofonen, sowie GCC aufgetragen über dem SNR. Oben:  $\beta = 0.5$ ; Unten:  $\beta = 0.6$

sehen. Desweiteren ist Sprache weder stationär noch weiß und weist somit keinen scharfen Peak in seiner Autokorrelationsfunktion auf, was eine exakte Quellenpositionsbestimmung erschwert.

Für die nun folgenden Untersuchungen werden gemessene Impulsantworten aus den Räumen N5.17, des Multimediaraumes (MMR) mit zu- und aufgezogenen Vorhängen sowie aus der Tiefgarage verwendet. Als Testsignal diente weißes Rauschen, gefärbtes Rauschen mit einem sprachähnlichen Spektrum und das Sprachsignal aus Abschnitt 4.1. Für beliebige Quellenpositionen (1, 2, 3, 4, 6, 9, 11) (siehe Abbildung 4.4) werden die Impulsantworten zu den Mikrofonen 1 bis 5 und 9 bis 13 mit den Testsignalen gefiltert um die Mikrofonsignale  $x_1$ ,  $x_2$ ,  $x_3$ ,

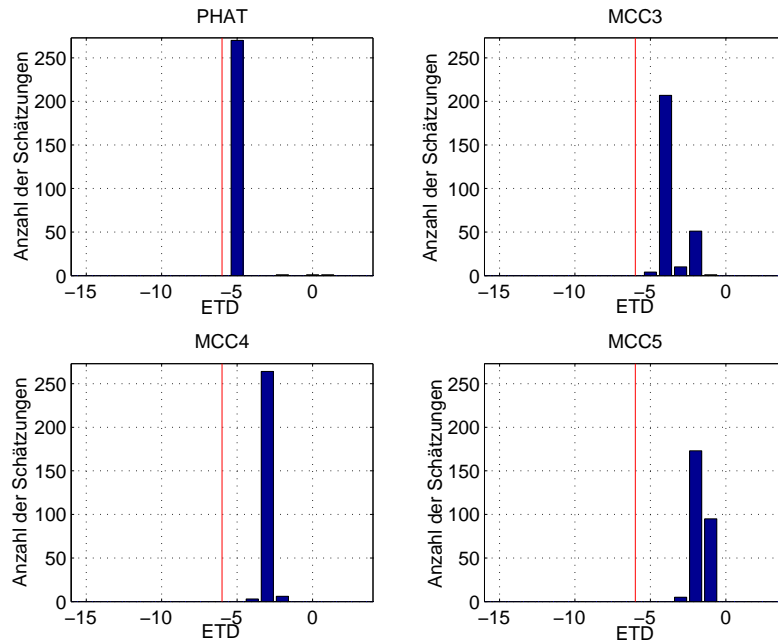


Abbildung 4.9: Verteilung der geschätzten Laufzeitdifferenzen zwischen Mikrofonpaar 1-2 mit Rauschen als Testsignal (N5.17, Quellenposition 1)

$x_4$  und  $x_5$  sowie  $x_9$ ,  $x_{10}$ ,  $x_{11}$ ,  $x_{12}$ , und  $x_{13}$  zu generiert (analog Gleichung (4.3)). Anschließend wird mit GCC, mit MCC 3, mit MCC 4 und MCC 5 der Laufzeitunterschied geschätzt. Bei dieser Testreihe werden allerdings beim Sprachsignal insgesamt 270 Blöcke von 1400 Samples (ca. 30ms) verwendet, bei den Rauschsignalen 273 Blöcke (die Abweichung rührt daher, dass das Sprachsignal nicht exakt 4 Sekunden lang ist).

Im Gegensatz zu den Messreihen, die unter ideal simulierten Bedingungen durchgeführt wurden, zeigen die Ergebnisse vor allem mit Sprache ein sehr widersprüchliches Verhalten der Algorithmen. Im Wesentlichen kann die Aussage getroffen werden, dass der mehrkanalige Korrelationsalgorithmus keine Verbesserung der Schätzgenauigkeit mit realen Impulsantworten und weißem Rauschen als Testsignal aufzeigt, was die Abbildung 4.9 exemplarisch für eine Vielzahl solcher gefundener Situationen belegt. Man gewinnt viel mehr den Eindruck, dass durch eine Zunahme der Mikrofonzahl die Abweichung vom realen Laufzeitunterschied,

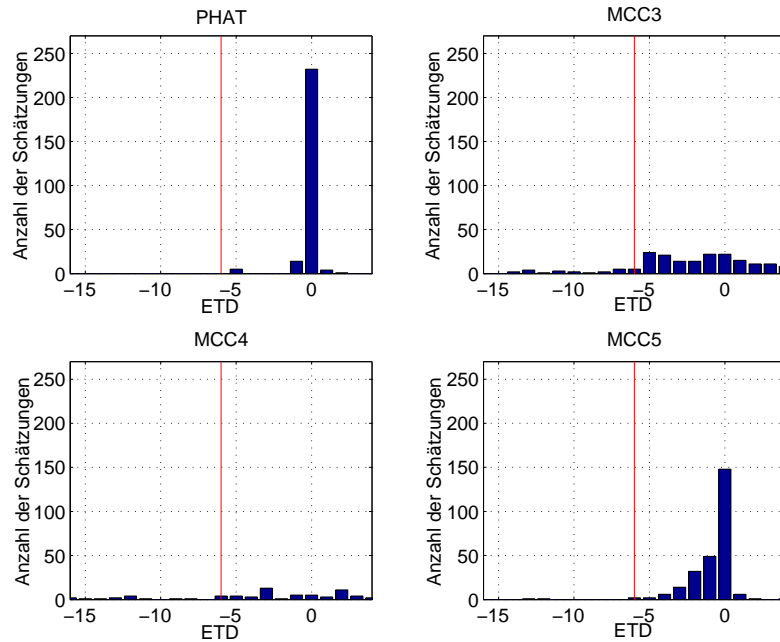


Abbildung 4.10: Verteilung der geschätzten Laufzeitdifferenzen zwischen Mikrofonpaar 1-2 mit Sprache als Testsignal ( $N5.17$ , Quellenposition 1)

