

FRIEDRICH-ALEXANDER-UNIVERSITÄT

ERLANGEN-NÜRNBERG

LEHRSTUHL FÜR MULTIMEDIAKOMMUNIKATION UND
SIGNALVERARBEITUNG

PROF. DR.-ING. WALTER KELLERMANN

Studienarbeit

Die Global Coherence Field Methode und
deren Anwendung für die akustische
Quellenlokalisierung

von Andreas Tautz

Mai 2016

Betreuer: M.Sc. Hendrik Barfuß

ERKLÄRUNG

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

Ort, Datum

Unterschrift

Inhaltsverzeichnis

1	Einleitung	1
2	Signalmodell	3
2.1	Impulsantwort	3
2.2	Lokalisierung mittels TDOA	5
3	Global Coherence Field Methode	7
3.1	Generalized Cross Correlation - Phase Transform	7
3.2	Global Coherence Field Methode	10
3.2.1	Prinzip	10
3.2.2	Implementierung	12
3.3	Programmablauf	13
3.4	Berechnung der GCF-Map	13
4	Simulationsparameter	16
4.1	Blocklänge	16
4.2	Fensterfunktion	19
4.3	Abtastrate	20
4.4	Abtastratenerhöhung	21
4.5	Gitterabstand	23
4.6	Mittelung	24
4.7	Rauschschwelle	26

INHALTSVERZEICHNIS

5 Experimente	28
5.1 Szenarien mit einer Schallquelle	28
5.1.1 Lineares Array zur Abstandsmessung	28
5.1.2 Simulierte Reflexion	30
5.1.3 3D Fall und mehrteilige Arrays	33
5.1.4 Kompakte Arrays und Nachhall	35
5.1.5 Messung am DICIT Array	37
5.2 Szenarien mit mehreren Schallquellen	41
5.2.1 Dominante Schallquelle	41
5.2.2 GCC-PHAT Maskierung	43
5.2.3 Alternatives Maskierungsverfahren	46
6 Zusammenfassung	54
Literaturverzeichnis	56

Abbildungsverzeichnis

2.1	Signalmodell der Mikrofonsignale	3
2.2	Lokalisierung mittels TDOA	5
3.1	Programmablauf inkl. Maskierung von Schallquellen	14
4.1	GCF-Map mit Blocklänge 512 bei 0dB SNR	17
4.2	GCF-Map mit Blocklänge 8192 bei 0dB SNR	18
4.3	RMS Fehler bei verschiedenen Blocklängen und SNR	19
4.4	Auswirkung verschiedener Fensterfunktionen	19
4.5	GCF-Map bei $f_s=16\text{kHz}$	21
4.6	GCF-Map bei $f_s=16\text{kHz}$ und 8facher Abtastratenerhöhung	22
4.7	Gitterabstand 10cm: Fehler 35cm	23
4.8	Fehler nach Verschiebung der Quelle um 5cm nur noch 5cm	24
4.9	Messfehler bei verschiedenen T_G und T_P	26
5.1	Erzeugung der Spiegelschallquellen	31
5.2	GCF-Map mit Spiegelschallquellen	32
5.3	GCF-Map ohne Spiegelschallquellen	32
5.4	Abbildung einer vertikal verschobenen Schallquelle auf die GCF-Map	33
5.5	GCF-Maps bei vertikal versetzter Schallquelle um 2m(!) mit 1 und 2 Arrays	34
5.6	Anordnung der 5 Messpositionen	35
5.7	Anordnung der Messpositionen und des Mikrofonarrays	38

ABBILDUNGSVERZEICHNIS

5.8 Fehler bei Nutzung der Mikrofone 4-12 bei einer Breite von 64cm	39
5.9 Fehler bei Nutzung der Mikrofone 3-13 bei einer Breite von 1,28m	39
5.10 Fehler bei Nutzung aller Mikrofone bei einer Breite von 2,56m . .	40
5.11 Fehler bei Nutzung des linearen Teilarrays bei einer Breite von 2,56m	41
5.12 Amplitudenverlauf zweier Sprachquellen	42
5.13 Maskierungsfunktion $\phi(\Delta k, 0)$ für verschiedene p und b	44
5.14 GCF-Map vor und nach der Maskierung	44
5.15 GCF-Map vor Anwendung der Maskierung	48
5.16 Maskierung nach Brutti et al. (5.3)	48
5.17 Vorgeschlagene Maskierung	49
5.18 Vergleich der Kreuzkorrelationsfunktionen	50
5.19 Vergleich der Kreuzkorrelationsfunktionen ohne Abtastratenerhöhung	51

Tabellenverzeichnis

5.1	Verschiedene Mikrofonanordnungen	29
5.2	Abstandsfehler[m] bei verschiedene Messpositionen	30
5.3	Messwerte mit und ohne Spiegelschallquellen	33
5.4	Winkelfehler[°] bei verschiedenen T_{60} mit beiden Arrays	36
5.5	Abstandsfehler[m] bei verschiedenen T_{60} mit beiden Arrays	36
5.6	Mikrofonanordnung DICIT Array	38
5.7	Messwerte dominante Schallquellen	42
5.8	Messwerte nach Suche des zweiten Maximums	45
5.9	Verdoppelte Blocklänge auf 256ms mit $T_G=0,2s$ und $0,5s$	46
5.10	Messwerte bei verschiedenen Maskierungsfunktionen	50
5.11	Messwerte bei verschiedenen Maskierungsfunktionen mit $U=1$	52

KURZFASSUNG

Die Global Coherence Field Methode (GCF) ist eine Methode zur Lokalisierung von Schallquellen mithilfe von Mikrofonarrays. Mit ihr werden akustische Karten (GCF-Maps) eines festgelegten Raumbereichs berechnet und damit die Positionen der Schallquellen bestimmt. Simulierte und gemessene Szenarien haben gezeigt wie die GCF auf Veränderungen der Berechnungsparameter und auf unterschiedliche akustische Situationen reagiert. Die GCF ist robust gegenüber Nachhall und Reflexion. Durch geeignete Parameterwahl ist auch die Nutzung bei schlechtem Signal-Rausch-Verhältnis möglich. Durch Nachbearbeiten der GCF-Maps ist es möglich mehrere Schallquellen gleichzeitig zu orten. In dieser Arbeit wird gezeigt wie sich die Wahl der Berechnungsparameter auf die Positionsbestimmung mittels GCF auswirkt und wie gut sie in simulierten und gemessenen Szenarien funktioniert. Zusätzlich wird eine neue Methode zur Ortung simultaner Schallquellen gezeigt und wann sie gegenüber bereits bekannten im Vorteil ist.

Formelzeichen

M	Anzahl der Mikrofone
s_n	Quellsignal n
$x_m(t), \mathbf{x}_m(\mathbf{k})$	Mikrofonsignal Mikrofon m
t	Zeit
n_m	Rauschsignal an Mikrofon m
$h_{nm}(t), H_{nm}(\omega)$	Raumimpulsantwort von Quelle n zu Mikrofon m im Zeit- und Frequenzbereich
ω	Kreisfrequenz
d, d_{12}	Abstandsdifferenz, Abs. zwischen Mikrofon 1 und 2
Δt	Laufzeitunterschied, Zeitabstand
c_s	Schallgeschwindigkeit
M_m	Mikrofon m
\mathcal{H}_{12}	Hyperbel zugehörig zu Mikrofon 1 und 2
P	Position einer Schallquelle
D, \hat{D}	Zeitverzögerung, Schätzung der Zeitverzögerung
α	Koeffizient
j	Imaginäre Einheit
$X_1(\omega), X_2(\omega)$	Fouriertransformation von $x_1(t), x_2(t)$
$H_1(\omega), H_2(\omega)$	Vorfaktoren für GCC
$\Psi(\omega), \Psi_{PHAT}(\omega)$	Vorfilter für GCC, GCC-PHAT
$G_{x_1x_2}(\omega), \hat{G}_{x_1x_2}(\omega)$	Kreuzleistungsdichtespektrum von x_1 und x_2 , Schätzung von $G_{x_1x_2}(\omega)$

FORMELZEICHEN

D_{12}	Abstand zwischen Mikrofon 1 und 2
f_s	Abtastrate
$y(t, q), Y(\omega, q)$	Ausgang eines Beamformers gerichtet auf q im Zeit- und Frequenzbereich
$P(q)$	Leistung eines Beamformers gerichtet auf q
q	Ort auf der GCF-Map
\hat{q}_s	Schätzung für Ort einer Schallquelle
$\Delta_n(q), \Delta_l(q), \Delta_k(q)$	Signallaufzeit von q zu Mikrofon n, l, k
$P_{GCF}(q)$	Leistung der GCF-Map am Ort q
$\hat{P}_{GCF}(q)$	Berechnete Leistung der GCF-Map am Ort q
\hat{q}_{GCF}	Ort des Maximums einer GCF-Map
$\text{round}(z)$	Rundungsoperator für reellwertiges Argument z
$\delta(t)$	Dirac-Funktion
$R(\tau)$	Kreuzkorrelationsfunktionen
$\hat{R}(\tau)$	Schätzung einer Kreuzkorrelationsfunktion
τ	Zeitvariable einer Kreuzkorrelationsfunktion
T	Zeitintervall
$\mathbf{R}_{12}(\Delta k), \mathbf{R}_{AB}(\Delta k)$	Berechnete Kreuzkorrelationsfunktion
Δk	Abtastwert einer Kreuzkorrelationsfunktion
N	Blocklänge in Abtastwerten
F_{RMS}	Quadratischer Mittelwert der Messfehler
Z	Anzahl der Einzelmessungen
U	Faktor der Abtastratenerhöhung
v_G, v_P	Vergessensfaktor
T_G	Zeitkonstante zur Mittelung der GCF-Map
T_P	Zeitkonstante zur Mittelung von Koordinaten
x_n, y_n	n -te gemessene x -, y -Koordinate
p_n	n -tes gemessenes Maximum
$\overline{W}_n, \overline{x}_n, \overline{y}_n, \overline{p}_n$	Mittelwert von W_n, x_n, y_n, p_n

FORMELZEICHEN

S	Schwellwert
M_P	Anzahl der Mikrofonpaare
y_P, z_P	y -, z -Koordinate von Punkt P
y'_P	gemessene y -Koordinate von Punkt P
$\phi(\Delta k, \nu)$	Maskierungsfunktion für Zeitdifferenz ν
ν	Zeitdifferenz
$R'_{AB}(\Delta k)$	Maskierte Kreuzkorrelationsfunktion
$\Delta_{AB}(q_{GCF})$	Zeitdifferenz bei Mikrofonpaar AB und Ort q_{GCF}
p, b	Parameter für Maskierungsfunktion
\mathbf{m}	Mitte für Maskierung einer Kreuzkorrelationsfunktion
\mathbf{k}	Variable für Abtastwerte

Kapitel 1

Einleitung

Es gibt vielerlei Anwendungen bei denen es hilfreich ist, die Positionen von Schallquellen zu kennen. Helfen kann diese Information zum Beispiel bei der Lärmbekämpfung und akustischer Überwachung. Für die Steuerung von Video- und Tonaufzeichnungen bei Auftritten vor Publikum, bei Sprachsteuerungssystemen, Videokonferenzen oder bei der Protokollierung von Sitzungen kann die Kenntnis über die Positionen von Sprechern genutzt werden.

Hierbei geht es darum, Positionsinformationen zu nutzen, um Sprecher zu unterscheiden, simultan gesprochene Sequenzen zu trennen und Störgeräusche zu unterdrücken. Erfolgt diese in Echtzeit, können damit Kameras ausgewählt und ausgerichtet werden, sowie akustische Beamformer zur Sprachübertragung auf die entsprechenden Quellen ausgerichtet werden. Diese Aufgaben würden von einer der Lokalisierung nachgeschalteten Signalverarbeitung, welche die Positionsdaten interpretiert, übernommen werden. Dieser Teil der Signalverarbeitung wird jedoch nicht in dieser Arbeit behandelt werden.

Die Global Coherence Field Methode (GCF) ist eine Methode zur Lokalisierung von Schallquellen mithilfe von Mikrofonarrays. Mit ihr berechnet man eine akustische Karte eines gewählten Raumbereichs. Aus dieser Karte lassen sich die Positionen einzelner oder mehrere simultaner Schallquellen bestimmen. In dieser Studienarbeit wird die GCF in MATLAB implementiert und ihr Verhalten

KAPITEL 1. EINLEITUNG

in vollständig simulierten Szenarien und mit in verschiedenen Räumen gemessenen Raumimpulsantworten untersucht. Ebenfalls werden verschiedene Verfahren gezeigt, wie mit der GCF mehrere simultan auftretende Schallquellen geortet werden können.

Kapitel 2

Signalmodell

2.1 Impulsantwort

Die in dieser Studienarbeit durchgeführten Simulationen basieren auf dem folgenden Signalmodell. Den angedachten Einsatzfällen entsprechend betrachten wir eine oder mehrere akustische Quellsignale, deren Inhalt und Ursprungsort für die spätere Ortung als unbekannt angenommen werden.

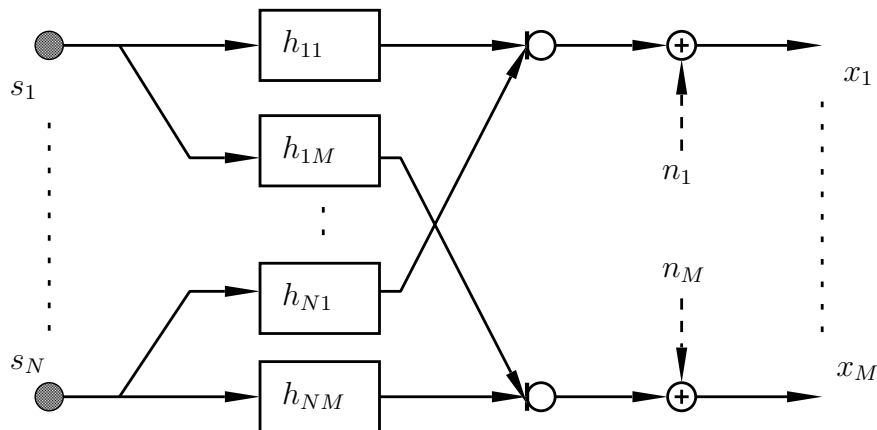


Abbildung 2.1: Signalmodell der Mikrofonsignale

Diese Signale s_1 bis s_N , dargestellt in Abbildung 2.1, treffen gefaltet mit ihren Raumimpulsantworten h_{11} bis h_{NM} an allen Mikrofonen ein. Dort überlagern sie sich mit den Rauschsignalen n_1 bis n_M zu den Mikrofonsignalen x_1 bis x_M .

KAPITEL 2. SIGNALMODELL

Für die Mikrofon-signale gilt:

$$x_m(t) = n_m + \sum_{n=1}^N s_n(t) * h_{nm}(t). \quad (2.1)$$

Die Rauschsignale ergeben sich aus Mikrofon-, Verstärker- und Wandlerrauschen. Sie werden als zueinander unkorreliertes weißes Rauschen angenommen bzw. simuliert.

Die Raumimpulsantworten beschreiben den Zusammenhang zwischen den ursprünglichen Quellsignalen und den am Mikrofon eintreffenden Signalen. Dieser beinhaltet die Abnahme der Amplitude mit dem Abstand und die dazugehörige Verzögerung. Dies gilt sowohl für den Direktschall als auch für einfache und mehrfache Reflexionen an Decke, Boden und Wänden (Nachhall) des betrachteten Raumes.

Werden die Raumimpulsantworten $h_{nm}(t)$ simuliert und nicht gemessen, bestehen sie aus einer um $\frac{D_{nm}}{c_s}$ verzögerten δ -Funktion und berücksichtigt die quadratische Leistungsabnahme mit dem Abstand D_{nm} zwischen der Quelle n und dem Mikrofon m durch die Verringerung der Amplitude um den Faktor $\frac{1}{D_{nm}}$:

$$h_{nm}(t) = \frac{1}{D_{nm}} \delta\left(t - \frac{D_{nm}}{c_s}\right). \quad (2.2)$$

Dabei steht c_s für die Schallgeschwindigkeit, die mit 343ms^{-1} bei 20°C simuliert wird.

Da die Simulation mit zeitdiskreten Signalen arbeitet, aber dennoch beliebige Verzögerungen vorkommen, erfolgt die Berechnung der verzögerten Quellsignale im Frequenzbereich. Verwendet werden dazu die Raumimpulsantworten $h_{nm}(t)$ im Frequenzbereich mit

$$H_{nm}(\omega) = \frac{1}{D_{nm}} e^{-j\omega \frac{D_{nm}}{c_s}}. \quad (2.3)$$

Die Verzögerung um $\frac{D_{nm}}{c_s}$ erfolgt nach Anwendung der diskreten Fouriertransformation (DFT) auf das Quellsignal durch eine Multiplikation mit $e^{-j\omega \frac{D_{nm}}{c_s}}$. Nach der Rücktransformation in den Zeitbereich wird anschließend das simulierte Rauschsignal dazuaddiert um die Mikrofon-signale zu erhalten.

2.2 Lokalisierung mittels TDOA

Da der Zeitpunkt und Inhalt der ausgesandten Signale einer Schallquellen nicht im vornherein bekannt sind, ist es nicht möglich ihren Abstand zu den Mikrofonen direkt über die Signallaufzeit zu messen. Betrachtet man zwei Mikrofone kann man durch Vergleichen der beiden aufgenommenen Signale die Differenz der Ankunftszeit an beiden Mikrofonen messen.

Mehrere Messungen der Differenz der Ankunftszeit (engl.: „Time Difference of Arrival“, TDOA) an verschiedenen Mikrofonpaaren können zur Positionsbestimmung der Schallquelle genutzt werden. Mit

$$d = \Delta t \cdot c_s \quad (2.4)$$

lässt sich die Abstandsdifferenz d aus dem Laufzeitunterschied Δt und der Schallgeschwindigkeit c_s berechnen.

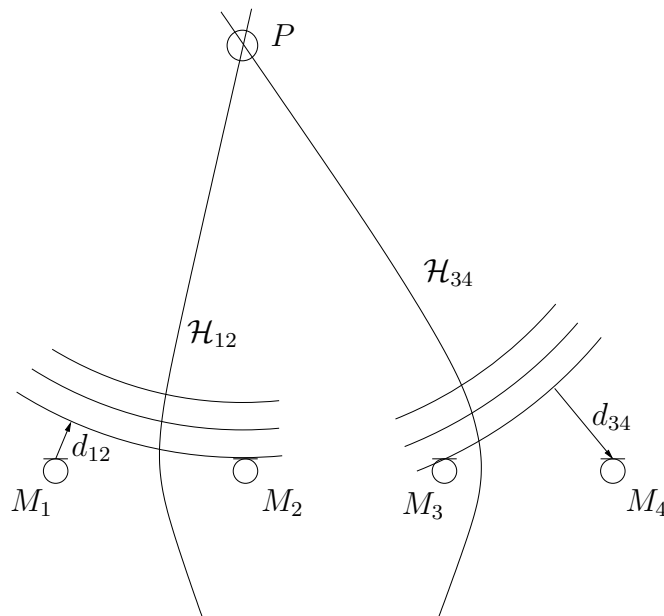


Abbildung 2.2: Lokalisierung mittels TDOA

Abbildung 2.2 zeigt die übliche Methode zur Lokalisierung mittels TDOA. Über Laufzeitunterschiede misst man die Abstandsdifferenzen d_{12} für das Mikrofonpaar (M_1, M_2) und d_{34} für das Mikrofonpaar (M_3, M_4) . Alle Punkte, die

KAPITEL 2. SIGNALMODELL

für die Mikrofone M_1 und M_2 dieselbe Abstandsdifferenz d_{12} haben, liegen auf der Hyperbel \mathcal{H}_{12} . Für die Mikrofone M_3 und M_4 liegen sie entsprechend auf der Hyperbel \mathcal{H}_{34} . Der Schnittpunkt P beider Hyperbeln ist die mittels TDOA bestimmte Position der Schallquelle.

Hat man wie im Beispiel ein lineares Array, bei dem alle Mikrofone auf einer Geraden liegen, schneiden sich die Hyperbeln auf beiden Seiten des Arrays. Für eine eindeutige Positionsbestimmung muss daher eine Seite von vornherein ausgeschlossen werden - z.B. durch die Lage des Arrays im Raum.

Für eine Positionsbestimmung in einer Ebene werden mindestens 2 Mikrofonpaare benötigt. Im dreidimensionalen Fall erhält man Hyperboloiden anstelle von Hyperbeln, und man benötigt mindestens 3 Mikrofonpaare. Jedes Mikrofon kann Teil mehrerer Mikrofonpaare sein. Meistens werden mehr als die benötigte Zahl an Mikrofonen genutzt, um eine bessere Positionsbestimmung bei vorhandenem Rauschen und Nachhall zu erreichen.

Für die Messung der Zeitdifferenzen wird in dieser Implementierung die Generalized Cross Correlation - Phase Transform (GCC-PHAT) eingesetzt. Sie wird im folgenden Kapitel beschrieben.

Kapitel 3

Global Coherence Field Methode

3.1 Generalized Cross Correlation - Phase Transform

Zwei gemessene Signale $x_1(t)$ und $x_2(t)$ mit der Verzögerung D können mit

$$x_1(t) = s_1(t) + n_1(t) \quad (3.1)$$

$$x_2(t) = \alpha s_1(t + D) + n_2(t) \quad (3.2)$$

beschrieben werden. Dabei wird das Quellsignal $s_1(t)$ und die additiven Rauschsignale $n_1(t)$ und $n_2(t)$ als zueinander unkorreliert angenommen. Eine übliche Methode zur Schätzung von D aus $x_1(t)$ und $x_2(t)$ nutzt die Schätzung der Kreuzkorrelationsfunktion

$$\hat{R}_{x_1x_2}(\tau) = \frac{1}{T - \tau} \int_{\tau}^T x_1(t)x_2(t - \tau) dt \quad (3.3)$$

über ein Intervall T . Die Stelle des Maximums in $\hat{R}_{x_1x_2}(\tau)$

$$\hat{D} = \underset{\tau}{\operatorname{argmax}} \hat{R}_{x_1x_2}(\tau) \quad (3.4)$$

ist eine Schätzung der Verzögerung D .

KAPITEL 3. GLOBAL COHERENCE FIELD METHODE

Die Kreuzkorrelationsfunktion lässt sich auch über die inverse Fouriertransformation des Kreuzleistungsdichtespektrums

$$G_{x_1x_2}(\omega) = X_1(\omega)X_2^*(\omega) \quad (3.5)$$

ermitteln. Dabei ist $X_1(\omega)$ mit

$$X_1(\omega) = \int_{-\infty}^{\infty} x_1(t)e^{-j\omega t} dt \quad (3.6)$$

die Fouriertransformierte von $x_1(t)$. $X_2^*(\omega)$ ist die komplex konjugierte Fouriertransformierte von $x_2(t)$. Die Kreuzkorrelationsfunktion lautet dann

$$R_{x_1x_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{x_1x_2}(\omega)e^{j\omega\tau} d\omega. \quad (3.7)$$

In der Praxis erhält man durch den begrenzten Beobachtungszeitraum nur eine Schätzung für $G_{x_1x_2}(\omega)$. Sie wird im weiteren mit $\hat{G}_{x_1x_2}(\omega)$ bezeichnet.

In [KC76] wird vorgeschlagen $X_1(\omega)$ und $X_2(\omega)$ vor der Berechnung von $R_{x_1x_2}(\tau)$ mit geeigneten Vorfaktoren $H_1(\omega)$ und $H_2(\omega)$ zu multiplizieren, um die Bestimmung von D zu vereinfachen. Dies nennt man Generalized Cross Correlation (GCC). Mit dem Vorfilter

$$\Psi(\omega) = H_1(\omega)H_2^*(\omega) \quad (3.8)$$

und der Schätzung für das Kreuzleistungsdichtespektrum $\hat{G}_{x_1x_2}(\omega)$ gilt dann

$$\hat{R}_{x_1x_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi(\omega)\hat{G}_{x_1x_2}(\omega)e^{j\omega\tau} d\omega. \quad (3.9)$$

Gewählt wird für die Global Coherence Field Methode (GCF) die Phasentransformation (GCC-PHAT) mit dem Vorfilter

$$\Psi_{PHAT}(\omega) = H_1(\omega)H_2^*(\omega) = \frac{1}{|\hat{G}_{x_1x_2}(\omega)|}. \quad (3.10)$$

Dieser sorgt dafür, dass das Produkt $\Psi_{PHAT}(\omega)\hat{G}_{x_1x_2}(\omega)$ ein weißes Frequenzspektrum erhält und $\hat{R}_{x_1x_2}(\tau)$ im Idealfall eine um D verschobene δ -Funktion

KAPITEL 3. GLOBAL COHERENCE FIELD METHODE

wird. Eine δ -Funktion ist durch ihr scharfes Maximum besonders gut zur Bestimmung von D geeignet. Durch diese Wahl von $\Psi_{PHAT}(\omega)$ werden aus $x_1(t)$ und $x_2(t)$ ausschließlich Phaseninformationen genutzt, um ihre zeitliche Relation zu bestimmen.

Die Verarbeitung der Mikrofonsignale in der Simulation findet im Frequenzbereich statt, um aufwendige Faltungsoperationen zu vermeiden. Aus den synchron abgetasteten Signalblöcken $\mathbf{x}_1(\mathbf{k})$ und $\mathbf{x}_2(\mathbf{k})$ wird mithilfe der Fast Fourier Transformation (FFT) die Schätzung der Kreuzkorrelationsfunktion $R_{12}(\Delta\mathbf{k})$ mit Phasentransformation durch

$$R_{12}(\Delta\mathbf{k}) \leftarrow \text{FFT}^{-1} \left(\frac{\text{FFT}(\mathbf{x}_1(\mathbf{k})) \cdot \text{FFT}^*(\mathbf{x}_2(\mathbf{k}))}{|\text{FFT}(\mathbf{x}_1(\mathbf{k})) \cdot \text{FFT}^*(\mathbf{x}_2(\mathbf{k}))|} \right) \quad (3.11)$$

berechnet. Die Multiplikation und Division erfolgen dabei elementweise mit den komplexwertigen Ergebnissen der FFTs.

Da zur Berechnung ein endlich langer Datenblock genutzt wird, sinkt die Anzahl der effektiv genutzten Abtastwerte mit zunehmendem Abstand zur Stelle $\Delta\mathbf{k}=0$ der Schätzung. Durch die Periodizität der FFT werden mit zunehmenden $|\Delta k|$ die Abtastwerte am linken Rand des einen Datenblocks mit den Abtastwerten am rechten Rand des jeweils anderen Datenblocks korreliert und tragen damit zu einer Verschlechterung der Schätzung bei. Die Blocklänge sollte deshalb deutlich länger sein als der benötigte Bereich der Kreuzkorrelationsfunktion. Die Länge des benötigten Bereichs $[\Delta k_{min}, \Delta k_{max}]$ entspricht mit

$$\Delta k_{min,max} = \mp \frac{D_{12} f_s}{c_s} \quad (3.12)$$

der doppelten Signallaufzeit zwischen beiden Mikrofonen. Dabei ist D_{12} der Abstand zwischen den Mikrofonen, f_s die Abtastrate und c_s die Schallgeschwindigkeit.

3.2 Global Coherence Field Methode

3.2.1 Prinzip

Die in [BW01] beschriebene Global Coherence Field Methode (GCF, dort „SRP-PHAT“) zur Positionsbestimmung einer Schallquelle basiert auf der Leistung („Steered Response Power“) eines gerichteten Beamformers unter Ausnutzung der GCC-PHAT und deren Robustheit unter realen akustischen Bedingungen.

Bei der GCF wird ein Beamformer auf alle möglichen Positionen einer Schallquelle gerichtet und die Signalleistungen bestimmt. Daraus erstellt man eine Karte (GCF-Map), die für jeden ihrer Gitterpunkte einen Leistungswert enthält. Die GCF nutzt im Prinzip die TDOA wie in Kapitel 2.2 beschrieben, jedoch werden die Laufzeitdifferenzen an den einzelnen Mikrofonpaaren nicht explizit ermittelt. Stattdessen werden die mit der GCC-PHAT berechneten Kreuzkorrelationsfunktionen genutzt um die GCF-Map zu berechnen. Durch eine Maximumsuche in der GCF-Map findet man den Gitterpunkt, der der Position der Schallquelle am ehesten entspricht.

Für den Ausgang $y(t, q)$ des Beamformers für den Gitterpunkt q gilt

$$y(t, q) = \sum_{n=1}^M x_n(t + \Delta_n(q)). \quad (3.13)$$

Dabei steht x_n für das Mikrofonsignal des n -ten Mikrofons und $\Delta_n(q)$ für die Signallaufzeit zwischen q und dem Mikrofon n . Für $y(t, q)$ erhält man im Frequenzbereich

$$Y(\omega, q) = \sum_{n=1}^M X_n(\omega) e^{j\omega \Delta_n(q)}. \quad (3.14)$$

Die Leistung für einen einzelnen Gitterpunkt q berechnet sich demnach mit

$$P(q) = \int_{-\infty}^{+\infty} Y(\omega, q) Y^*(\omega, q) d\omega. \quad (3.15)$$

Bereits jetzt vor der Anwendung der Phasentransformation erhalte man durch

KAPITEL 3. GLOBAL COHERENCE FIELD METHODE

die Suche des Ortes mit der maximalen Signalleistung einen geschätzten Quellort

$$\hat{q}_s = \operatorname{argmax}_q P(q) \quad (3.16)$$

für die zu ortende Schallquelle.

Setzt man Gleichung (3.14) in (3.15) ein erhält man

$$P(q) = \int_{-\infty}^{+\infty} \sum_{l=1}^M X_l(\omega) e^{j\omega\Delta_l(q)} \sum_{k=1}^M X_k^*(\omega) e^{-j\omega\Delta_k(q)} d\omega. \quad (3.17)$$

Eine Umstellung der Summen nach vorne führt zu

$$P(q) = \sum_{l=1}^M \sum_{k=1}^M \int_{-\infty}^{+\infty} X_l(\omega) X_k^*(\omega) e^{j\omega(\Delta_l(q) - \Delta_k(q))} d\omega. \quad (3.18)$$

Das Integral im Inneren der Doppelsumme in (3.18) entspricht bis auf die Normierung mit $\frac{1}{2\pi}$ einer Kreuzkorrelationsfunktion wie in Gleichung (3.7), jedoch von $X_l(\omega)$ und $X_k(\omega)$ an der Stelle $\tau = \Delta_l(q) - \Delta_k(q)$.

Fügt man in die Berechnung dieser Kreuzkorrelationsfunktion wie bei der GCC-PHAT das Vorfilter $\Psi_{PHAT}(\omega)$ ein, lautet die Gleichung dafür

$$R_{lk}(\Delta_l(q) - \Delta_k(q)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_{PHAT}(\omega) X_l(\omega) X_k^*(\omega) e^{j\omega(\Delta_l(q) - \Delta_k(q))} d\omega. \quad (3.19)$$

Das Vorfilter für die Phasentransformation lautet in diesem Fall

$$\Psi_{PHAT}(\omega) = \frac{1}{|X_l(\omega) X_k^*(\omega)|}. \quad (3.20)$$

Genauso wie bei der GCC-PHAT fügen wir dieses Vorfilter in die Gleichung des Beamformers ein. Damit besteht die GCF Methode aus der Erstellung einer in der Praxis ortsdiskreten GCF-Map mit der „Leistung“ des modifizierten Beamformers

$$P_{GCF}(q) = \sum_{l=1}^M \sum_{k=1}^M \int_{-\infty}^{+\infty} \Psi_{PHAT}(\omega) X_l(\omega) X_k^*(\omega) e^{j\omega(\Delta_l(q) - \Delta_k(q))} d\omega \quad (3.21)$$

und der anschließenden Suche des geschätzten Ortes der Schallquelle \hat{q}_{GCF} mit

$$\hat{q}_{GCF} = \operatorname{argmax}_q P_{GCF}(q). \quad (3.22)$$

In diesem Fall hat der Begriff „Leistung“ durch die Anwendung des nichtlinearen Vorfilters $\Psi_{PHAT}(\omega)$ jedoch keine physikalische Bedeutung mehr.

3.2.2 Implementierung

Um veränderliche und sich bewegende Schallquellen zu berücksichtigen, erfolgt die Signalverarbeitung blockweise. Dabei sind Signalblöcke in der Länge von Zweierpotenzen vorteilhaft, da für solche Längen die FFT besonders effizient berechnet werden kann und dadurch eine mögliche Verarbeitung in Echtzeit begünstigt wird. Während eines Durchlaufs werden also synchronisiert aufgenommene Signalblöcke $x_m[k]$ der Länge $N = 2^n$ für jedes einzelne Mikrofon m verarbeitet.

Aus den Gleichungen (3.19) und (3.21) ergibt sich für die Berechnung von $P_{GCF}(q)$ im Zeitbereich

$$P_{GCF}(q) = 2\pi \sum_{l=1}^M \sum_{k=1}^M R_{lk}(\Delta_l(q) - \Delta_k(q)) \quad (3.23)$$

mit der Kreuzkorrelationsfunktion R_{lk} für das Mikrofonpaar (l, k) und M Mikrofonen.

Anstelle von R_{lk} wird ihre Schätzung \hat{R}_{lk} blockweise wie in (3.11) berechnet. Es ist nicht nötig Autokorrelationen von Mikrofonsignalen zu berechnen, da sie lediglich einen Gleichanteil zur GCF-Map beitragen würden. Jede mögliche Kombination aus verschiedenen Mikrofonen l und k wird nur einmal berücksichtigt, da

$$\hat{R}_{lk}(\Delta_l(q) - \Delta_k(q)) = \hat{R}_{kl}(\Delta_k(q) - \Delta_l(q)). \quad (3.24)$$

Die Normierung mit 2π wird ignoriert. Die Kreuzkorrelationsfunktion \hat{R}_{lk} ist ebenso wie die Eingangssignale ein zeitdiskretes Signal mit der Abtastrate f_s . Die zum Ort q gehörige Stelle muss daher gerundet werden. Damit wird die tatsächlich berechnete GCF-Map mit

$$\hat{P}_{GCF}(q) = \sum_{(l,k)} \hat{R}_{lk} \left(\frac{1}{f_s} \text{round} \left(f_s (\Delta_l(q) - \Delta_k(q)) \right) \right) \quad (3.25)$$

und dem Rundungsoperator

$$\text{round}(z) = \text{sgn}(z) \lfloor |z| + 0,5 \rfloor \quad (3.26)$$

berechnet. Der Operator `round` rundet sein Argument auf die nächstgelegene Ganzzahl. Die realisierte Berechnung von $\hat{P}_{GCF}(q)$ in Gleichung (3.25) ist daher nur eine Näherungslösung für $P_{GCF}(q)$.