in den Grafiken gekennzeichnet durch die vertikale Linie in der Mitte, zunimmt. Ein möglicher Grund für dieses Verhalten kann die nicht exakte Erfüllung der Fernfeldannahme sein. Die gemessenen Impulsantworten wurden nur mit einer endlichen Abtastrate ( $f_s = 48kHz$ ) aufgenommen. Dies kann dazu führen, dass die Direktschallpeaks nicht äquidistanten Abstand voneinander haben, vor allem dann, wenn der Peak bei der Abtastung nicht genau getroffen wurde. Da der MCC-Algorithmus aber von Fernfeldbedingungen ausgeht, tendiert er dazu mit einer Zunahme der Mikrofonzahl falsch zu schätzen. Um diese Theorie zu bekräftigen wird die Aperturgröße (Mikrofonabstand) vergrößert. Es werden nun die Mikrofone 1, 3, 5, und 7, sowie 9, 11, 13 und 15 verwendet. Durch die Vergrößerung der Apertur verschlechtert sich die Fernfeldbedingung, weil der Mikrofonabstand im Verhältnis zum Abstand der Schallquelle größer wird. Nun kann allerdings der MMC-Algorithmus mit fünf Mikrofonen nicht mehr verwendet werden, weil ein fünftes Mikrofon im gleichen Abstand nicht vorhanden ist. Abbildungen 4.11 und

4.12 zeigen das Verhalten der Algorithmen.

In Abbildung 4.11 schätzen die GCC-Methode und MCC 3 in etwa den gleichen Laufzeitunterschied, der um ein Sample (dies entspricht einer Differenz von etwa 0,7cm) vom wahren Laufzeitunterschied abweicht (vertikale Linie in der Mitte). Die Schätzungen von MCC 4 liegen um 3 Samples neben dem richtigen Wert. Abbildung 4.12 zeigt, dass nach Vergrößerung der Apertur die Schätzungen des MCC 3 und MCC 4 sich je um ein weiteres Sample verschlechtern.

Die Algorithmen zeigen mit dem gefärbtem Rauschen ein ganz ähnliches Verhalten, wie Abbildungen 4.13 und 4.14 darstellen. Das Spektrum der verwendeten Rauschsequenz ist ähnlich dem Spektrum von Sprache. Mit Sprache scheinen die Algorithmen insgesamt nur noch sehr mangelhaft zu funktionieren, was Abbildung 4.10 exemplarisch belegt. Da das gefärbte Rauschen als Testsignal nicht größere Unterschiede bei den Schätzungen erzeugt als weißes Rauschen, muss die mangelhafte Funktion der Algorithmen mit dem Sprachsignal auf dessen Instationarität zurückgeführt werden und nicht auf den begrenzten Spektralbereich.

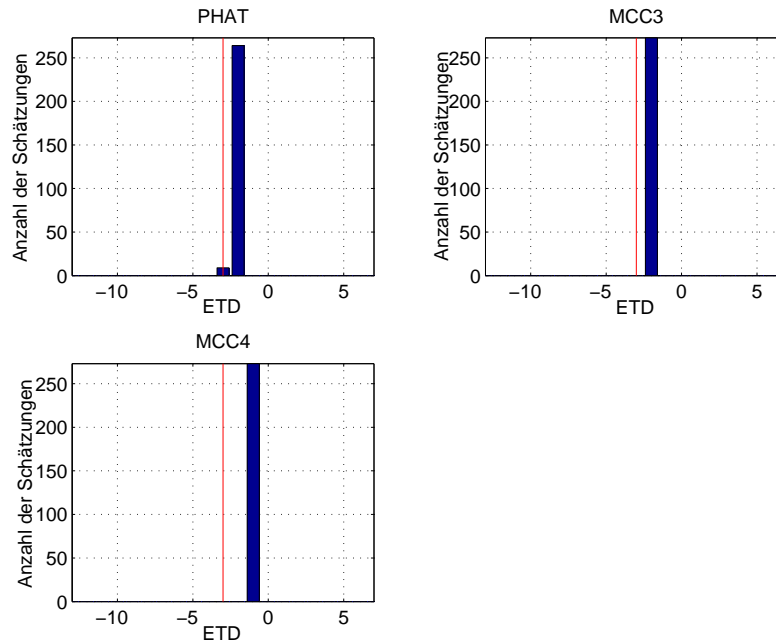


Abbildung 4.11: Verteilung der geschätzten Laufzeitdifferenzen zwischen Mikrofonpaar 9-10 mit Rauschen als Testsignal (MMR, Quellenposition 2)

### 4.3 Vergleich von zwei- und mehrkanaligen Adaptionsalgorithmen

Die in Kapitel 2 vorgestellten adaptiven Algorithmen schätzen blind die Kanalübertragungsfunktionen bzw. die Impulsantworten von der Quelle zu den Mikrofonen. Die Verfahren beruhen auf dem allgemeinen Schallausbreitungsmodell (Gleichung (2.2)) und sollten daher von vorne herein eine bessere Leistungsfähigkeit in halligen Räumen erwarten lassen, als die Korrelationsmethoden, deren Funktion sich auf das ideale Schallausbreitungsmodell stützt. Da der zweikanalige AED-Algorithmus ebenso wie der Mehrkanal-AED-Algorithmus weder Nah- noch Fernfeldbedingungen voraussetzen, wird darauf verzichtet die Untersuchungen bezüglich ihrer Leistungsfähigkeit mit Impulsantworten durchzuführen, die mit der Spiegelquellenmethode erstellt wurden. Es wird ein Satz von gemessenen Impulsantworten verwendet. Als Testsignal wird ausschließlich mit dem