Die Abstände der Mikrofone zu allen Gitterpunkten q werden nur einmalig berechnet und in Tabellen abgelegt. Damit und mit der Schallgeschwindigkeit werden dann die Signallaufzeiten $\Delta_k(q)$ und $\Delta_l(q)$ berechnet.

3.3 Programmablauf

Das MATLAB Programm dient zur Simulation einer möglichen Echtzeitanwendung. Zunächst werden alle Mikrofonensignale vollständig berechnet. Dies erfolgt wahlweise aus der Geometrie eines frei wählbaren Szenarios mit gegebenen Sprachsignalen oder aus der Faltung bekannter Raumimpulsantworten mit den Sprachsignalen. Gemessene Mikrofonensignale können auch direkt geladen werden.

Diese Mikrofonensignale werden dann so lange in der Hauptschleife des Programms abgearbeitet, bis ein gewählter Zeitpunkt erreicht ist oder das Ende der Signale erreicht wurde. Die Verarbeitung erfolgt in Blöcken, die sich überlappen können. Für jeden dieser Blöcke wird eine GCF-Map berechnet und ausgewertet. Die Ergebnisse werden für jeden Zeitabschnitt abgespeichert und gegebenenfalls während der Programmausführung dargestellt. Es wird keine konkrete Anwendung simuliert. Es besteht die Möglichkeit Sprechpausen vor der Berechnung aller Mikrofonensignale im genutzten Beispielsignal entfernen zu lassen.

3.4 Berechnung der GCF-Map

Abbildung 3.1 zeigt den Ablauf der Berechnung einer einzelnen GCF-Map. Zuerst werden Blöcke aus den Mikrofonensignalen mit einer Fensterfunktion (siehe 4.2) multipliziert. Für jede mögliche Kombination der Mikrofone wird daraus die Kreuzkorrelationsfunktion mit Phasentransformation mithilfe der FFT wie in

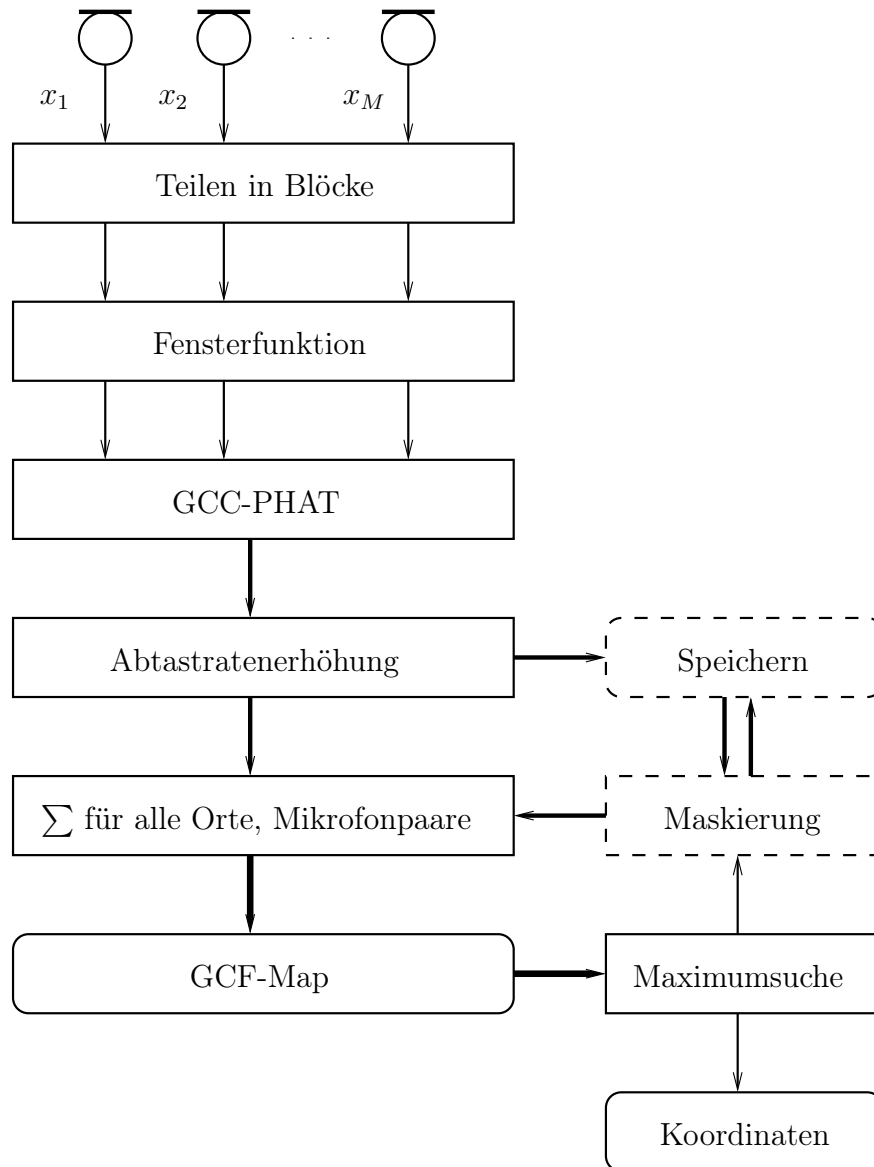


Abbildung 3.1: Programmablauf inkl. Maskierung von Schallquellen

(3.11) berechnet. Diese werden mit einem wählbaren Faktor in ihrer Abtastrate erhöht. Für jeden Ort in der GCF-Map wird nun die Summe aller zugehörigen Werte aus allen Kreuzkorrelationsfunktionen wie in Gleichung (3.25) berechnet. Die Speicherung der Kreuzkorrelationsfunktionen und deren Maskierung werden nur für die Behandlung simultaner Schallquellen wie in Kapitel 5.2 benötigt. Eine Suche des globalen Maximums in der GCF-Map führt zu den Koordinaten der

KAPITEL 3. GLOBAL COHERENCE FIELD METHODE

Schallquelle. Der Wert an der Stelle des Maximums entspricht im Idealfall mit einer einzelnen Schallquelle und rauschfreien Mikrofonsignalen der Anzahl der ausgewerteten Mikrofonpaare.

Kapitel 4

Simulationsparameter

In diesem Kapitel werden die Parameter zur Berechnung der GCF-Map vorgestellt und ihre Auswirkungen auf die berechneten GCF-Maps untersucht.

4.1 Blocklänge

Die Verarbeitung der Mikrofon-signale erfolgt blockweise. Eine größere Blocklänge N führt zu einer größeren Toleranz gegenüber Rauschen, und die Berechnung der GCF-Map erfolgt seltener. Dies führt zu einem geringeren Rechenaufwand, jedoch auch zu einer größeren Reaktionszeit bei Veränderungen der Schallquelle. Aus den Signalblöcken werden die Kreuzkorrelationsfunktionen berechnet. Deren nutzbare Breite wird umso größer je länger N ist. Größere Mikrofonabstände führen zu größeren Laufzeitunterschieden und erfordern eine größere Mindestlänge für N .

Alle Testdurchläufe verwenden identisches simuliertes weißes Rauschen. Die Abbildung 4.1 zeigt eine GCF-Map mit einem mittleren Signal-Rausch-Verhältnis (SNR) von 0dB an jedem Mikrofon und einer Blocklänge von 512 Abtastwerten zur Verarbeitung der Signale. Die Abtastrate f_s beträgt 16kHz. Die weißen Halbkreise oben im Bild stellen die 7 Mikrofone dar. In der Mitte des schwarzen Kästchens befindet sich die simulierte Schallquelle. Das weiße Kreuz zeigt die Stelle des gefundenen Maximums für den dargestellten Zeitabschnitt.

KAPITEL 4. SIMULATIONSPARAMETER

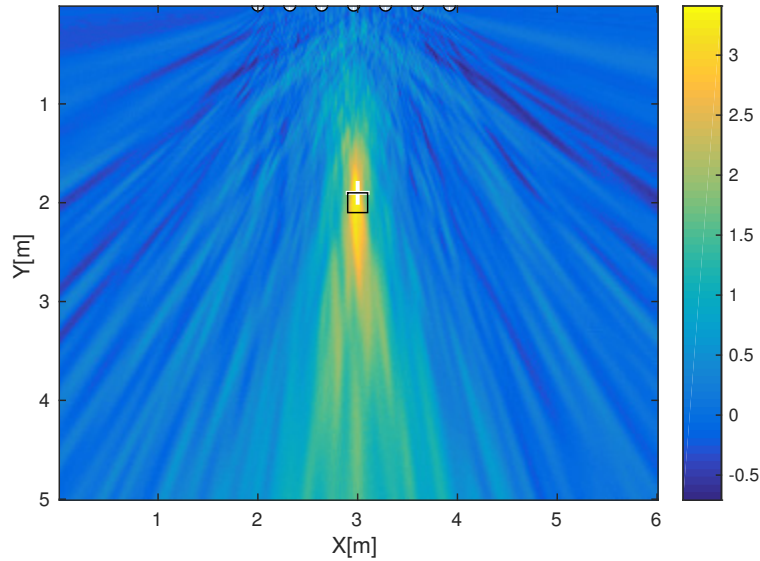


Abbildung 4.1: GCF-Map mit Blocklänge 512 bei 0dB SNR

Die Abbildung 4.2 zeigt eine GCF-Map für dasselbe Szenario, jedoch mit einer Blocklänge von 8192 Abtastwerten und einem SNR von ebenfalls 0dB. Man erkennt ein deutlich saubereres Bild. Das gefundene Maximum befindet sich am Ort der simulierten Schallquelle. An der Skala auf der rechten Seite kann man die höhere Amplitude des Maximums mit dem Wert von ca. 4,6 erkennen. Ebenfalls lassen sich jetzt die in Kapitel 2.2 beschriebenen Hyperbeln erahnen, die oben in der Mitte jedes Mikrofonpaares beginnen und sich alle beim Maximum kreuzen. Weiter unten laufen sie wieder auseinander.

Messfehler werden als quadratischer Mittelwert (RMS) aus Z Einzelmessungen mit

$$F_{RMS} = \sqrt{\frac{1}{Z} \sum_{n=1}^Z \overline{q_n q_0}^2} \quad (4.1)$$

berechnet. Dabei ist $\overline{q_n q_0}$ ist der Abstand zwischen ermittelten Maxima bei q_n und dem simulierten Quellsignal bei q_0 .

Die Abbildung 4.3 zeigt die Auswirkung von verschiedenen Signalblocklängen auf die Messgenauigkeit jeweils für 0dB, 10dB und 30dB SNR. Auf der Hochachse

KAPITEL 4. SIMULATIONSPARAMETER

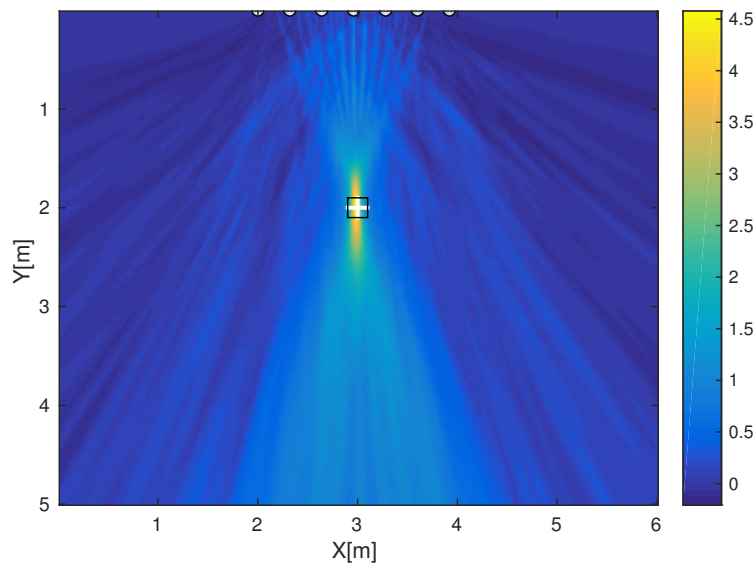


Abbildung 4.2: GCF-Map mit Blocklänge 8192 bei 0dB SNR

aufgetragen ist der nach (4.1) ermittelte Messfehler in der Einheit Meter. Da in allen Fällen ein Signal der Länge 10s ausgewertet wird, stehen bei größeren Blocklängen eine entsprechend geringere Anzahl von Messwerten zur Verfügung. Als Testsignal wurde ein Sprachsignal verwendet, bei dem die Sprechpausen zuvor entfernt wurden. GCF-Maps, die nur aus Rauschsignalen berechnet wurden, sind deshalb nicht Teil der Auswertung.

Klar zu erkennen ist der Zusammenhang zwischen SNR und Messgenauigkeit. Erwartungsgemäß sind die Ergebnisse bei höherem SNR deutlich genauer. Ebenfalls gut zu erkennen ist das durch eine größere Blocklänge ein kleines SNR kompensiert werden kann und sich so auch bei einem SNR von 0dB gute Ergebnisse im Zentimeterbereich erzielen lassen. Beispielsweise führt eine Blocklänge von 512 und 30dB SNR zu einer ähnlichen Genauigkeit wie eine Blocklänge von 4096 bei 10dB SNR. Die Wahl der Blocklänge 4096 bei einer Abtastrate von 16kHz ermöglicht hier relativ gute Messungen im Takt von ungefähr einer viertel Sekunde für 10dB SNR.

KAPITEL 4. SIMULATIONSPARAMETER

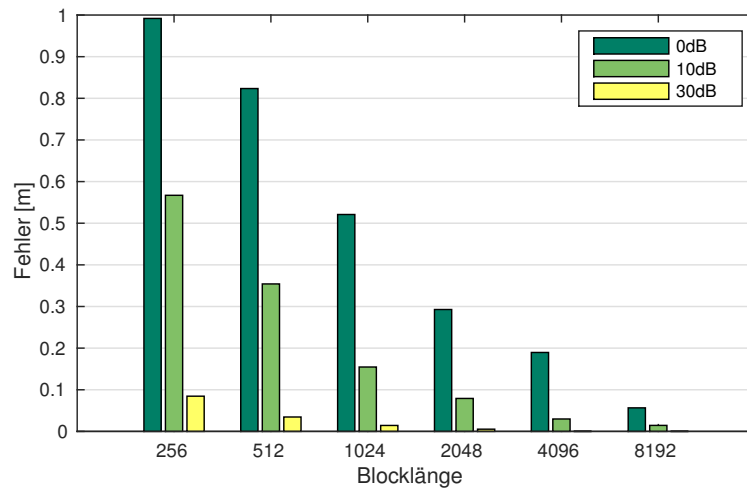


Abbildung 4.3: RMS Fehler bei verschiedenen Blocklängen und SNR

4.2 Fensterfunktion

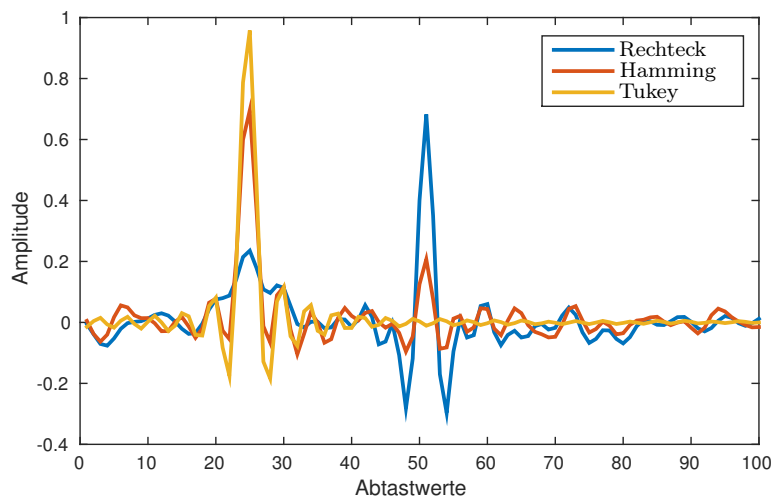


Abbildung 4.4: Auswirkung verschiedener Fensterfunktionen

Die FFT eines Signalblocks berechnet die Zeitdiskrete Fouriertransformation (DTFT) des zu beiden Seiten bis ins Unendliche periodisch wiederholten Signalblocks. Je nach Signalform können dadurch an den Rändern des Signalblocks Sprünge entstehen. Da diese Sprünge für alle Mikrofonsignale an derselben Stelle

KAPITEL 4. SIMULATIONSPARAMETER

auftreten, verursachen sie eine Signalspitze an der Stelle 0 der Kreuzkorrelationsfunktionen und verursachen fehlerhafte GCF-Maps. Abbildung 4.4 zeigt einen Ausschnitt dreier Kreuzkorrelationsfunktionen. Diese wurden aus denselben Signalblöcken, die jedoch mit verschiedenen Fensterfunktion multipliziert wurden, errechnet. Die blaue Kurve zeigt die Variante mit Rechteckfenster. Diese zeigt einen großen Ausschlag in der Mitte bei Abtastwert 50. Diese Stelle steht für eine Zeitdifferenz von 0 zwischen den Signalen beider Mikrofone. Die Schallquelle in diesem Beispiel sollte jedoch eine Spitze bei ca. 24 verursachen. Die rote Kurve zeigt einen Hauptausschlag bei Abtastwert 24. Ein kleiner Ausschlag in der Mitte ist jedoch auch hier erkennbar, denn das verwendete Hamming-Fenster fällt an seinen Rändern ebenfalls nicht bis 0 ab. Die orange Kurve zeigt einen noch größeren Ausschlag bei 24 als mit dem Hamming-Fenster. In der Mitte ist keine Signalspitze mehr erkennbar. Verwendet wird ein Tukey-Fenster welches an den Rändern bei 0 beginnt, zur Mitte hin kosinusförmig ansteigt und über einen großen Teil um die Mitte 1 bleibt. Die Parameter des Fensters werden dabei so gewählt, das bei sich überlappenden Signalblöcken jeder Abtastwert gleich gewichtet wird. Die Überlappung bei allen Simulationen dieser Arbeit beträgt 20%.

Die Effekte einer schlecht gewählten Fensterfunktion werden allerdings durch ein besonders gleichmäßiges weißes Quellsignal oder ein kleines Signal-Rausch-Verhältnis abgemildert.

4.3 Abtastrate

Die Abtastrate f_s der Mikrofonssignale sollte möglichst gering gewählt werden, um die Datenmengen und den Rechenaufwand möglichst klein zu halten. Da aber auch die Signalenergie der Quellsignale möglichst gut ausgenutzt werden soll darf die Abtastrate nicht zu klein sein. Für Sprachsignale wird hier meistens $f_s=16\text{kHz}$ gewählt.

4.4 Abtastratenerhöhung

Wie in Kapitel 3.2.2 beschrieben sind die Kreuzkorrelationsfunktionen der Mikrofonpaare zeitdiskrete Signale. Bei der Zuordnung eines Gitterpunktes der GCF-Map zu einer Stelle in der Kreuzkorrelationsfunktion ist daher eine Rundung zum nächstgelegenen Abtastwert wie in Gleichung (3.25) nötig. Jeder Abtastwert repräsentiert bei der Berechnung der GCF-Map dadurch einen Zeitbereich mit der Länge des Abtastintervalls. Jede Stelle einer Kreuzkorrelationsfunktion wird in der GCF-Map auf eine Hyperbel wie in 2.2 abgebildet. Da die Funktionen zeitdiskret sind wird ein einzelner Abtastwert auf eine Schar von Hyperbeln abgebildet. In großem Abstand zum Mikrofonpaar ist die Breite der Schar ungefähr proportional zum Abstand des Mikrofonpaares. Der Beitrag eines jeden Abtastwertes zur GCF-Map besteht also näherungsweise aus einem Dreieck mit gleichem Wert.

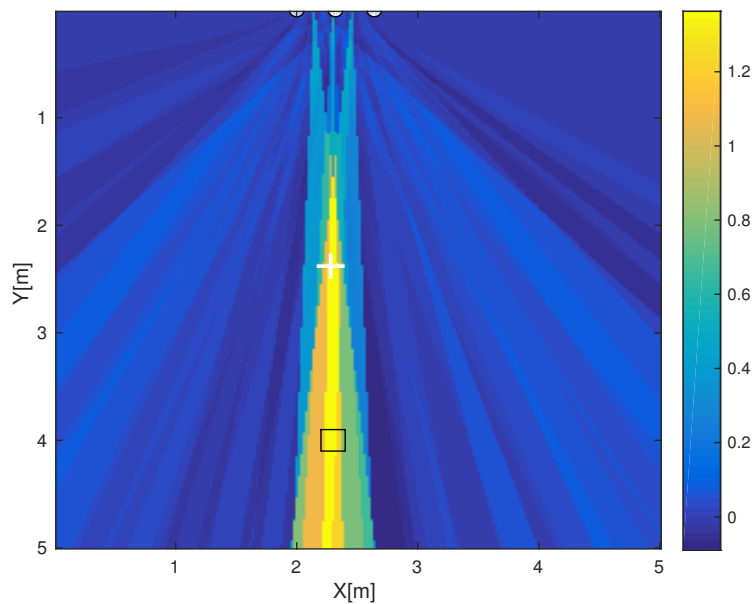


Abbildung 4.5: GCF-Map bei $f_s=16\text{kHz}$

Abbildung 4.5 zeigt ein Beispiel mit einem 64cm kurzen Array mit 3 Mikrofonen und einer mit 4m relativ weit entfernten Schallquelle. Ausgehend von den

KAPITEL 4. SIMULATIONSPARAMETER

Mittelpunkten zwischen allen Mikrofonpaaren an der Oberkante der GCF-Map kann man die Dreiecksflächen des jeweils maximalen Abtastwertes der dazugehörigen Kreuzkorrelationsfunktionen erkennen. Überlagern sich die drei Dreiecke bilden sie aufsummiert das Maximum des gesamten GCF-Map. Das Maximum ist hier die längliche gelbe Fläche mit gleichem Wert. Eine Maximumsuche führt nicht zu eindeutigen Koordinaten oder wie hier zu Koordinaten, die vom verwendeten Suchalgorithmus abhängig sind. Die längliche Form der Maximumsfläche macht es insbesondere schwierig, den Abstand der Schallquelle zum Mikrofonarray abzuschätzen.

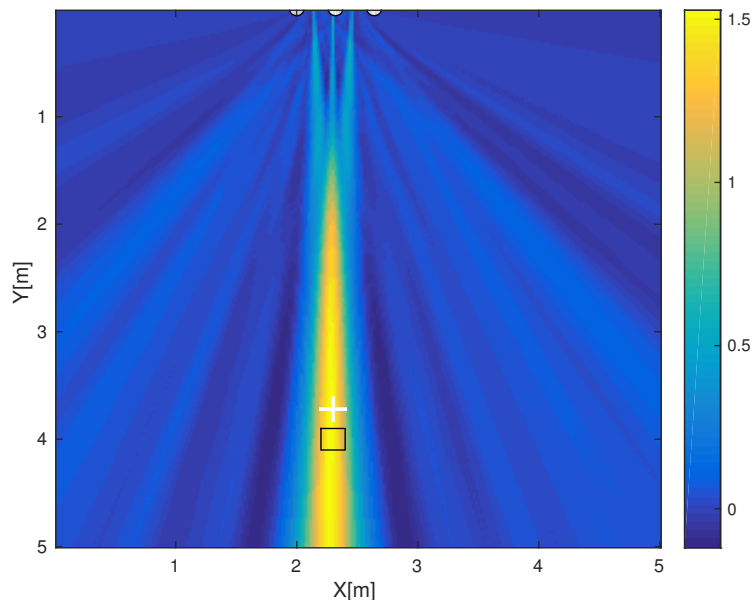


Abbildung 4.6: GCF-Map bei $f_s=16\text{kHz}$ und 8facher Abtastratenerhöhung

Eine Möglichkeit diese Maximumsfläche zu verkleinern um die Schallquelle genauer zu orten, ist die Umrechnung der Kreuzkorrelationsfunktionen auf eine höhere Abtastrate. Die Hyperbelschar jedes Abtastwertes wird dadurch um den Faktor der Abtastratenerhöhung U schmaler. Abbildung 4.6 zeigt die GCF-Map aus Abbildung 4.5 zum selben Zeitpunkt berechnet mit einer um den Faktor $U=8$ erhöhten Abtastrate der Kreuzkorrelationsfunktionen. Die neue GCF-Map

hat keine harten Kanten und zeigt deutlich mehr Farbstufen. Das weiße Kreuz markiert das Ergebnis der Maximumsuche. Es liegt jetzt näher am wahren Ort der Schallquelle, welcher mit einem schwarzen Kästchen markiert ist.

4.5 Gitterabstand

Die GCF-Map wird für kartesische Koordinaten mit gleichmäßig verteilten Gitterpunkten berechnet. Dem Gitterpunkt mit dem maximalen Wert wird der Quellort der Schallquelle zugeordnet. Dieser Gitterpunkt ist eine Näherungslösung für das Maximum einer kontinuierlichen GCF-Map. Ist die Form des „Maximumhügels“ schmal aber sehr lang und hat einen kleinen Winkel zu den Koordinatenachsen kann der Fehler bei der Positionsbestimmung größer sein als der Gitterabstand. Ein kleinerer Gitterabstand führt zu einer genaueren Bestimmung des Quellorts, allerdings auch zu einer Vergrößerung des Rechenaufwandes.

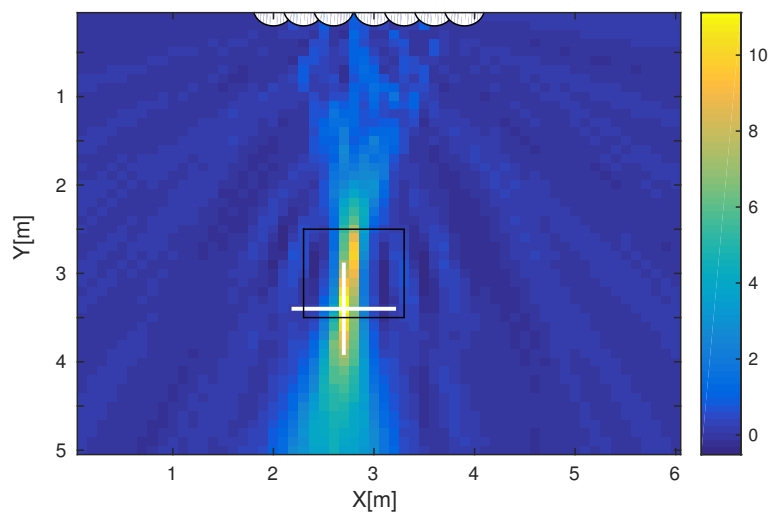


Abbildung 4.7: Gitterabstand 10cm: Fehler 35cm

Abbildung 4.7 zeigt ein Beispiel mit einem Gitterabstand von 10cm. Die wahre Position der Schallquelle befindet sich in der Mitte des schwarzen Kastens. Oberhalb der Mitte hat die GCF-Map ein lokales Maximum. Unterhalb befindet sich

KAPITEL 4. SIMULATIONSPARAMETER

das globale Maximum, das mit einem weißen Kreuz markiert ist. Der Abstand zur simulierten Position beträgt 35cm.

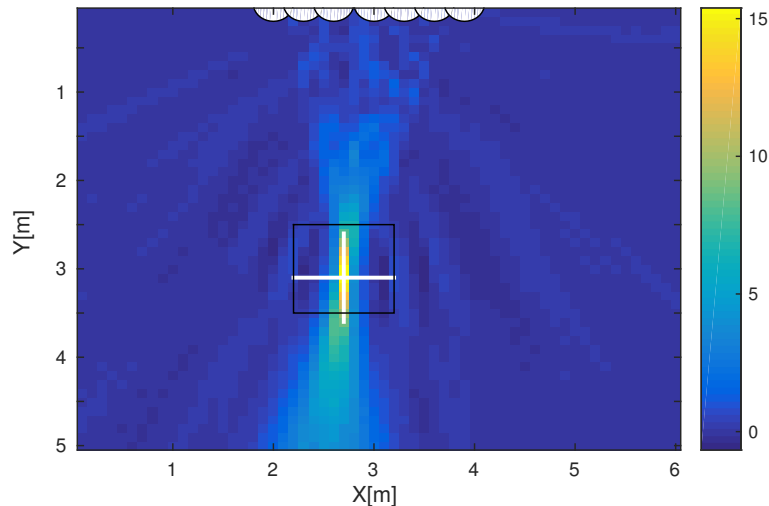


Abbildung 4.8: Fehler nach Verschiebung der Quelle um 5cm nur noch 5cm

Abbildung 4.8 zeigt das vorherige Beispiel mit einer um 5cm verschobenen simulierten Schallquelle. In diesem Fall reduziert sich der Fehler der Positionsbestimmung auf 5cm.

4.6 Mittelung

Um die Robustheit gegenüber Rauschen zu erhöhen ist es möglich die Blocklänge N zu erhöhen. Alternativ dazu ist es auch möglich mehrere aufeinanderfolgende GCF-Maps zu mitteln.

Eine einfache Möglichkeit eine Mittelung zu implementieren besteht darin, den letzten gemittelten Wert mit einem Vergessensfaktor v_G zu multiplizieren und zu dem mit $(1 - v_G)$ multiplizierten neuen Wert zu addieren um einen neuen Mittelwert zu erhalten. Der Vergessensfaktor v_G wird aus einer zuvor gewählten Zeitkonstante T_G mit

$$v_G = e^{-\frac{\Delta t}{T_G}} \quad (4.2)$$

KAPITEL 4. SIMULATIONSPARAMETER

berechnet, wobei Δt der Zeitabstand zweier Messwerte ist.

Der neue Mittelwert \overline{W}_n berechnet man aus dem neuen Messwert W_n und dem alten Mittelwert \overline{W}_{n-1} mit

$$\overline{W}_n = (1 - v_G)W_n + v_G\overline{W}_{n-1}. \quad (4.3)$$

Die so erhaltene Mittelwerte \overline{W}_n entsprechen der mit einem Tiefpass 1. Ordnung gefilterten Messwertreihe W_n .

Angewendet wird diese Methode in dieser MATLAB Simulation auf jeden Abtastwert der Kreuzkorrelationsfunktionen der einzelnen Mikrofonpaare, aus denen anschließend die GCF-Map berechnet wird. Dies entspricht der Mittelung der gesamten GCF-Map.

Alternativ lässt sich dieses Verfahren auf die Koordinaten der gefundenen Maxima in den GCF-Maps der aufeinanderfolgenden Signalblöcke anwenden. Für die gemittelten Koordinaten \overline{x}_n und \overline{y}_n gilt dann

$$\overline{x}_n = (1 - v_P)x_n + v_P\overline{x}_{n-1} \quad (4.4)$$

$$\overline{y}_n = (1 - v_P)y_n + v_P\overline{y}_{n-1} \quad (4.5)$$

und für den gemittelten Wert des Maximums \overline{p}_n gilt

$$\overline{p}_n = (1 - v_P)W_n + v_P\overline{p}_{n-1}. \quad (4.6)$$

Der Vergessensfaktor v_P berechnet sich wie in Gleichung (4.2) aus einer gewählten Zeitkonstante T_P .

Abbildung 4.9 zeigt den Vergleich beider Varianten. Bei dem Experiment wurde ein Sprachsignal ohne Sprechpausen bei einer SNR von 0dB genutzt. Die Blocklänge betrug 2048 Abtastwerte bei $f_s=16\text{kHz}$. In Grün dargestellt ist der RMS Positionsfehler für die Mittelung der GCF-Map für verschiedene Zeitkonstanten T_G . Gelb dargestellt ist der Fehler für gemittelte Positionen der Maxima mit der Zeitkonstanten T_P . Für Zeitkonstanten ab 0,1s liefert die Mittelung der GCF-Map deutlich genauere Ergebnisse als die Mittelung der Koordinaten.

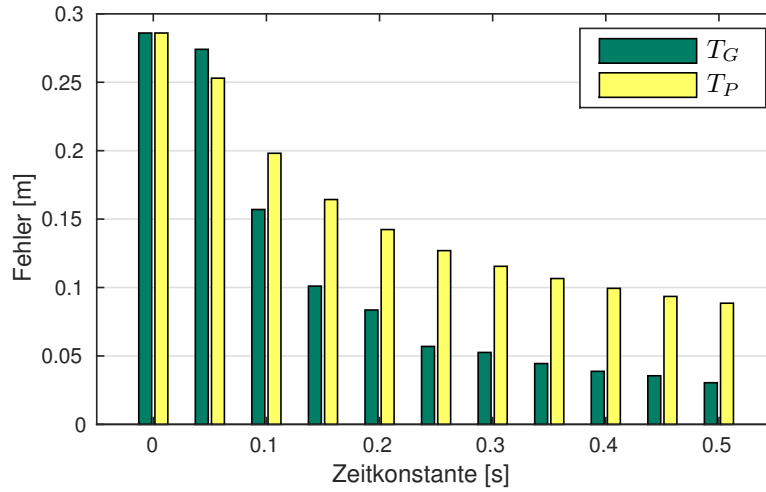


Abbildung 4.9: Messfehler bei verschiedenen T_G und T_P

Ein zusätzlicher Nachteil der Koordinatenmittelung besteht darin, dass sie bei schlechtem Signal-Rausch-Verhältnis nicht dabei hilft, Schallquellen überhaupt zu erkennen. Problematisch ist auch die Entscheidung ob die Koordinaten des Maximums einer GCF-Map und die Koordinaten der darauf folgenden GCF-Map zur selben Schallquelle gehören.