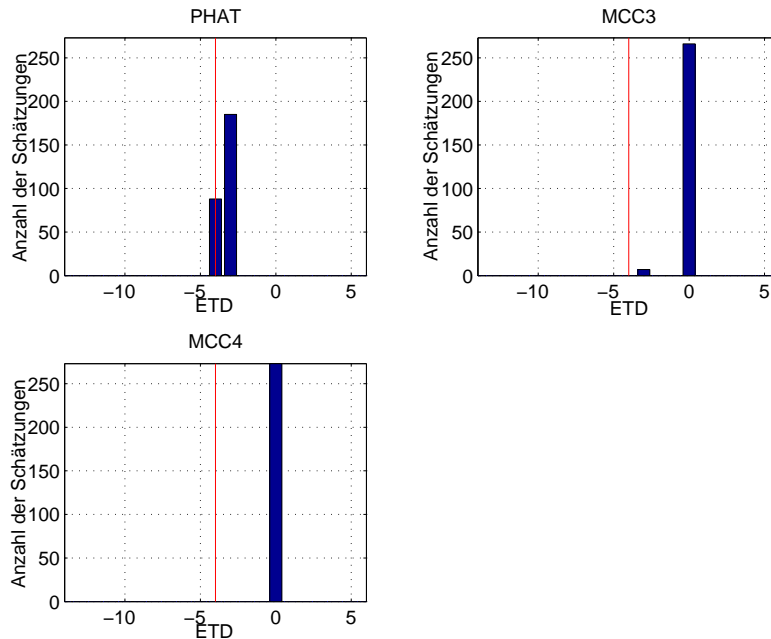


Abbildung 4.12: Verteilung der geschätzten Laufzeitdifferenzen zwischen Mikrofonpaar 9-11 mit Rauschen als Testsignal (MMR, Quellenposition 2)

Sprachsignal aus Abschnitt 4.1 gearbeitet. Auf weiterführende Untersuchungen mit Rauschsignalen wird verzichtet, da erste Tests keine wesentlichen Unterschiede im Verhalten der Algorithmen erkennen ließen, ob sie nun mit Rauschen oder Sprache durchgeführt wurden. Somit lässt sich sagen, dass die adaptiven Algorithmen unter realistischen Bedingungen untersucht werden und die erhaltenen Ergebnisse auch für reale Anwendungen aussagekräftig sind.

Für den Vergleichstest werden die Mikrofone 3 bis 7 (siehe Abbildung 4.4) ausgewählt. Als Quellenpositionen dienen Position 1, 2, 4, 7, 9, 11 und 12. Somit ist der Fall, dass sich die Quelle in nahem Abstand befindet mit den Positionen 12 und 7 abgedeckt. Ein mittlerer Abstand kann mit den Impulsantworten, die mit den Quellenpositionen 1, 2 und 11 gemessen wurden, simuliert werden. Für die Modellierung eines größeren Abstands treten die Quellenpositionen 4 und 9 ein. Die Positionierung der Quelle reicht aus Sicht der Mikrofone ebenso von weit links außen (11, 12), über mittige Positionen (1, 7, 9), bis hin zu weit rechts außen



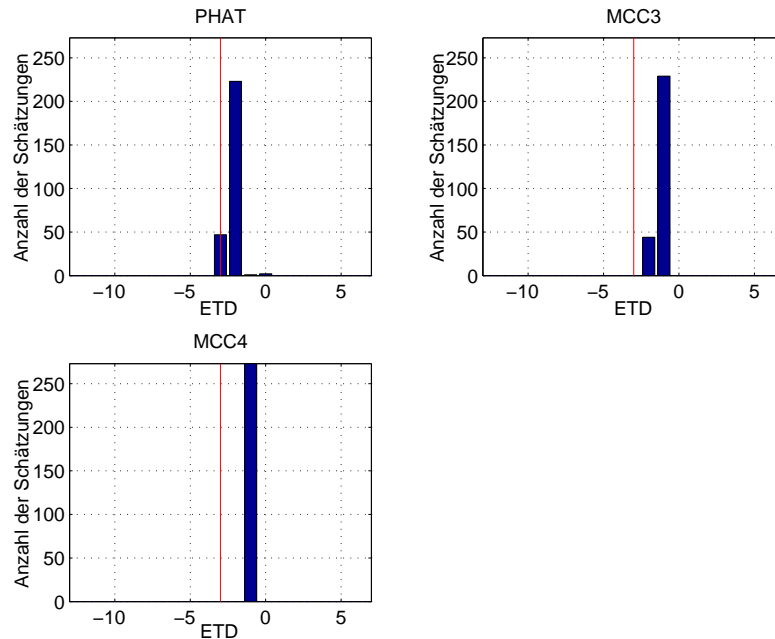


Abbildung 4.13: Verteilung der geschätzten Laufzeitdifferenzen zwischen Mikrofonpaar 9-10 mit gefärbtem Rauschen als Testsignal (MMR, Quellenposition 2)

(2, 4). Diese Auswahl an Quellenpositionen soll eine möglichst von der Position der Quelle unabhängige Beurteilung des Verhaltens der Algorithmen erlauben. Es werden entsprechend die Impulsantworten aus dem Multimediarraum mit auf- und zugezogenen Vorhängen, aus dem Raum N5.17 und aus der Tiefgarage verwendet. Der Multimediarraum mit geschlossenen Vorhängen hat in dieser Reihe die kleinste Nachhallzeit ( $T_{60} \approx 250\text{ms}$ ). Eine immer noch kleine, aber etwas größere Nachhallzeit liegt im Multimediarraum vor, wenn die Vorhänge aufgezogen sind ( $T_{60} \approx 350\text{ms}$ ), da nun die Wände und Fenster eine bessere Reflexion (größeren Reflexionsfaktor) bieten. Eine weitere Steigerung der Nachhallzeit findet man im Raum N5.17 ( $T_{60} \approx 1000\text{ms}$ ), dort liegt kein Teppich aus. In der Tiefgarage ist der Nachhall am größten ( $T_{60} \approx 2000\text{ms}$ ).