4.7 Rauschschwelle

Zur Auswertung der Positionsmessungen mit der GCF Methode ist ein Schwellwert nötig um zu unterscheiden ob während der Dauer eines Signalblocks eine Schallquelle aktiv ist, oder nur Rauschsignale verarbeitet wurden. Ist keine Schallquelle aktiv transformiert sich das Rauschen der Mikrofonssignale in ein Rauschen der gesamten GCF-Map. Als Schwellwert werden in den folgenden Simulationen empirisch ermittelte Werte genutzt, die vom Maximum einer aus Rauschen berechneten GCF-Map mit hoher Wahrscheinlichkeit nicht erreicht werden. Da die zur Berechnung verwendete GCC-PHAT Amplitudeninformationen ignoriert, hängt der gewählte Schwellwert ausschließlich von der Blocklänge N und der Anzahl der genutzten Mikrofonpaare M_P ab. Ein relativ sicherer Schwellwert S ohne

KAPITEL 4. SIMULATIONSPARAMETER

Gefahr einer falsch positiven Erkennung einer Schallquelle berechnet sich mit

$$S = 12\sqrt{\frac{M_P}{N}}. \quad (4.7)$$

Wählt man für die Mittelung der GCF-Map eine Zeitkonstante mit $T_G > 0$ s, sind auch kleinere Werte geeignet.

Ist der Wert eines Maximums der GCF-Map höher als der Schwellwert werden seine Koordinaten als Messung einer vorhandenen Schallquelle interpretiert und bei der Berechnung des Messfehlers einbezogen. Ist z.B. während einer Sprechpause der Wert des Maximums kleiner als der Schwellwert werden seine Koordinaten ignoriert.

Kapitel 5

Experimente

In diesem Kapitel werden mehrere Szenarien mit einzelnen oder mehreren simultanen Schallquellen untersucht. Dabei werden die Mikrofonsignale vollständig simuliert oder basieren auf gemessenen Raumimpulsantworten.

5.1 Szenarien mit einer Schallquelle

Zunächst werden hier einfache Szenarien mit einer einzelnen Schallquelle untersucht. Dies dient zur Darstellung der Auswirkungen verschiedener Mikrofonanordnungen und von Nachhall auf das Positionsergebnis.

5.1.1 Lineares Array zur Abstandsmessung

Lineare Arrays sind die einfachste Form die Mikrofone anzuordnen. Zwei Mikrofone sind die einfachste Variante davon. Da jedoch nur eine Hyperbel im TDOA verfahren generiert wird lässt sich nur die Richtung, nicht jedoch der Abstand zum Mikrofonpaar bestimmen. Ab zwei Mikrofonpaaren mit mindestens 3 Mikrofonen lässt sich der Abstand durch den Schnittpunkt mehrerer Hyperbeln bestimmen. Die geschieht durch die Maximumsuche in der GCF-Map automatisch. Es ist naheliegend das ein breiteres Array besser zur Abstandsmessung geeignet

KAPITEL 5. EXPERIMENTE

Nr.	Abstände[m]							Breite[m]
1	-0,9	-0,6	-0,3	0	0,3	0,6	0,9	1,8
2	-0,9	-0,45	-0,15	0	0,15	0,45	0,9	1,8
3	-0,9	-0,75	-0,15	0	0,15	0,75	0,9	1,8
4		-0,6	-0,3	0	0,3	0,6		1,2
5	-0,6	-0,4	-0,2	0	0,2	0,4	0,6	1,2
6	-0,3	-0,2	-0,1	0	0,1	0,2	0,3	0,6

Tabelle 5.1: Verschiedene Mikrofonanordnungen

ist als ein schmales, da sich die Hyperbeln damit in einem größeren Winkel zueinander kreuzen. Eine große Anzahl an Mikrofonen erhöht den Aufwand bei der Berechnung der GCF-Map und bei der Hardware des Arrays. Das Signal-Rausch-Verhältnis der GCF-Map wird jedoch durch mehr Mikrofone verbessert.

Tabelle 5.1 zeigt eine Auswahl mehrerer Mikrofonanordnungen für ein lineares Array. Nr. 1 bis 3 haben die gleiche Gesamtbreite von 1,80m jedoch eine unterschiedlichen Verteilung der 7 Mikrofone. Bei Nr. 1 haben alle Mikrofone den gleichen Abstand. Bei Nr. 2 handelt es sich um ein logarithmisches Array mit einer höheren Mikrofondichte in der Mitte. Die Verteilung bei Nr. 3 ist ein Versuch den Abstand der Mikrofone aller Paare zu maximieren und gleichzeitig die Mittelpunkte der Paare möglichst weit von der Mitte des Arrays entfernt zu halten, um die Abstandsmessung zu verbessern. Nr. 4 und 5 haben eine reduzierte Breite von 1,20m. Bei Nr. 4 ist auch die Mikrofonanzahl auf 5 reduziert. Anordnung Nr. 7 hat lediglich ein Breite von 60cm.

Die Mitte des Mikrofonarrays in der Simulation befindet sich bei $x=2,9$ und $y=0$ im simulierten Raum. Tabelle 5.2 zeigt den Abstandsfehler für drei verschiedene simulierte Schallquellen. Die Fehler sind quadratische Mittelwerte wie in Gleichung (4.1). Die Anordnung 1* besteht aus denselben Mikrofonen wie bei Nr. 1 jedoch werden die linken 4 und die rechten 4 Mikrofone als unabhängige

Position (x,y)[m]	Anordnung						
	1	1*	2	3	4	5	6
(1, 4)	0,47	0,32	0,24	0,29	0,48	0,36	0,98
(1, 1)	0,07	0,12	0,23	0,14	0,26	0,26	0,61
(3, 4)	0,13	0,16	0,20	0,25	0,33	0,26	0,84

Tabelle 5.2: Abstandsfehler[m] bei verschiedene Messpositionen

Arrays betrachte und ihre jeweiligen GCF-Maps addiert. Die Anzahl der zu verarbeitenden Mikrofonpaare sinkt dadurch von 21 auf 12. Die Messgenauigkeit wird dadurch geringfügig verschlechtert, im Falle der Position (1, 4) sogar verbessert. Der Wert von 0,47m bei Anordnung 1 ist auf einen einzelnen Ausreißer in der Messreihe zurückzuführen, der durch das simulierte Rauschen verursacht wird. Die mittlere SNR an den Mikrofonen beträgt 11,6dB. Maßnahmen wie die Verlängerung der Blocklänge auf mehr als die hier verwendeten 256ms oder eine Mittelung aufeinanderfolgender GCF-Maps würden alle Werte dieser Tabelle verbessern.

Die logarithmische Anordnung 2 und die experimentelle Anordnung 3 zeigen eine Verschlechterung gegenüber Anordnung 1 mit der gleichen Breite und gleichmäßig verteilten Mikrofonen. Die Abstandsmessungen mit den verkürzten Anordnungen 4 und 5 liefern erwartungsgemäß weniger genaue Ergebnisse. Wegen des Rauschsignals ist Anordnung 5 durch die größere Mikrofonanzahl überlegen. Das kürzeste Array der Anordnung 6 zeigt eine deutliche Verschlechterung bei der Genauigkeit der Abstandsmessung.

5.1.2 Simulierte Reflexion

Es ist möglich ein an einer Fläche reflektiertes Schallsignal nachzubilden, indem man eine weitere Schallquelle der Simulation hinzufügt. Das Signal der hinzugefügten Schallquelle ist identisch mit der Originalschallquelle. Die Koordinaten

KAPITEL 5. EXPERIMENTE

der Schallquelle werden an einer gedachten Reflexionsebene gespiegelt. Spiegelt man die Originalschallquelle und die Spiegelschallquelle erneut an einer weiteren Reflexionsebene erhält man die Simulation einer Zweifachreflexion der Originalschallquelle. Dieser Vorgang ließe sich beliebig fortführen um Mehrfachreflexionen nachzubilden.

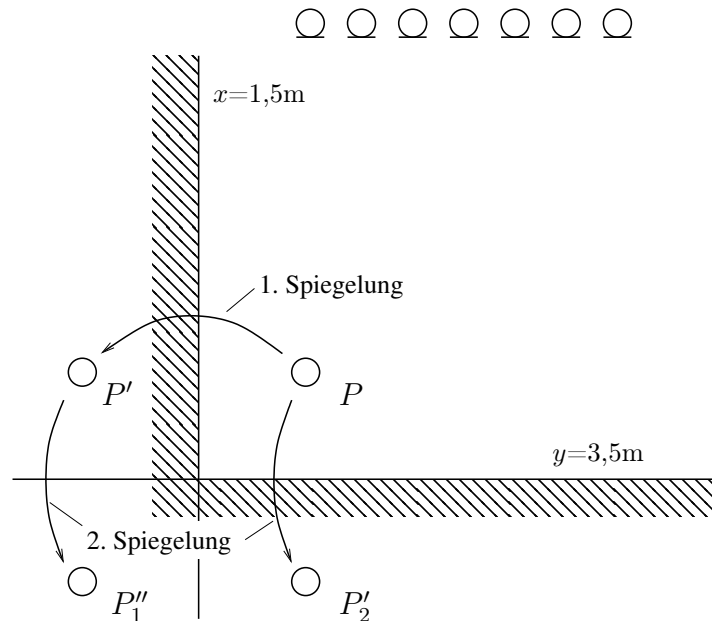


Abbildung 5.1: Erzeugung der Spiegelschallquellen

Abbildung 5.1 zeigt die Vorgehensweise in diesem Beispiel. Zunächst wird die Schallquelle P bei $(2, 3)$ an der Reflexionsebene bei $x=1,5\text{m}$ gespiegelt und man erhält die Spiegelschallquelle P' bei $(1, 3)$. Bei der 2. Spiegelung werden P und P' an der Reflexionsebene $y=3,5\text{m}$ gespiegelt. Man erhält dadurch P'_1 bei $(1, 4)$ und P'_2 bei $(2, 4)$.

Abbildung 5.2 zeigt die GCF-Map der Schallquelle P und ihrer Spiegelschallquellen. Das weiße Kreuz markiert das Maximum der GCF-Map. Diese fällt mit der gedachten „wahren“ Schallquelle zusammen, da die sich am nächsten am Mikrofonarray oben in der Mitte befindet. Trotz des identischen Signals aller vier simulierter Schallquellen wirkt die resultierende GCF-Map wie eine Überlagerung der GCF-Maps von vier unabhängigen Schallquellen.

KAPITEL 5. EXPERIMENTE

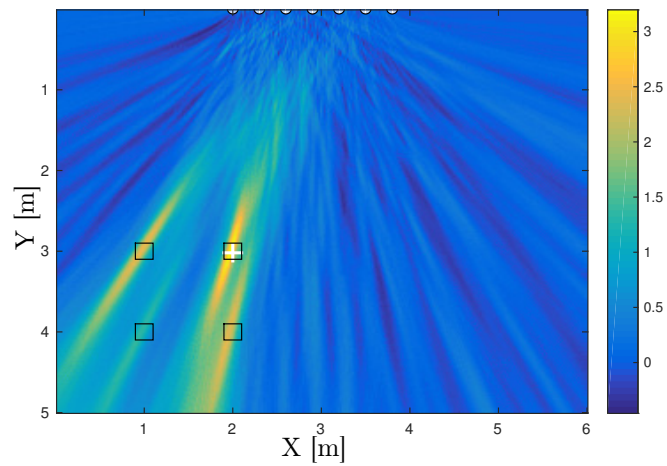


Abbildung 5.2: GCF-Map mit Spiegelschallquellen

Abbildung 5.3 zeigt denselben Zeitpunkt ohne simulierte Reflexion. Der Wert des Maximums der GCF-Maps ist ungefähr doppelt so groß wie der mit Spiegelschallquellen. Die Auswirkungen auf die Genauigkeit der Positionsbestimmung ist jedoch gering, wie Tabelle 5.3 zeigt.

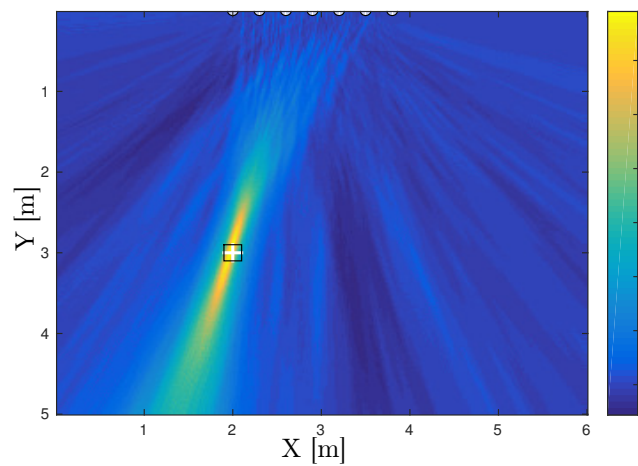


Abbildung 5.3: GCF-Map ohne Spiegelschallquellen

RMS Fehler	Winkel[°]	Abstand[m]
Ohne Spiegelschallquellen	0,15	0,10
Mit Spiegelschallquellen	0,30	0,13

Tabelle 5.3: Messwerte mit und ohne Spiegelschallquellen

5.1.3 3D Fall und mehrteilige Arrays

Bisher befanden sich in den Simulationen die Schallquellen, die Mikrofone und die GCF-Map in derselben Ebene. Befindet sich eine Schallquelle ober- oder unterhalb dieser Ebene und ist also in der z -Achse verschoben, kann man mehrere Fälle unterscheiden.

Benutzt man ein lineares Array zur Positionsbestimmung werden praktisch alle möglichen GCF-Maps, die sich aus einer Rotation ihrer Ebene um die Längsachse des Arrays bilden lassen, auf einmal berechnet. Befindet sich das Array bei $y=0$ und eine Schallquelle bei Punkt P , vergrößert sich die gemessene y -Koordinate y'_P gegenüber der tatsächlichen y -Koordinate y_P mit

$$y'_P = \sqrt{y_P^2 + z_P^2} \quad (5.1)$$

bei einem vertikalen Abstand z_P zur Ebene der GCF-Map. Abbildung 5.4 zeigt

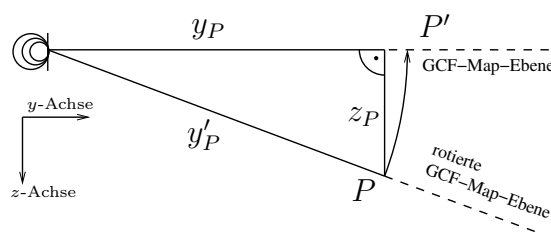


Abbildung 5.4: Abbildung einer vertikal verschobenen Schallquelle auf die GCF-Map

den Punkt P innerhalb einer möglichen rotierten Map-Ebene. Da er sich um z_P außerhalb der berechneten GCF-Map befindet, wird er als P' mit der y -Koordinate y'_P in die GCF-Map abgebildet.

KAPITEL 5. EXPERIMENTE

Hat man ein zweidimensionales Mikrofonarray sind die Auswirkungen weniger leicht vorhersehbar. Die Möglichkeit, dass sich alle Hyperbeln aller Mikrofonpaare in exakt demselben Punkt treffen, ist dann nicht mehr gegeben, denn bei dreidimensionaler Betrachtung werden aus den Hyperbeln mit gleicher Abstandsdifferenz Hyperboloiden, deren Schnittpunkt außerhalb der GCF-Map-Ebene liegt. Das Maximum in der GCF-Map wird weniger scharf und verschiebt sich gegenüber dem Abbild der Schallquelle auf die GCF-Map. Die Richtung und Weite der Verschiebung sind abhängig von der Lage der Schallquelle und der Mikrofone.

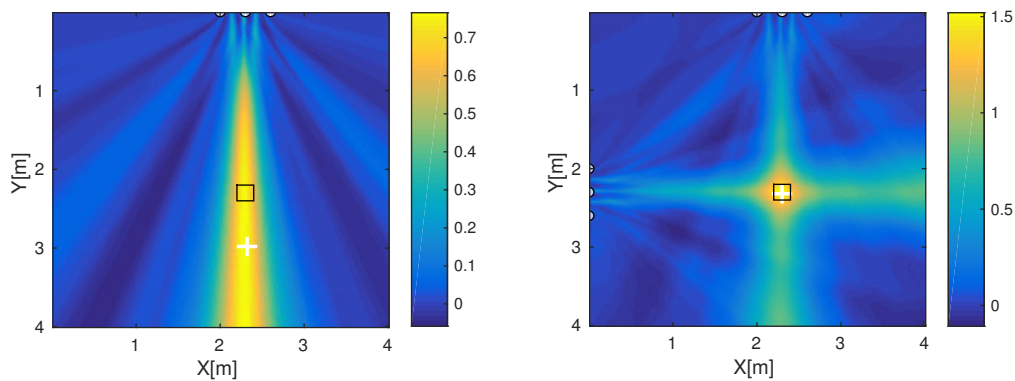


Abbildung 5.5: GCF-Maps bei vertikal versetzter Schallquelle um 2m(!) mit 1 und 2 Arrays

Sind lineare Arrays kurz im Vergleich zum Abstand der Schallquellen sind sie schlecht dazu geeignet den Abstand zu messen. Besser geeignet sind sie zur Winkelmessung. Die resultierende Spitze in der GCF-Map ist daher scharf bezüglich des Winkels und unscharf bezüglich des Abstandes. Man kann eine GCF-Map auch aus der Summe der GCF-Maps zweier einzelner GCF-Maps berechnen. Befinden sich zwei kurze lineare Array rechtwinklig zueinander innerhalb der Map-Ebene, z.B. an den Wänden eines rechtwinkligen Raumes, besteht der Hauptbeitrag der einzelnen Arrays in der Winkelmessung. Der Abstand der gemessenen Position der Schallquelle von ihrem Abbild auf die GCF-Map wird dadurch geringer als bei der Nutzung eines einzelnen Arrays. Liegt die Quelle mittig vor beiden Arrays wird der Positionsfehler praktisch 0. Abbildung 5.5 zeigt die GCF-Maps

KAPITEL 5. EXPERIMENTE

eines Extremfalls, bei dem die Schallquelle um 2m vertikal zur Map-Ebene versetzt ist, dabei aber nur einen Abstand von 2,30m zu beiden 60cm breiten Mikrofonarrays hat. Im linken Bild wird nur das Array bei $y=0$ genutzt und die gemessene y -Koordinate vergrößert sich wie in Gleichung (5.1). Im rechten Bild liegt die Schallquelle mittig vor beiden Arrays und es wird die korrekte Position in der Map-Ebene bestimmt, da die Einzelarrays hauptsächlich der Winkelmessung dienen.

5.1.4 Kompakte Arrays und Nachhall

Bei diesem Versuch wurden die Mikrofonsignale aus einem Sprachsignal und gemessenen Raumimpulsantworten berechnet. Gemessen wurde mit einem linearen Array aus 14 Mikrofonen in einem Abstand von jeweils 2,54cm und einer Gesamtbreite von 33cm und mit einem zirkularem Array aus 16 Mikrofonen und einem Durchmesser von 30cm. Die Abtastrate der Mikrofonsignale beträgt 32kHz. Die Abtastratenerhöhung der Kreuzkorrelationsfunktionen beträgt $U=4$. Die Blocklänge beträgt 8192 Abtastwerte oder 256ms. Simuliertes Rauschen wurde nicht hinzugefügt.

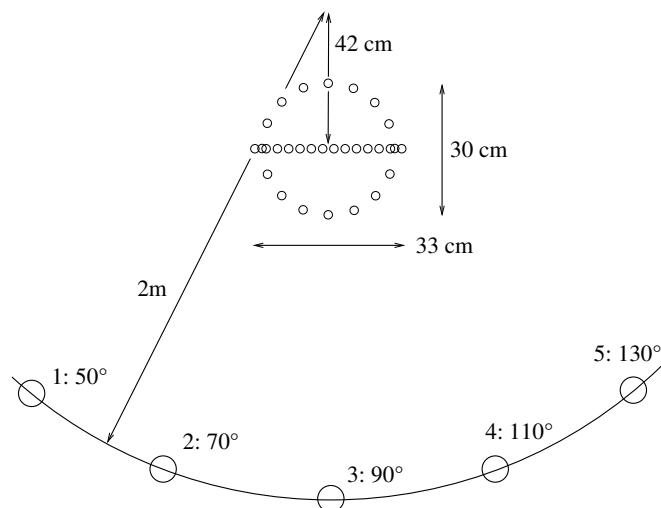


Abbildung 5.6: Anordnung der 5 Messpositionen

KAPITEL 5. EXPERIMENTE

Abbildung 5.6 zeigt die fünf Messpositionen relativ zu den Mikrofonarrays. Alle fünf Messpositionen befinden sich im Abstand von 20° auf einem Kreisbogen mit Radius 2m. Der Kreismittelpunkt liegt 42cm hinter den Mikrofonarrays. Position 3 befindet sich mittig vor den Mikrofonarrays.

		T_{60} [s] lineares Array					T_{60} [s] zirkulares Array				
		0,00	0,30	0,60	0,78	1,30	0,00	0,30	0,60	0,78	1,30
Position	1	4,9	5,1	1,2	4,7	4,0	5,8	6,9	0,0	3,8	5,4
	2	5,5	3,2	1,0	0,1	1,9	5,8	3,9	0,6	0,7	2,2
	3	3,5	0,0	2,3	0,8	0,0	3,6	0,0	2,2	0,7	0,0
	4	4,0	3,1	0,6	2,8	3,1	4,1	3,0	0,5	2,4	3,0
	5	7,2	4,1	2,2	1,2	5,0	7,0	5,3	1,5	0,9	5,6

Tabelle 5.4: Winkelfehler[°] bei verschiedenen T_{60} mit beiden Arrays

Vermessen wurden diese Anordnung in fünf Räumen mit Nachhallzeiten (T_{60}) von 0,0s bis 1,3s. Die Tabellen 5.4 und 5.5 zeigen den Winkelfehler und Abstandsfehler an den verschiedenen Positionen in den fünf Räumen für beide Mikrofonarrays.

		T_{60} [s] lineares Array					T_{60} [s] zirkulares Array				
		0,00	0,30	0,60	0,78	1,30	0,00	0,30	0,60	0,78	1,30
Position	1	0,63	0,55	0,29	0,14	0,01	1,85	1,18	2,09	2,11	1,20
	2	0,10	0,11	0,20	0,24	0,14	1,14	2,70	0,52	1,04	1,34
	3	0,28	0,35	0,08	0,16	0,00	1,29	1,70	0,96	1,58	1,70
	4	0,24	0,02	0,08	0,05	0,02	1,56	0,11	0,10	0,45	0,11
	5	0,19	0,32	0,09	0,16	0,73	1,30	0,82	0,09	0,32	0,15

Tabelle 5.5: Abstandsfehler[m] bei verschiedenen T_{60} mit beiden Arrays

Beim Abstand und bei den Winkeln lässt sich kein offensichtlicher Zusammenhang zwischen den Abweichungen und der Nachhallzeit des entsprechenden

KAPITEL 5. EXPERIMENTE

Raumes feststellen. Die Messfehler der Winkel werden womöglich teilweise durch Ungenauigkeiten im Versuchsaufbau verursacht. Der Unterschied zwischen beiden Arrays ist bei den Winkeln kleiner als 2° und damit sehr gering.

Der Abstand von 2m ist im Vergleich zu den Breiten der Arrays von 33cm und 30cm relativ groß, deshalb ist die Abstandsmessung schwierig. Trotz ihrer ähnlichen Größe fällt der Unterschied zwischen beiden Arrays deutlich aus. Beim linearen Array liegt der Abstandsfehler bei fast allen Messungen unter 40cm. Beim zirkularen Array ist der Abstandsfehler jedoch bei über der Hälfte aller Messungen größer als 1m.

In der Praxis werden die Genauigkeiten des linearen Arrays bei der Abstandsmessung schwer zu erreichen sein. Durch Rauschen und dem geringen Anteil an hohen Frequenzen bei Sprachsignalen kann die hier simulierte Abtastrate von 32kHz kaum ausgenutzt werden.

5.1.5 Messung am DICIT Array

Ebenfalls getestet wurde die GCF Methode an gemessenen Raumimpulsantworten mit dem Mikrofonarray des DICIT Projekts bei verschiedenen Anzahlen genutzter Mikrofone. Die Mikrofonsignale wurden aus dem Beispielsignal und den Impulsantworten der einzelnen Messpositionen errechnet. Ein simuliertes Rauschsignal wurde dabei nicht hinzugefügt. Die Impulsantworten wurden in einem Raum mit ca. 270ms Nachhallzeit (T_{60}) ermittelt. Verwendet wurde hier eine Abtastrate von 16kHz, eine Abtastraterhöhung um den Faktor $U=2$ und ein Sprachsignal einer männlichen Stimme. Ausgewertet wurden 10 Sekunden. Die Blocklänge betrug 256ms.

Das DICIT Array besteht aus 15 Mikrofonen. In Tabelle 5.6 sind ihre Positionen in x -Richtung angegeben. Im simulierten Raum befindet es sich bei $y=0$. Um die Mitte herum bei Mikrofon 8 gibt es einen logarithmischen Teil von Mikrofon 3 bis 13. Insgesamt beträgt die Breite 256cm.

KAPITEL 5. EXPERIMENTE

	Mikrofonpositionen														
Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x[\text{cm}]$	-128	-96	-64	-32	-16	-8	-4	0	4	8	16	32	64	96	128

Tabelle 5.6: Mikrofonanordnung DICIT Array

Abbildung 5.7 zeigt die Lage der mit Kreisen markierten Messpositionen für die Impulsantworten relativ zu den Mikrofonen. Die Messpositionen befanden sich in einem Bereich von 1m links der Mittelachse bis 2m rechts davon in Abständen von 1m, 2m und 3,50m.

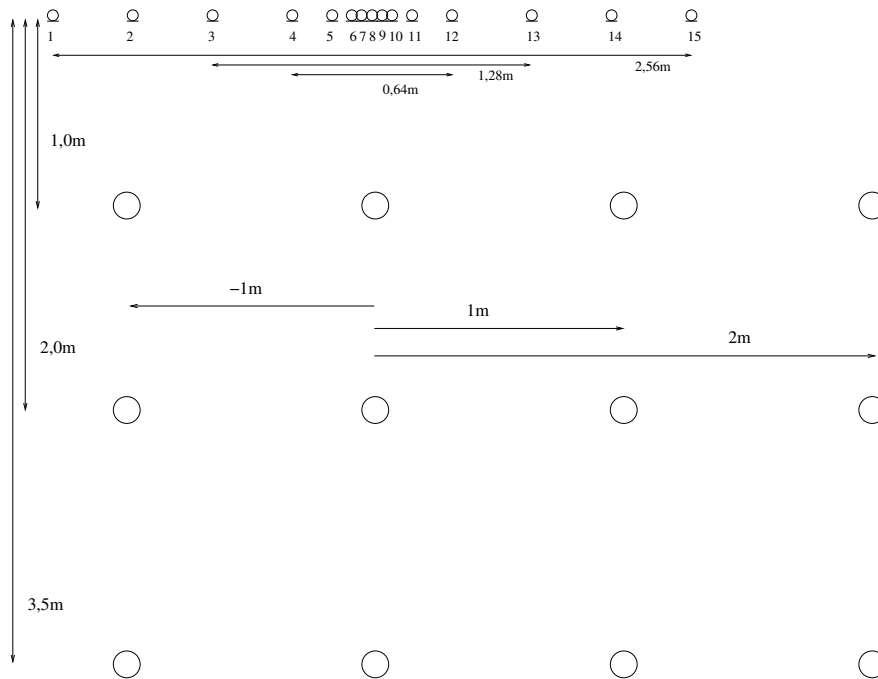


Abbildung 5.7: Anordnung der Messpositionen und des Mikrofonarrays

Das Diagramm in Abbildung 5.8 zeigt den RMS Positionsfehler in Metern für alle Messpositionen bei Berücksichtigung der Mikrofone 4 bis 12. Bei diesem mit 64cm relativ kleinen Array steigt der Positionsfehler mit größer werdenden Abstand vom Array deutlich auf 0,66m bei (0/3,5) an. Durch den großen Winkel zur Mittelachse wird das Mikrofonarray bei der Position (2/1,0) effektiv schmaler

KAPITEL 5. EXPERIMENTE

und verursacht durch die erschwerte Abstandsmessung einen großen Fehler von 0,65m.

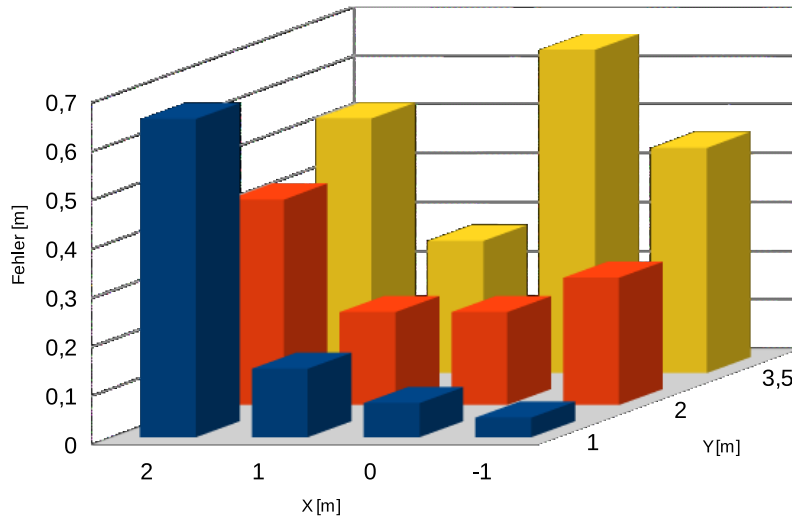


Abbildung 5.8: Fehler bei Nutzung der Mikrofone 4-12 bei einer Breite von 64cm

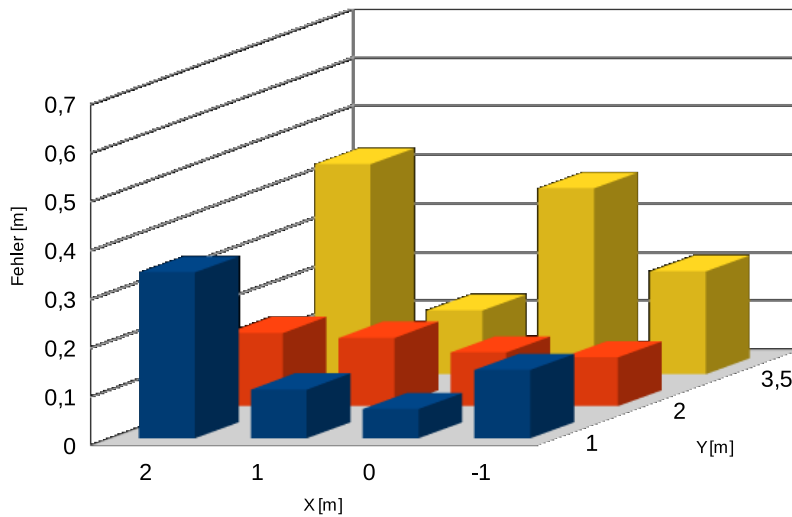


Abbildung 5.9: Fehler bei Nutzung der Mikrofone 3-13 bei einer Breite von 1,28m

Nutzt man zusätzlich die Mikrofone 3 und 13 des Arrays erhält man die Positionsfehler in Abbildung 5.9. Die größere Anzahl der Mikrofone und das mit 1,28m doppelt so breite Array führen erwartungsgemäß zu signifikant geringeren

KAPITEL 5. EXPERIMENTE

Fehlern von maximal 0,43m. Größere Abstände und große Winkel zur Mittelachse des Arrays führen dennoch zu einer erkennbaren Reduzierung der Positionsgenauigkeit.

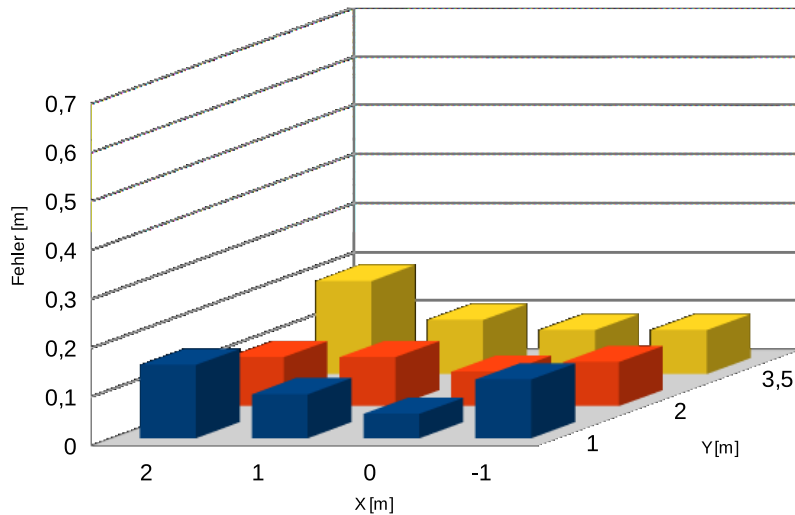


Abbildung 5.10: Fehler bei Nutzung aller Mikrofone bei einer Breite von 2,56m

Abbildung 5.10 zeigt die Positionsfehler bei der Verwendung aller Mikrofone. Die nochmals verdoppelte Breite des Arrays beträgt 2,56m. Der Fehler liegt nun bei allen Messpositionen unter 0,2m. Diese befinden sich durch die Größe des Arrays praktisch alle im Nahfeld. Als Grund für die verbleibenden Fehler bleiben lediglich Fehler im Messaufbau und Auswirkungen des Nachhalls.

Ignoriert man die Mikrofone 5-7 und 9-11 erhält man ein lineares Array mit einem gleichmäßigen Mikrofonabstand von 32cm bei einer gleichbleibenden Breite von 256cm. Die Positionsfehler ändern sich dadurch kaum, wie in Abbildung 5.11 zu sehen ist. Durch die Reduzierung der Mikrofonpaare halbiert sich die Rechenzeit von MATLAB, der Messfehler vergrößert sich im schlimmsten Fall jedoch nur um 3cm.