Es werden nun der zweikanalige AED-Algorithmus, implementiert in der Matlabdatei AED.m, mit dem mehrkanaligen AED-Algorithmus (MAED), implementiert in der Matlabdatei multiAED5.m, verglichen. Der MAED-Algorithmus wird

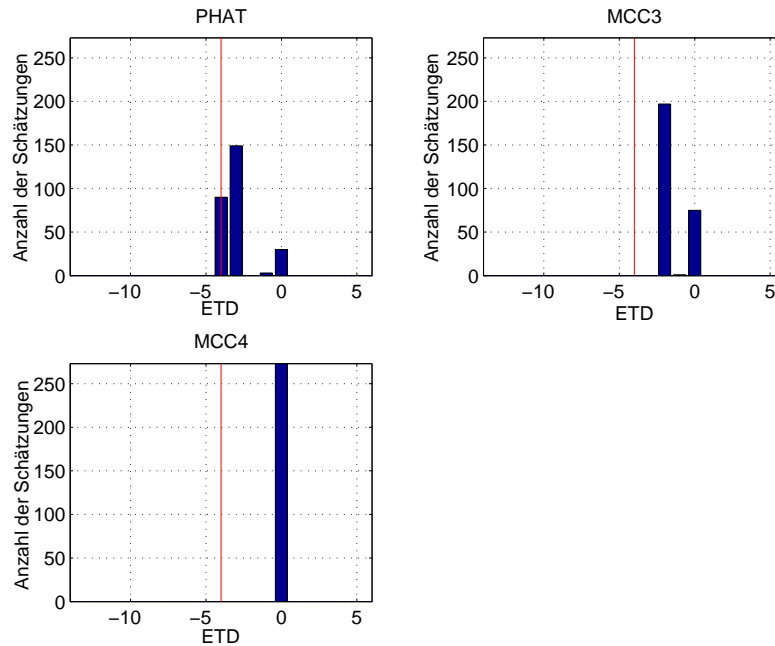


Abbildung 4.14: Verteilung der geschätzten Laufzeitdifferenzen zwischen Mikrofonpaar 9-11 mit gefärbtem Rauschen als Testsignal (MMR, Quellenposition 2)

mit 2 (MAED 2), 3 (MAED 3), 4 (MAED 4) und 5 (MAED 5) Mikrofonsignalen betrieben. Zum einen soll herausgestellt werden, ob es Unterschiede bei der Schätzgenauigkeit zwischen dem AED-Algorithmus und dem MAED-Algorithmus gibt. Zum anderen steht im Interesse der Untersuchungen, inwieweit sich die Schätzgenauigkeit des MAED-Algorithmus verbessert, wenn man weitere Mikrofonsignale hinzunimmt. Als erstes werden dazu die Algorithmen für den Fall, dass sich die Quelle an Position 1 befindet, so eingestellt (Anpassung der Schrittweiten), dass sie bei der Schätzung der Laufzeitdifferenzen richtige Ergebnisse liefern. Für eine beliebige weitere Quellenposition wird die Einstellung überprüft und falls nötig korrigiert. Der MAED-Algorithmus mit 2 Mikrofonen benötigt unter Umständen eine andere Schrittweite  $\mu_f$  (vgl. Gleichung 3.22) als der MAED-Algorithmus mit 3, 4 oder 5 Mikrofonen. So werden die Schrittweiten für jeden Fall extra eingestellt. Eine Übersicht der verwendeten Schrittweiten liefert Tabelle 4.1. Das Sprachsignal wird mit den Impulsantworten für die Mikrofone 3 bis 7 gefiltert

Tabelle 4.1: *verwendete Schrittweiten  $\mu$* 

<b>Raum</b>	<b>AED</b>	<b>MAED 2</b>	<b>MAED 3</b>	<b>MAED 4</b>	<b>MAED 5</b>
MMRcclosed <sup>a</sup>	0.1	0.1	0.1	0.1	0.1
MMRcopen <sup>b</sup>	0.1	0.1	0.5	0.5	0.5
N5.17	0.1	0.1	0.1	0.1	0.5
Garage	0.8	0.8	0.8	0.8	0.8

<sup>a</sup>Multimediarraum mit zugezogenen Vorhängen

<sup>b</sup>Multimediarraum mit aufgezogenen Vorhängen

um die Mikrofonsignale  $x_1, x_2, x_3, x_4$  und  $x_5$  zu generieren. Anschließend schätzen der AED-Algorithmus und der MAED-Algorithmus die Laufzeitdifferenzen zwischen den Mikrofonpaaren 3-4, 3-5, 3-6, 3-7, 4-5, 4-6, 4-7, 5-6, 5-7 und 6-7. Da der MAED-Algorithmus mit mehr als zwei Mikrofonsignalen die Impulsantworten aller beteiligten Kanäle auf einmal versucht zu identifizieren, dürfen nicht alle möglichen Laufzeitdifferenzen, die theoretisch aus den geschätzten Impulsantworten zu ermitteln wären, für den Vergleich mit den zweikanaligen Verfahren herangezogen werden. Es werden immer nur die Laufzeitdifferenzen zwischen dem vorinitialisierten Kanal und den anderen vollständig selbst geschätzten Kanälen ausgewertet. Wäre ein Kanal vollkommen falsch identifiziert, so würden alle Laufzeitdifferenzen zwischen ihm und den anderen Kanälen ebenso falsch sein, auch wenn die anderen Kanäle richtig geschätzt wären. Erst durch diese Einschränkung ist ein Vergleich der Lesitungsfähigkeit mit den zweikanalig arbeitenden Algorithmen fair. Tabelle 4.2 zeigt die Laufzeitdifferenzen, die mit den einzelnen Algorithmen ermittelt wurden. MAED 3 verwendet die Mikrofontripel (3,4,5), (4,5,6) und (5,6,7), MAED 4 die Quartette (3,4,5,6) und (4,5,6,7), MAED 5 verwendete alle Mikrofone (3,4,5,6,7). Es werden nacheinander alle Laufzeitdifferenzen für alle Quellenpositionen und für jeden Raum ermittelt. Die Laufzeitdifferenzen werden mit den wahren Zeitunterschieden der Direktschallpeaks der Impulsantworten verglichen. Liegt die Differenz zwischen Schätzung und wahren Zeitunterschied unterhalb