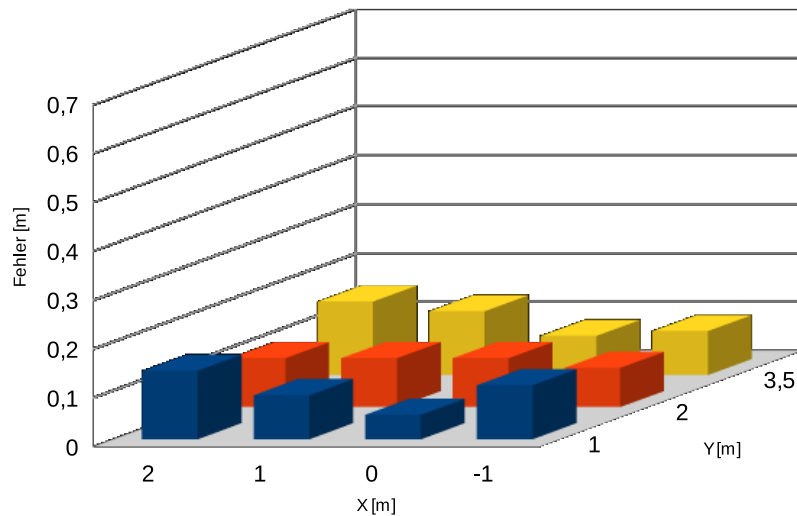


Abbildung 5.11: Fehler bei Nutzung des linearen Teilarrays bei einer Breite von 2,56m

5.2 Szenarien mit mehreren Schallquellen

Oft treten in der Praxis mehrere Schallquellen gleichzeitig auf. Mehrere Sprecher können gleichzeitig sprechen oder es wird ein Gespräch im Hintergrund geführt. Ebenfalls können Störgeräusche auftreten. Für eine weitere Verarbeitung der aufgenommenen Signale kann das Wissen über die Positionen der vorhandenen Schallquellen nützlich sein.

5.2.1 Dominante Schallquelle

Im Falle von mehreren Sprachquellen lassen sich die Eigenschaften der Sprachsignale ausnutzen, um sie zu orten. Die Ungleichmäßigkeit der Lautstärke eines Sprachsignals führt dazu, dass auch eine im Mittel leisere oder eine entfernte Sprachquelle kurzzeitig am Ort der Mikrofone lauter erscheint. Die gewählte Blocklänge sollte dazu eine achteil Sekunde oder weniger betragen, da sonst keine Möglichkeit besteht, beide Schallquellen getrennt zu orten.

Abbildung 5.12 zeigt zwei ähnlich laute Sprachsignale in rot und blau. Dabei

KAPITEL 5. EXPERIMENTE

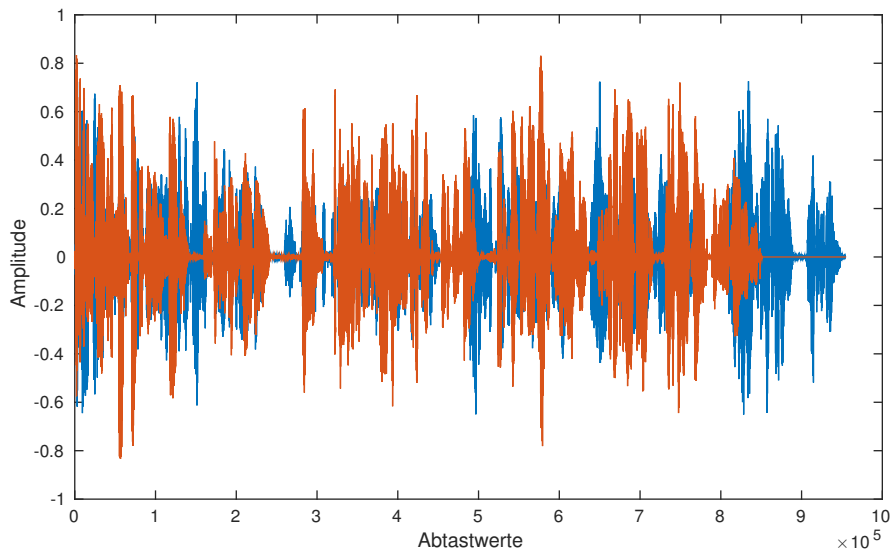


Abbildung 5.12: Amplitudenverlauf zweier Sprachquellen

wird das blaue, hinter dem roten dargestellte Signal dennoch zu manchen Zeitpunkten sichtbar, weil es dort lauter ist. Das in einem bestimmten Block im Mikrofonsignal lauter erscheinende Signal verursacht meistens das globale Maximum in der GCF-Map und lässt sich dadurch orten. Die GCF-Map darf bei diesem Verfahren nicht über mehrere Signalblöcke gemittelt werden, da es sonst passieren kann, dass Schallquellen gar kein oder über längere Zeitabschnitte kein globales Maximum verursachen, und sie dadurch nicht mehr mit einer einfachen Maximumsuche gefunden werden können.

RMS Fehler	dB	Winkel[°]	Abstand[m]	Anzahl
Sprecher 1	-2,3	0,53	0,11	48
Sprecher 2	0	0,16	0,04	44
Sprecher 1	-2,3	0,46	0,10	78
Sprecher 2	-10	0,23	0,04	11

Tabelle 5.7: Messwerte dominante Schallquellen

Tabelle 5.7 zeigt die Auswertung einer Simulation mit zwei Sprechern im

gleichen Abstand zu einem linearen Mikrofonarray; einmal haben beide mit einem Unterschied von 2,3dB ungefähr die gleiche Lautstärke und einmal wurde der zweite Sprecher um 10dB abgedämpft. Ausgewertet wurde ein Zeitraum von 10s mit einer Blocklänge von 128ms. Das simulierte SNR beträgt ca. 13,6dB im ersten, und ca. 10dB mit dem leiseren zweiten Sprecher. Die Ortungsgenauigkeit in dieser Simulation ist sehr genau. Ausgewertet wurden Maxima mit einem Wert größer als dem nach Gleichung (4.7) berechneten Schwellwerts. Die Anzahl der ausgewerteten Maxima ist im Falle der ungefähr gleich lauten Sprecher mit 48 und 44 sehr ausgewogen. Beim Beispiel mit dem bedämpften Sprecher 2 erhält man allerdings nur 11 Positionswerte in einem Zeitraum von 10s.

5.2.2 GCC-PHAT Maskierung

Eine weitere Methode mehrere Quellen zu Orten wurde in [BOS08] vorgeschlagen. Diese besteht darin, das gefundene Maximum aus der GCF-Map zu entfernen und im Resultat erneut das Maximum zu suchen.

Dazu wird zuerst der Ort des globalen Maximums \hat{q}_{GCF} bestimmt. Daraus wird für jedes Mikrofonpaar \mathbf{AB} die entsprechende Zeitdifferenzen $\Delta_{\mathbf{AB}}(\hat{q}_{GCF})$ berechnet. Alle Kreuzkorrelationsfunktionen $\mathbf{R}_{\mathbf{AB}}(\Delta k)$ werden nun mit der dazugehörigen Maskierungsfunktion $\phi(\Delta k, \Delta_{\mathbf{AB}}(\hat{q}_{GCF}))$ multipliziert und man erhält die maskierten Kreuzkorrelationsfunktionen $\mathbf{R}'_{\mathbf{AB}}(\Delta k)$ mit

$$\mathbf{R}'_{\mathbf{AB}}(\Delta k) = \phi\left(\Delta k, \Delta_{\mathbf{AB}}(\hat{q}_{GCF})\right) \cdot \mathbf{R}_{\mathbf{AB}}(\Delta k). \quad (5.2)$$

Die Maskierungsfunktion $\phi(\Delta k, \nu)$ für die Zeitdifferenz ν beim Abtastwert Δk und ist definiert mit

$$\phi(\Delta k, \nu) = \alpha \left[1 - e^{-\frac{|\Delta k - \nu|^p}{b}} \right]. \quad (5.3)$$

Die Gleichung (5.3) stammt aus [BOS09] und ist eine Verfeinerung der in [BOS08] angegebenen Formel. Aus [BOS09] entnommen wurden die hier verwendeten Parameterwerte $p=2$ und $b=3,5$ da sie für lineare und verteilte Mikrofonanordnungen gut geeignet sind.

KAPITEL 5. EXPERIMENTE

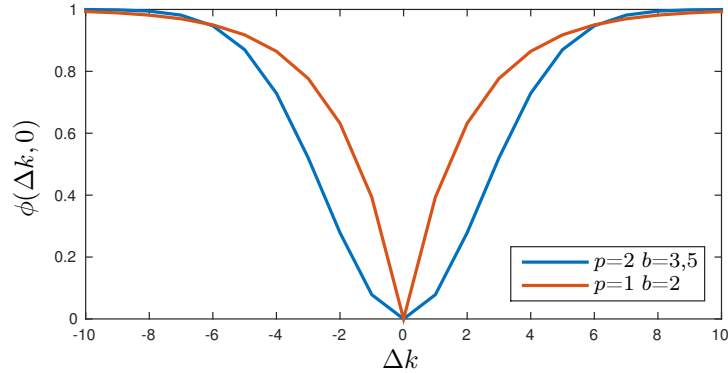


Abbildung 5.13: Maskierungsfunktion $\phi(\Delta k, 0)$ für verschiedene p und b

Abbildung 5.13 zeigt zwei Maskierungsfunktionen für verschiedene Kombinationen aus p und b für eine Zeitdifferenz $\nu=0$. Der Parameter p bestimmt dabei die Krümmung am Minimum. Der Parameter b bestimmt die Breite der Kerbe der Maskierungsfunktion. Der Normierungsfaktor α wird in dieser Implementation nicht verwendet und hat deshalb den Wert 1 bei allen Simulationen. Aufgabe der Maskierungsfunktion ist es, den Anteil der zuerst gefundenen Schallquelle in der Kreuzkorrelationsfunktion vollständig zu entfernen, jedoch den Rest möglichst unverändert zu lassen. Aus diesen so nachbearbeiteten Kreuzkorrelationsfunktionen wird eine neue GCF-Map berechnet. Eine Maximumsuche führt zum Ort der Schallquelle mit dem neuen globalen Maximum in der GCF-Map.

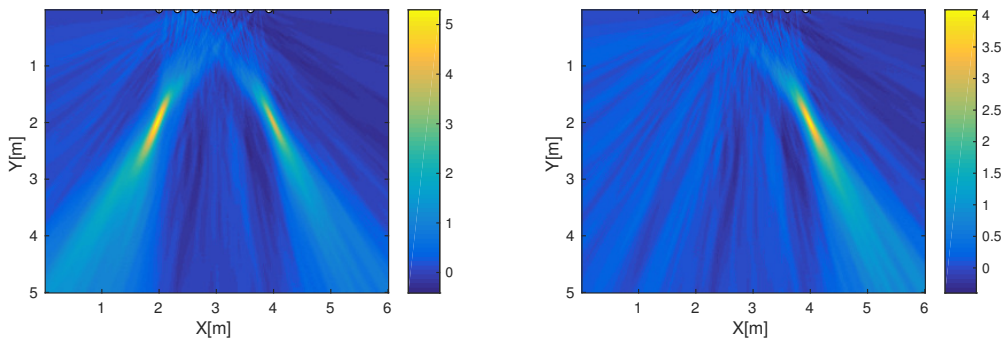


Abbildung 5.14: GCF-Map vor und nach der Maskierung

Abbildung 5.14 zeigt links eine GCF-Map mit dem globalen Maximum der

KAPITEL 5. EXPERIMENTE

RMS Fehler	dB	Winkel[°]	Abstand[m]	Anzahl
Sprecher 1	-2,3	0,49	0,12	83
Sprecher 2	0	0,24	0,07	69
Sprecher 1	-2,3	0,47	0,12	84
Sprecher 2	-10	0,40	0,18	34

Tabelle 5.8: Messwerte nach Suche des zweiten Maximums

ersten Schallquelle bei (2, 2) und das lokale Maximum der zweiten Schallquelle bei (4, 2). Auf der rechten Seite befindet sich die Neuberechnete GCF-Map nach der Maskierung der ersten Schallquelle. Die Farbskala der rechten Abbildung wurde an das neue globale Maximum angepasst.

Wendet man diese Methode auf das Experiment aus dem vorherigen Kapitel an erhält man die Ergebnisse in Tabelle 5.8. Vergleicht man diese mit 5.7, erkennt man erwartungsgemäß eine deutlich größere Anzahl an verwendeten Messwerten. Die Genauigkeit der einzelnen Werte hat sich allerdings leicht verringert, da auch das kleinere und weniger scharfe lokale Maximum der momentan leiseren Schallquelle in der GCF-Map ausgewertet wurde.

Da man jetzt nicht mehr darauf angewiesen ist jede Schallquelle zu erfassen, wenn sie gerade am lautesten ist, kann man die Blocklänge erhöhen und über mehrere GCF-Maps mitteln, um Auswirkungen des Rauschanteils im Mikrofonsignal zu reduzieren. Tabelle 5.9 zeigt das Ergebnis nach Erhöhung der Blocklänge auf 256ms und der Mittelung aufeinanderfolgender GCF-Maps mit einer Zeitkonstanten von $T_G=0,2s$ und $0,5s$. Die Verdoppelung der Blocklänge reduziert den Einfluss des Rauschens, halbiert jedoch die Anzahl der berechneten GCF-Maps von 98 auf 49 und damit die Anzahl der erzeugten Messwerte. Die Messgenauigkeit wird durch diese Maßnahmen deutlich verbessert. Die Erhöhung der Zeitkonstanten T_G von $0,2s$ auf $0,5s$ verbessert die Genauigkeit ebenfalls. Der Anteil der genutzten GCF-Maps für die Lokalisierung des um 10dB bedämpften Sprechers 2 ist mit $30/49$ bzw. $32/49$ größer als bei der Variante mit der kürzeren

RMS Fehler	dB	Winkel[°]	Abstand[m]	Anzahl
Sprecher 1	-2,3	0,16	0,04	48
Sprecher 2	-10	0,35	0,09	30
Sprecher 1	-2,3	0,00	0,01	49
Sprecher 2	-10	0,18	0,04	32

Tabelle 5.9: Verdoppelte Blocklänge auf 256ms mit $T_G=0,2s$ und $0,5s$

Blocklänge mit 34/98.

Prinzipiell lässt sich die Maskierung beliebig oft wiederholen, um weitere Schallquellen zu Orten. Je mehr simultane Schallquellen vorhanden sind, desto weniger scharf ausgeprägt werden ihre lokalen Maxima in der GCF-Map und desto ungenauer wird ihre Positionsbestimmung.

5.2.3 Alternatives Maskierungsverfahren

Je nach Lage der Schallquellen im Raum können die ihnen zugehörigen Stellen in den Kreuzkorrelationsfunktionen dicht beieinander liegen. Ein blindes Maskieren des Signalanteils der zuerst gefundenen Schallquelle würde den Signalanteil der anderen ebenfalls ganz oder teilweise auslöschen. Deshalb schlage ich eine andere Form der Maskierung vor, die diesen Nachteil in geringerem Maße aufweisen soll.

Dieses Verfahren geht von der Annahme aus, dass die von einer Schallquelle verursachte Signalspitze in der Kreuzkorrelationsfunktion symmetrisch ist und einseitig mit den Ausläufern der Signalspitze einer anderen Schallquelle überlagert ist. Es wird hier versucht, nur den symmetrischen Anteil zu entfernen und den Anteil anderer Schallquellen möglichst zu erhalten.

KAPITEL 5. EXPERIMENTE

Die Maskierungsvorschrift für die Kreuzkorrelationsfunktionen lautet:

$$\begin{array}{l}
 R_{AB}(m) \quad \leftarrow \quad 0 \\
 R_{AB}(m+k) \quad \leftarrow \quad \left\{ \begin{array}{l} 0 \\ R_{AB}(m+k) - R_{AB}(m-k) \end{array} \right. \quad \left| \quad \begin{array}{l} R_{AB}(m+k) < R_{AB}(m-k) \\ \text{sonst} \end{array} \\
 R_{AB}(m-k) \quad \leftarrow \quad \left\{ \begin{array}{l} R_{AB}(m-k) - R_{AB}(m+k) \\ 0 \end{array} \right. \quad \left| \quad \begin{array}{l} R_{AB}(m+k) < R_{AB}(m-k) \\ \text{sonst} \end{array}
 \end{array} \quad (5.4)$$

$$k \in [1; 1+5U].$$

In (5.4) bezeichnet R_{AB} die Kreuzkorrelationsfunktion der Mikrofone A und B. Der zur maskierenden Schallquelle zugehörige Abtastwert in R_{AB} ist hier mit m bezeichnet. Er wird zu null gesetzt. Vergleichen wir nun die beidseitig um k von m entfernten Abtastwerte in R_{AB} . Die jeweils kleineren Werte setzen wir zu null. Auf der jeweils anderen Seite erhalten wir nur die Differenz der beiden Werte, um den sich überlappenden Signalanteil einer möglichen weiteren nebenan liegenden Schallquelle zu erhalten.

Im Idealfall mit nur einer Schallquelle und ohne Rauschen produziert die GCC-PHAT einen um die entsprechende Zeit verschobenen sinc-Funktion mit der „Breite“ von einem Abtastwert. Sind mehrere Schallquellen und Rauschen vorhanden, verbreitert sich dieser Ausschlag. In der Praxis scheint ein Arbeitsbereich von 5 Abtastwerten links und rechts von m geeignet zu sein. Bei Anwendung der Abtaststratenerhöhung mit dem Faktor U ergibt sich für k ein Wertebereich von 1 bis $1+5U$.