Tabelle 4.2: mit (M)AED-Algorithmen geschätzte Laufzeitdifferenzen

Algorithmus	Mikrofonpaare
AED	3-4, 3-5, 3-6, 3-7, 4-5, 4-6, 4-7, 5-6, 5-7, 6-7
MAED 2	3-4, 3-5, 3-6, 3-7, 4-5, 4-6, 4-7, 5-6, 5-7, 6-7
MAED 3	3-4, 3-5, 4-5, 4-6, 5-6, 5-7
MAED 4	3-4, 3-5, 3-6, 4-5, 4-6, 4-7
MAED 5	3-4, 3-5, 3-6, 3-7

eines definierten kritischen Limits, so wird die Schätzung als richtig eingestuft. Die Abbildungen 4.15 bis 4.18 zeigen das Ergebnis dieser Testreihe. Das kritische Limit ist hier 2 Samples. Somit wird eine Abweichung vom wahren Laufzeitunterschied um 1 Sample (geometrisch entspricht dies 0,7cm) toleriert.

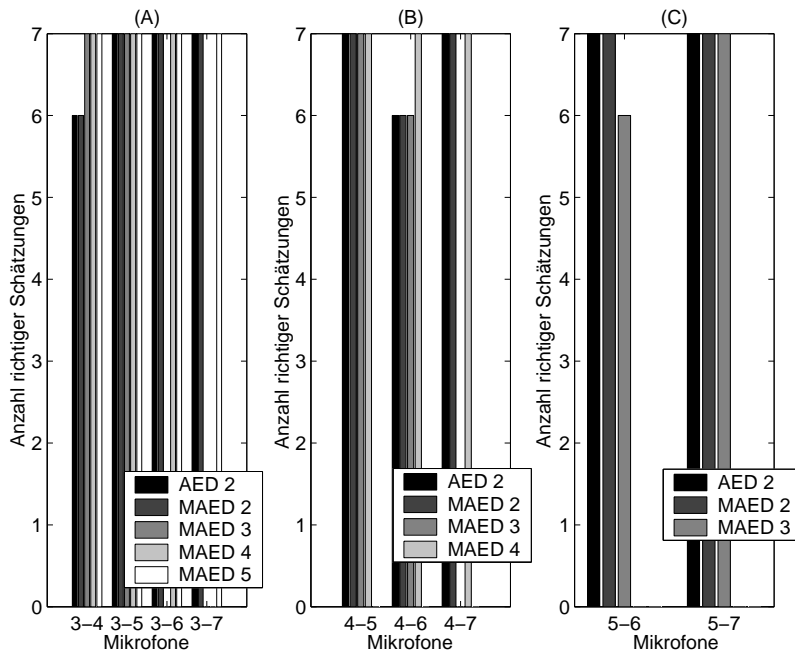


Abbildung 4.15: AED vs. MAED: Multimediaraum mit zugezogenen Vorhängen

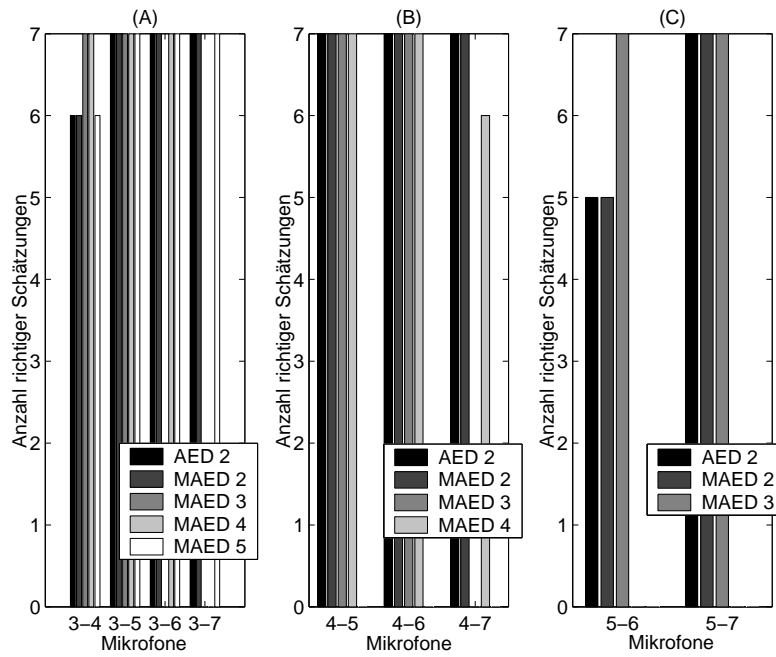


Abbildung 4.16: AED vs. MAED: Multimediarraum mit aufgezogenen Vorhängen

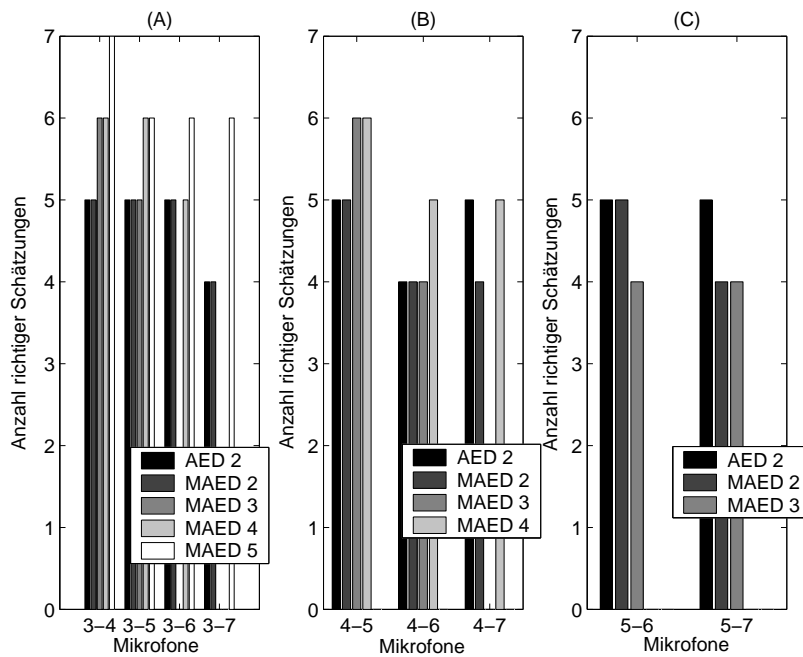


Abbildung 4.17: AED vs. MAED: Raum N5.17

Die Grafiken zeigen die Anzahl der als richtig klassifizierten Schätzungen auf-

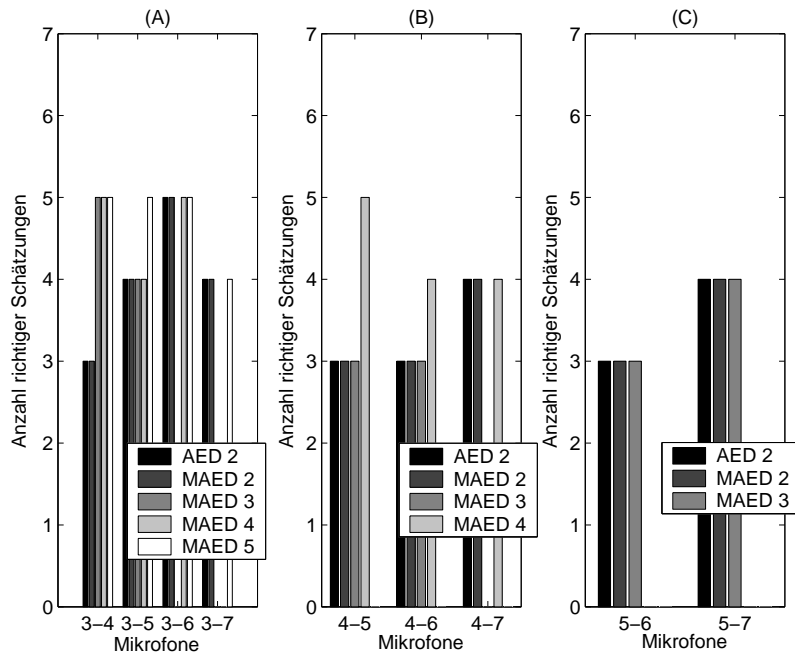


Abbildung 4.18: *AED vs. MAED: Garage*

getragen über dem Mikrofonpaar, an dem sie ermittelt wurden. Da pro Raum sieben Quellenpositionen betrachtet wurden, können die Algorithmen somit maximal sieben richtige Schätzungen aufweisen. Im Multimediaraum (Abbildungen 4.15 und 4.16) zeigen die Algorithmen ähnlich gute Genauigkeit bei der Laufzeitdifferenzschätzung. Die Abweichungen sind minimal. Da die Nachhallzeit im Multimediaraum jedoch nicht sehr hoch ist, war zu erwarten, dass alle Algorithmen, unabhängig von der Anzahl der verwendeten Kanäle, gut funktionieren würden. Mit den Impulsantworten aus dem Raum N5.17 und aus der Tiefgarage zeigen alle Algorithmen Schwächen auf. So geht die Zahl der richtig eingestuften Laufzeitdifferenzschätzungen insgesamt zurück.

Im Raum N5.17 zeigt der MAED 5 - Algorithmus ein besseres Verhalten (vgl. Abbildung 4.17 (A)), denn er schätzt alle Laufzeitdifferenzen mindestens einmal mehr richtig als die zweikanaligen Verfahren. Auch der MAED 4 - Algorithmus scheint besser mit der erhöhten Nachhallzeit zurechtzukommen als AED und MAED 2. So liegt er bei fünf von sechs Schätzungen einmal mehr richtig als AED

und MAED 2 (vgl. Abbildung 4.17 (A) und (B)). Der MAED 3 - Algorithmus hingegen ist gegenüber AED und MAED 2 nicht besser. Zwar hat er bei den Mikrofonpaaren 3-4 und 4-5 eine richtige Schätzung mehr, allerdings bei den Mikrofonpaaren 5-6 und 5-7 eine richtige Schätzung weniger als AED und MAED 2 (vgl. Abbildung 4.17).

Abbildung 4.18 zeigt das Ergebnis das mit den Impulsantworten aus der Tiefgarage gewonnen wurde. Es ist ein allgemeiner Rückgang der Treffergenauigkeit gegenüber dem Raum N5.17 zu beobachten, der auf den erhöhten Nachhall zurückzuführen ist. Die Tendenzen bezüglich der Performanz, wie sie mit den Impulsantworten im Raum N5.17 beobachtet wurden, lassen sich auch hier, wenn auch in geschwächerter Form erkennen. In zwei von vier Fällen ist der MAED 5 - Algorithmus um eine richtige Schätzung besser als AED und MAED 2 (vgl. Abbildung 4.18 (A)). MAED 4 schafft im Vergleich zu AED und MAED 2 in drei von sechs Fällen eine Schätzung mehr richtig (vgl. Abbildung 4.18 (A) und (B)). MAED 3 kann nur beim Mikrofonpaar 3-4 einmal mehr den Laufzeitunterschied richtig schätzen als AED und MAED 2.