Betrachten wir nun ein simuliertes Beispiel mit einem linearen Array von $2m$ Breite bei $y=0$ und zwei Sprechern im Abstand von $2m$ und $3m$ mittig davor. Die beiden Sprecher liegen aus der Sicht des Arrays direkt hintereinander. Dies führt dazu, dass sich ihre Signalspitzen in den Kreuzkorrelationsfunktionen bei vielen Mikrofonpaaren ganz oder teilweise überlagern. Das macht die Maskierung besonders schwierig. Das mittlere SNR an den Mikrofonen beträgt 12,6dB.

KAPITEL 5. EXPERIMENTE

Die zugehörige GCF-Map zu einem bestimmten Zeitpunkt ist in Abbildung 5.15 dargestellt.

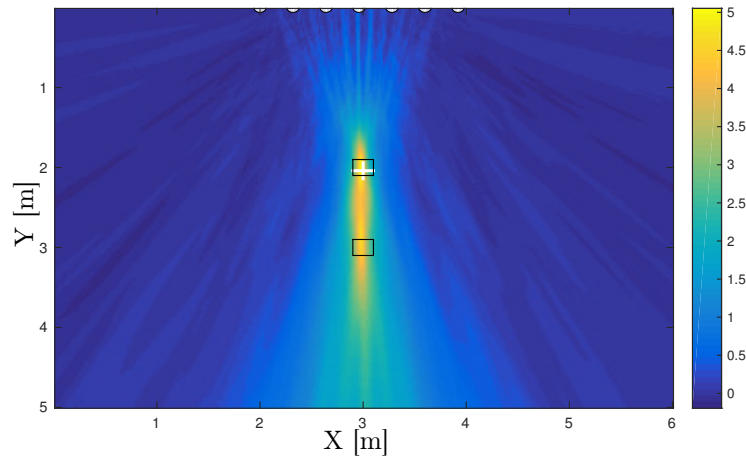


Abbildung 5.15: GCF-Map vor Anwendung der Maskierung

Mit Rechtecken markiert sind die Orte der simulierten Sprecher. Das weiße Kreuz markiert den Ort des Maximums der mit einer Zeitkonstanten von $T_G=0,5s$ gemittelten GCF-Map. Es markiert ziemlich gut den Ort der näheren und damit von den Mikrofonen lauter aufgenommenen Schallquelle.

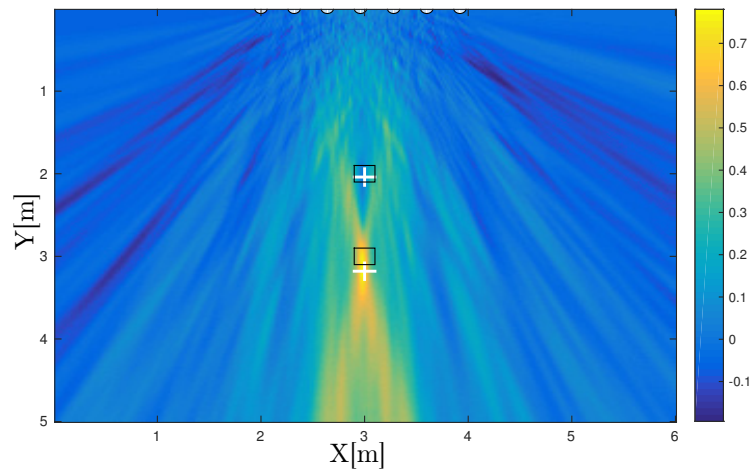


Abbildung 5.16: Maskierung nach Brutti et al. (5.3)

KAPITEL 5. EXPERIMENTE

Abbildung 5.16 zeigt die Neuberechnete GCF-Map nach der Maskierung mit der Funktion (5.3) von Brutti et al. wie im vorherigen Kapitel 5.2.2 mit den Parametern $p=2$ und $b=3,5$. Ebenfalls markiert mit einem Kreuz ist hier das Maximum der modifizierten GCF-Map, das zur zweiten Schallquelle gehört.

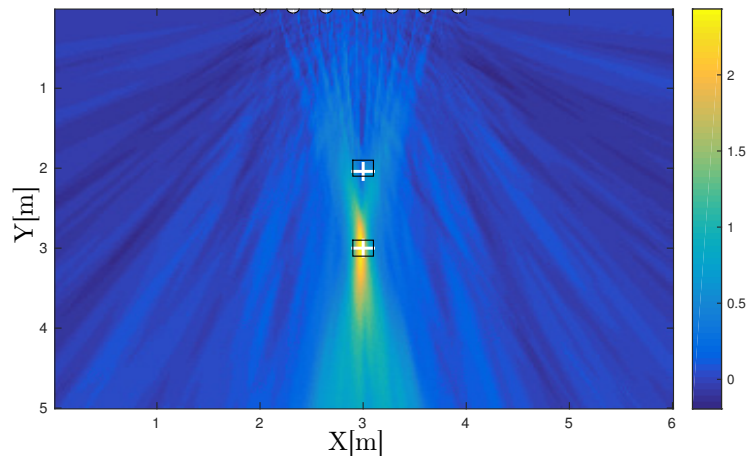


Abbildung 5.17: Vorgeschlagene Maskierung

Abbildung 5.17 zeigt die Neuberechnete GCF-Map mit dem hier vorgeschlagenem Maskierungsverfahren. Das Maximum hat hier eine größere Amplitude und sein Ort stimmt genauer mit der zu ortenden zweiten Schallquelle überein.

Abbildung 5.18 zeigt als Beispiel die Kreuzkorrelationsfunktion der Mikrofone 4 und 6 im Originalzustand (blau), nach der Maskierung durch die Funktion von Brutti et al. (rot) und mit dem hier gezeigten Maskierungsverfahren (orange).

In diesem dargestellten Ausschnitt beträgt $m=68$. Man sieht die orange Kurve hat ein größeres Maximum als die rote Kurve. Dieses Maximum befindet sich geringfügig näher an m . Das Maximum der leiseren Schallquelle wurde also weniger weit durch die Maskierung der lautereren „verdrängt“. Dieses „Verdrängen“ kann man auch in Abbildung 5.16 als eine Verschiebung des Maximums (weißes Kreuz) nach unten in der GCF-Map sehen.

Tabelle 5.10 zeigt einen Vergleich der beiden gezeigten Methoden. Das vorgeschlagene Maskierungsverfahren ist in diesem Beispiel bei der aus der Sicht des

KAPITEL 5. EXPERIMENTE

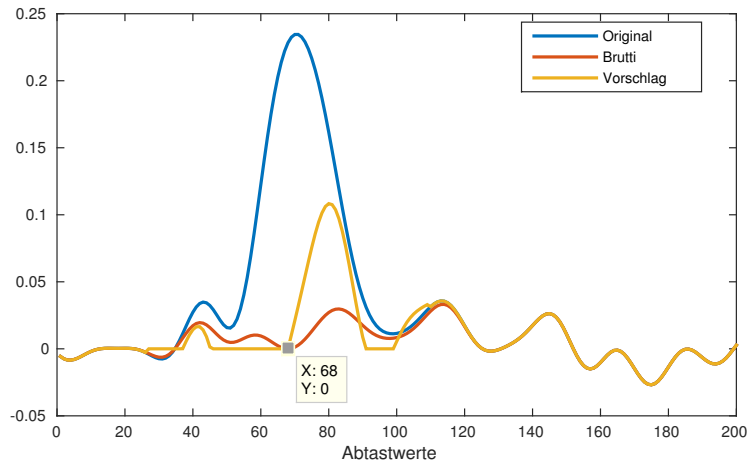


Abbildung 5.18: Vergleich der Kreuzkorrelationsfunktionen

RMS Fehler	Nr.	Winkel[°]	Abstand[m]	Anzahl	FP
Brutti et al.	1	0,10	0,14	49	3
$p=2$ $b=3,5$	2	0,22	0,52	45	–
Brutti et al.	1	0,05	0,08	49	49
$p=1$ $b=2$	2	0,24	0,22	48	–
Vorschlag	1	0,06	0,09	49	2
	2	0,13	0,23	48	–

Tabelle 5.10: Messwerte bei verschiedenen Maskierungsfunktionen

Mikrofonarrays hinten liegenden Position Nr. 2 mehr als doppelt so genau wie die bisherige Methode mit den empfohlenen Parametern $p=2$ und $b=3,5$.

Ändert man die Parameter auf z.B. $p=1$ und $b=2$ erhält man eine schmalere und spitzere Maskierungsfunktion. Diese wird ebenfalls in Abbildung 5.13 gezeigt. Damit erhält man fast so gute Genauigkeitswerte wie bei der vorgeschlagenen Methode.

Die Maskierungsfunktion soll nicht nur lokale Maxima in der GCF-Map von noch nicht identifizierten Schallquellen möglichst gut erhalten, sondern muss auch das bereits gefundene Maximum in der GCF-Map vollständig maskieren. Wird

KAPITEL 5. EXPERIMENTE

es nicht vollständig maskiert, ist es möglich, dass die „Überreste“ erneut als eine eigenständige Schallquelle erkannt werden.

Macht man nun die Gegenprobe und deaktiviert die Schallquelle 2, erhält man 49 falsch positive Erkennungen (FP) für aktive Schallquellen nach der Maskierung mit $p=1$ und $b=2$. Dies ist darauf zurückzuführen, dass die Signalspitzen der zuerst erkannten Schallquelle bei diesen Parametern nicht vollständig aus den Kreuzkorrelationsfunktionen entfernt wurden. Bei den ursprünglich gewählten Parametern lag diese Zahl bei 3, und bei der neu vorgestellten Methode bei 2.

Für einen praktischen Einsatz sollte also ein höherer Schwellwert für die Maximalwerte der GCF-Map zur Erkennung einer aktiven Schallquelle gewählt werden. Verdoppelt man den Schwellwert verursachen die Parameter $p=1$ und $b=2$ immernoch 35 falsch positive Erkennungen, während sich die Zahl für die anderen beiden Varianten auf 0 reduziert.

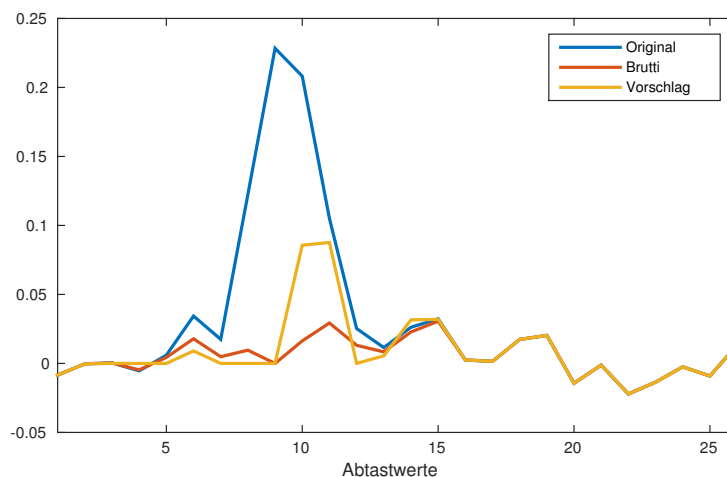


Abbildung 5.19: Vergleich der Kreuzkorrelationsfunktionen ohne Abtastratenerhöhung

Die in Tabelle 5.10 gezeigten Ergebnisse und die in Abbildung 5.18 gezeigten Kreuzkorrelationsfunktionen beziehen sich auf Mikrofonsignale mit 16kHz und

KAPITEL 5. EXPERIMENTE

RMS Fehler	Nr.	Winkel[°]	Abstand[m]	Anzahl	FP
Brutti et al.	1	0,55	0,15	50	5
$p=2$ $b=3,5$	2	0,21	0,59	43	–
Vorschlag	1	0,27	0,11	52	42
	2	0,18	0,31	46	–

Tabelle 5.11: Messwerte bei verschiedenen Maskierungsfunktionen mit $U=1$

einer Abtastratenerhöhung (Kap. 4.4) um den Faktor $U=8$. Die Maskierung einer Schallquelle in einer Kreuzkorrelationsfunktion erfolgt bei beiden Methoden mit einer Genauigkeit von einem Abtastwert. Die hier vorgeschlagene Methode funktioniert mit der Annahme, daß Signalspitzen symmetrisch sind. Die Abtastratenerhöhung hilft dabei, genauer in der Mitte einer Signalspitze mit der Maskierung zu beginnen. Abbildung 5.19 zeigt dieselben Maskierungen wie in Abbildung 5.18 jedoch ohne Abtastratenerhöhung.

Tabelle 5.11 zeigt die Auswertung des gleichen Szenarios mit dem gleichen Schwellwert wie bei Tabelle 5.10 nur ohne Abtastratenerhöhung. Die Genauigkeit hat sich bei beiden Methoden verschlechtert, jedoch ist immer noch ein Vorteil bei der vorgeschlagenen Methode vorhanden. Für die Entfernung der zuerst gefundenen Schallquelle aus der GCF-Map wird das Weglassen der Abtastratenerhöhung bei beiden Methoden zum Nachteil. So ist die Anzahl der Schallquelle 1 zugeordneten Messwerte mit 50 und 52 größer als die Zahl der ausgewerteten Signalblöcke, welche 49 beträgt. In manchen Fällen lag also das gefundene Maximum in der GCF-Map nach der Maskierung näher an Schallquelle 1 und wurde daher erneut dieser zugeordnet. Macht man auch hier die Gegenprobe mit deaktivierter Schallquelle 2 führt die Methode von Brutti et al. zu 5 falsch positiven Erkennungen und die hier vorgeschlagene Methode sogar zu 42.

Will man mehrere Schallquellen in einer GCF-Map zuverlässig erkennen und lokalisieren ist die Abtastratenerhöhung bei beiden Maskierungsverfahren

KAPITEL 5. EXPERIMENTE

hilfreich, und bei der vorgeschlagenen Methoden sogar notwendig, um Schallquellen aus GCF-Maps vollständig zu entfernen und um den Schwellwert für die Erkennung einer Schallquelle möglichst niedrig wählen zu können.

Kapitel 6

Zusammenfassung

In dieser Arbeit wurde gezeigt, dass die GCF gut geeignet ist um Schallquellen zu lokalisieren. Die Wirkung von Rauschen lässt sich durch eine Verlängerung der Beobachtungsdauer über die Blocklänge oder die Mittelung aufeinanderfolgender GCF-Maps kompensieren. Auf Nachhall und Reflexionen der Schallquellen reagiert die GCF sehr robust. Die Genauigkeit von Abstandsmessungen hängt erwartungsgemäß stark von der Breite des Arrays und von der Entfernung der Schallquelle ab. Interessant ist der Unterschied zwischen einem linearen und einem zirkularem Array vergleichbarer Breite. Das lineare Array liefert hierbei die zuverlässigeren Ergebnisse.

Mithilfe der GCF lassen sich auch simultan aktive Schallquellen lokalisieren. Das dazu hier neu vorgestellte Maskierungsverfahren bietet bessere Ergebnisse als das zuvor bekannte, sofern man die Abtastratenerhöhung für die Kreuzkorrelationsfunktionen nutzt.

Prinzipiell ist die GCF sehr rechenaufwendig, dennoch ist ein aktueller PC in der Lage bei entsprechend gewählten Parametern und bei nicht übermäßig großen Arrays die GCF-Map Berechnung in Echtzeit mit MATLAB durchzuführen. Die Experimente in dieser Arbeit sind sehr theoretisch und zielten auf keinen konkreten Anwendungsfall ab. Praktische Versuche zur Wahl einer geeigneten Mikrofonanordnung und praktikablen Berechnungsparametern sind vor einem Einsatz noch

KAPITEL 6. ZUSAMMENFASSUNG

notwendig. Optimierungen in der Signalverarbeitung sind ebenfalls noch möglich und sollten untersucht werden. Dies könnten z.B. verschiedene Auflösungsstufen der GCF-Map und dynamisch veränderliche Regionen mit hoher Auflösung sein. Die Auswahl der ausgewerteten Mikrofonpaare könnte gleichfalls dynamisch an die aktuelle akustische Situation angepasst werden. Es wird nur ein relativ kleiner Bereich in der Mitte der Kreuzkorrelationsfunktionen benötigt. Dies birgt weiteres Einsparungspotential. Da die GCF-Maps meistens unterschiedlich scharf sind bezüglich des Winkels und des Abstandes zum Mikrofonarray, könnte eine Berechnung in Polarkoordinaten helfen die Anzahl der zu berechnenden Gitterpunkte zu reduzieren ohne eine geringere Genauigkeit in Kauf zu nehmen. Die Erkennung von aktiven Schallquellen, deren Zuordnung und Weiterverarbeitung müsste ebenfalls noch für einen konkreten Einsatz gestaltet werden.

Literaturverzeichnis

- [BOS08] A. Brutti, M. Omologo, and P. Svaizer. Localization of multiple speakers based on a two step acoustic map analysis. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4349–4352, March 2008.
- [BOS09] A. Brutti, M. Omologo, and P. Svaizer. A sequential monte carlo approach for tracking of overlapping acoustic sources. In *Signal Processing Conference, 2009 17th European*, pages 2559–2563, Aug 2009.
- [BW01] M. Brandstein and D. Ward, editors. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, New York, 2001.
- [KC76] C.H. Knapp and G.C. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-24(4):320–327, Aug. 1976.