Es ist zu bemerken, dass AED und MAED 2 bis auf zwei Fälle (siehe Abbildung 4.17 (B) und (C)) gleiche Lesitungsfähigkeit aufweisen. Diese minimale beobachtete Abweichung kann zufällig sein und daher wird davon ausgegangen, dass der MAED-Algorithmus mit zwei Kanälen ebenso funktioniert, wie der AED-Algorithmus.

Zusammenfassend lässt sich sagen, dass eine leichte Tendenz dahingehend aufgezeigt wurde, dass der mehrkanalige Adaptionsalgorithmus in Szenarien mit erhöhtem Nachhall etwas besser funktioniert als der mit zwei Kanälen arbeitende AED-Algorithmus. Kritisch muss allerdings an dieser Stelle erwähnt werden, dass die hier durchgeführten Simulationen nicht ausreichen um eine eindeutige Aussage zu treffen. In einer Vielzahl von Versuchen, die in dieser Arbeit nicht explizit dargestellt werden, die allerdings nötig waren um die Funktion der Algorithmen zu überprüfen und ihre Schrittweiten anzupassen, konnte beobachtet werden,

dass der zweikanalig arbeitende AED-Algorithmus oft gleich gut funktionierte, wie der Mehrkanal-AED-Algorithmus. Um die sich hier andeutende Tendenz zu untermauern, müsste eine Vielzahl von weiteren Quellenpositionen untersucht werden. Erst so kann ausgeschlossen werden, dass es sich bei den hier aufgezeigten Fällen, in denen der MAED besser funktioniert als AED, nicht um Ausnahmen handelt. Weitere unterschiedliche Sprachsignale könnten ebenso dazu beitragen eine stärker Aussagekraft zu erhalten.

## 4.4 Abschließende Bemerkungen

In diesem Kapitel wurden bei weitem nicht alle im Rahmen dieser Arbeit entstandenen Ergebnisse explizit dargestellt. Vielmehr wurde versucht die Aussagen die über die Leistungsfähigkeit der Algorithmen getroffen wurden an Hand von ausgewählten Beispielen zu belegen.

Es sei hier auch erwähnt, dass es durchaus Fälle gibt, in denen das MCC-Verfahren unter realistischen Bedingungen nicht schlechter abschneidet als die GCC-Methode (PHAT). Eine offene Fragestellung bleibt auch, in wie fern sich eine Verbesserung der Implementierung der Testalgorithmen auf die Leistungsfähigkeit des MCC-Algorithmuses mit Sprache auswirken könnte. In dieser Arbeit wurde keine Bandpassfilterung des Sprachsignales vorgenommen. Auf den Einsatz eines Sprachpausendetektors wurde ebenso verzichtet. Desweiteren könnte untersucht werden, wie sich eine Veränderung der Blockgröße oder die Verwendung unterschiedlicher Fenster bei der Segmentierung auswirken (hier wurde stets ein Rechteckfenster benutzt).

Es mag ebenso als hart empfunden werden, dass bei den Korrelationsmethoden eine Schätzung als falsch klassifiziert wurde, wenn sie nicht exakt mit dem wahren Zeitunterschied übereinstimmte. Selbstverständlich könnte man durch die Einführung einer Toleranzgrenze, die einen gewissen Fehler von wenigen Samples zulässt, ein besseres Abschneiden der Korrelationsmethoden erreichen. Allerdings



könnten dann die Tendenzen wie sie hier aufgezeigt wurden, nämlich dass das Mehrkanal-Korrelationsverfahren mit zunehmender Mikrofonzahl dazu neigt, den Laufzeitunterschied falsch zu schätzen im verborgenen bleiben, vor allem dann, wenn der wahre Laufzeitunterschied recht gering ist (Nahe bei Null). Bei einer Vielzahl von Simulationen fiel auf, dass die Korrelationsmethoden oft fälschlicherweise den Laufzeitunterschied zu Null schätzten (vgl. z.B. Abbildung 4.10). Auf einen direkten Vergleich der adaptiven Algorithmen mit den Korrelationsmethoden wurde letzten Endes verzichtet. Gerade weil die adaptiven Algorithmen unter realistischen Bedingungen gut funktionieren und die Korrelationsmethoden weniger gut, ist ein aussagekräftiger Vergleich schwierig. Weiteren darf kritisch bemerkt werden, dass die Algorithmen nur mit einem Sprachsignal getestet wurden, und nicht mit einer Auswahl vieler verschiedener Sprachsignale, die zudem länger sind als vier Sekunden.

# Kapitel 5

## Zusammenfassung und Ausblick

In den vorangegangenen Kapiteln wurden zwei neue Verfahren zur Laufzeitdifferenzschätzung vorgestellt, die mehr als zwei Mikrofone verwenden, das MCC-Verfahren und der MAED-Algorithmus.

Der Mehrkanalkorrelations-Algorithmus kann als Verallgemeinerung und Erweiterung der zweikanaligen GCC-Methode verstanden werden. Die Korrelationsverfahren stützen sich auf das ideale Schallausbreitungsmodell und haben somit Schwierigkeiten in realen Anwendungen mit erhöhtem Nachhall den Laufzeitunterschied zwischen zwei Mikrofonen richtig zu schätzen. Das MCC-Verfahren zeigte unter idealisierten Bedingungen eine deutlich bessere Leistungsfähigkeit bei zunehmendem Nachhall wie das GCC-Verfahren, das mit der PHAT, als eines seiner Varianten, das Referenzverfahren darstellte (vgl. Abschnitt 4.2.1). Auch in Szenarien mit einem niedrigen SNR konnte das MCC-Verfahren besser abschneiden. Allerdings ist das MCC-Verfahren unter realen Bedingungen nicht weniger unzuverlässig als die GCC-Methode, was in Abschnitt 4.2.2 aufgezeigt werden konnte. Die Ursache hierfür wird in der nicht hundertprozentigen Einhaltung der Fernfeldbedingungen gesehen. Eine hochgenaue Kalibrierung und Justierung der Mikrofone wäre unbedingt notwendig. Mit Sprache als Testsignal konnte keine zuverlässige Aussage getroffen werden, da dies in der Arbeit nur exemplarisch untersucht wurde und die Tests sehr widersprüchliche Ergebnisse lieferten.

Eine Möglichkeit zur Verbesserung der Leistungsfähigkeit des MCC-Verfahrens mit Sprache könnte ein Sprachpausendetektor bringen. Desweiteren müsste der Einfluss der verwendeten Blockgröße bei der Segmentierung und die Anwendung von frequenzbandbegrenzenden Filtern untersucht werden.

Der AED-Algorithmus und seine Erweiterung auf mehrere Kanäle stützen sich auf das allgemeine Schallausbreitungsmodell. Beide Algorithmen wurden unter realistischen Bedingungen mit Hilfe von gemessenen Impulsantworten aus verschiedenen Räumen an der Universität Erlangen-Nürnberg und mit einem Sprachsignal getestet und verglichen. Sie zeigen in Umgebungen mit wenig bis mittleren Nachhall ( $T_{60} \leq 350\text{ms}$  im Multimediaraum) eine sehr hohe Zuverlässigkeit bei der Laufzeitdifferenzschätzung. Mit zunehmendem Nachhall verschlechterte sich die Schätzgenauigkeit von AED und MAED gleichermaßen, doch konnte eine leichte Tendenz des MAED beobachtet werden in stark halligen Umgebungen etwas besser zu funktionieren. Für die Aussage, dass mehrere Kanäle nun wirklich eine Robustheitssteigerung gegenüber Nachhall nachsichziehen, fehlt es dieser Arbeit allerdings an Beweiskraft, da nur eine begrenzte Zahl an Quellenpositionen verwendet wurde. Eine Vielzahl weiterer Untersuchungen mit unterschiedlichen Szenarien und mehr unterschiedlichen Sprachsignalen könnte diese offene Fragestellung beantworten.

Die Korrelationsmethoden und adaptiven Algorithmen wurden nicht direkt miteinander verglichen. Dennoch zeigt vor allem das bessere Abschneiden des AED- und MAED-Algorithmuses unter realen Bedingungen, dass diese sich offenbar besser zur Laufzeitdifferenzschätzung eignen als das GCC- und MCC-Verfahren.

# Anhang A

## Matlabfunktionen

Die wichtigsten Matlabfunktionen die zum Erstellen dieser Arbeit verwendet wurden sind hier aufgeführt.

<code>AED.m</code>	Implementierung des zweikanaligen AED-Algorithmus. Von Herbert Buchner und Jochen Stenglein.
<code>calcdet2.m</code>	Implementierung des MCC-Algorithmus. Liefert die Determinante der Kreuzkorrelationskoeffizientenmatrix und schätzt den Laufzeitunterschied zwischen Eingangssignal 1 und 2. Maximal können bis zu fünf Eingangssignale übergeben werden.
<code>correlate.m</code>	berechnet die Kreuzkorrelation zweier Vektoren mit beliebigem Shiftfaktor. Implementierung im Zeitbereich nach [8].
<code>ir5.m</code>	erstellt fünf, auf Fernfeldannahme getrimmte Impulsantworten mit der Spiegelquellenmethode.
<code>multiaED5.m</code>	Implementierung des MAED-Algorithmus. Berechnet die Laufzeitunterschiede zwischen dem ersten und den n-ten Eingangssignalvektoren. Die Anzahl der Eingangssignale ist prinzipiell unbegrenzt.

# Literaturverzeichnis

- [1] J. Chen, Y. Huang, J. Benesty, “Time Delay Estimation“, *Audio Signal Processing*, chap. 8: 197-227, Kluwer Academic Publishers, 2004
- [2] Y. Huang and J. Benesty, “Adaptive Multichannel Time Delay Estimation Based on Blind System Identification for Acoustic Source Localization“, *Adaptive Signal Processing: Applications to Real-World Problems*, chap. 8: 227-247, Springer, 2003
- [3] Jacob Benesty, Yiteng Huang and Gary W. Elko, “Microphone Arrays for Video Camera Steering“, in Steven Gay and Jacob Benesty (Editors): *Acoustic Signal Processing for Telecommunication*, chap. 11: 239-259, Kluwer Academic Publishers, Boston, 2000
- [4] Gish and Cochran, “Generalized Coherence“, *International Conference on Acoustics, Speech, and Signal Processing*, vol. 5: 2745-2748, New York, 1988
- [5] J.B. Allen and D.A. Berkeley, “Image Method for Efficiently Simulating Small-Room Acoustics“, *Journal of the Acoustical Society of America*, vol. 65(4): 943-950, April 1979
- [6] Bronstein, Semendjajew, u.a., “Taschenbuch der Mathematik“, Verlag Harri Deutsch, 5. Auflage, 2001

- [7] Roland Barthel, “Ein robustes akustisches Quellenortungsverfahren für multimediale Anwendungen in verhallter Umgebung“, Diplomarbeit, Erlangen, 2000
- [8] Walter Kellermann, “Digital Signal Processing“, Skript zur Vorlesung Digitale Signalverarbeitung, Erlangen, Sommer 2003