

# Bachelorarbeit

Katja Kirstein

Halbautomatischer Koordinatenabgleich zwischen einem  
optischen Positionserfassungssystem und einer  
Wellenfeldsynthese-Anlage

Katja Kirstein

# Halbautomatischer Koordinatenabgleich zwischen einem optischen Positionserfassungssystem und einer Wellenfeldsynthese-Anlage

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung  
im Studiengang Bachelor of Science Technische Informatik  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr.-Ing. Wolfgang Fohl  
Zweitgutachter: Prof. Dr. Andreas Meisel

Eingereicht am: 1. Februar 2021

**Katja Kirstein**

**Thema der Arbeit**

Halbautomatischer Koordinatenabgleich zwischen einem optischen Positionserfassungssystem und einer Wellenfeldsynthese-Anlage

**Stichworte**

Wellenfeldsynthese, optische Lokalisierung, Infrarot-Tracking, Audiolokalisierung, Ambisonics, Soundfield-Mikrofon, homogene Transformation

**Kurzzusammenfassung**

In dieser Arbeit wird eine Methode zur Bestimmung der Koordinatentransformation zwischen einer Wellenfeldsynthese-Anlage und dessen optischem Ortungssystem untersucht. Hierbei dient ein Soundfield-Mikrofon als Schnittstelle der beiden Systeme, um mittels Triangulation die akustische Quelle im visuellen Ortungssystem zu bestimmen. Die Grundlagen der Systeme werden ebenso erläutert, wie die Vorgehensweise bei der Datenerfassung. Anschließend folgt die Auswertung der in Bezug auf die homogene Transformation stehenden Daten. Insbesondere wird hier auf die Datenqualität des Mikrofons eingegangen.

**Katja Kirstein**

**Title of Thesis**

Semi-automatic coordinate alignment between an optical positioning system and a wave field synthesis system

**Keywords**

Wave field synthesis, optical positioning, infrared tracking, audio localization, homogeneous coordination transformation, Ambisonics, Soundfield microphone

**Abstract**

This thesis examines a method to determine the coordination transformation between a wave field synthesis system and an optical positioning system. An Ambisonics Soundfield-Microphone is used to gather and triangulate coordinates to create a foundation for both systems. The basics of both systems will be explained as well as the approach to collect data. Subsequently, the data is evaluated in relation to the coordination transformation, especially in terms of data quality of the microphone.

# Inhaltsverzeichnis

Abbildungsverzeichnis	v
Abkürzungen	viii
Symbolverzeichnis	ix
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Struktur der Arbeit . . . . .	2
<b>2 Grundlagen</b>	<b>3</b>
2.1 Wellenfeldsynthese . . . . .	4
2.1.1 Audioquellen . . . . .	5
2.1.2 Probleme der WFS-Anlage . . . . .	6
2.2 Positionserfassungssystem . . . . .	7
2.2.1 Targets . . . . .	8
2.2.2 Tracking- und WFS-Koordinatensystem . . . . .	9
2.2.3 Motion Tracker-Wave Field Synthesis-Connector . . . . .	11
2.3 Ambisonics . . . . .	12
2.3.1 Kugeloberflächenfunktionen . . . . .	12
2.3.2 Soundfield-Mikrofon und Berechnung des B-Formats . . . . .	15
2.3.3 Speicherformate . . . . .	16
2.3.4 Koordinatensystem des Soundfield-Mikrofons . . . . .	17
<b>3 Schallquellenlokalisierung und Koordinatentransformation</b>	<b>18</b>
3.1 Richtungsbestimmung mit Ambisonics . . . . .	19
3.2 Triangulation . . . . .	20
3.3 Homogene Transformation . . . . .	23

<b>4</b>	<b>Aufbau und Durchführung der Messungen</b>	<b>26</b>
4.1	Hardware-Komponenten . . . . .	26
4.2	Datenaufnahme und -verarbeitung . . . . .	29
4.2.1	Erfassen und Speichern der Daten . . . . .	29
4.2.2	Verarbeitung der Daten . . . . .	32
4.3	Durchführung der Messung . . . . .	34
<b>5</b>	<b>Auswertung der Messungen</b>	<b>35</b>
5.1	Messtechniken . . . . .	35
5.1.1	Messgenauigkeit . . . . .	35
5.1.2	Sinussignal vs weißes Rauschen . . . . .	37
5.1.3	Unterschiedliche Punktquellen . . . . .	40
5.1.4	MoWeC Persontracking . . . . .	41
5.2	Durchschreitung der fokussierten Punktquelle . . . . .	43
5.3	Bewertung der unterschiedlichen Messungen . . . . .	45
5.4	Berechnung der Transformation . . . . .	45
<b>6</b>	<b>Fazit und Ausblick</b>	<b>50</b>
<b>A</b>	<b>Anhang</b>	<b>56</b>
A.1	Config-File . . . . .	56
A.2	Häufige Fehler bei den Messungen . . . . .	57
	<b>Selbstständigkeitserklärung</b>	<b>59</b>

# Abbildungsverzeichnis

2.1	Skizze des Wellenfeldsynthese (WFS)-Labors der HAW [Nog12]	3
2.2	WFS Labor der HAW Hamburg [Foh]	5
2.3	Schallwellenentwicklung unterschiedlicher virtueller Quellen (externe Punktquelle, lineare Quelle und fokussierte Punktquelle) [Mon11]	6
2.4	Korrektur Wahrnehmungsbereich einer fokussierten Punktquelle [Mon11]	7
2.5	Optisches Tracking-Prinzip des A.R.T.-Systems [ART12]	8
2.6	Ungefähre Lage der Koordinatensysteme im WFS-Labor	10
2.7	Tracking-Datenübertragung zur WFS	11
2.8	Sphärische Koordinaten <sup>1</sup>	13
2.9	Ambisonics-Ordnung nach Furse Malham <sup>2</sup>	14
2.10	Tetraeder	15
2.11	Soundfield-Mikrofon SPS200 <sup>3</sup>	15
2.12	Soundfield-Mikrofon SPS200 mit den Achsen nach der Umrechnung ins B-Format	17
3.1	Messungen für die akustische Triangulation	20
3.2	Punkt P des Koordinatensystems B dargestellt in A	23
4.1	Komponenten-Diagramm der Hardware	27
4.2	Halterung mit Target TT5 und SPS200 Soundfield-Mikrofon	28
4.3	Sequenzdiagramm der Datenerfassung für ein Datenset	30
4.4	Ablauf der Datenverarbeitung von einem Messdurchlauf	33
5.1	Messung einer fokussierten Punktquelle mit weißem Rauschen auf einer geraden Linie zur Quelle	36
5.2	Exemplarischer Konfidenzintervall von drei sich schneidenden Geraden	37
5.3	Messung einer fokussierten Punktquelle mit einem Sinussignal (220 Hz) ohne Persontracking	39

5.4	Messung einer fokussierten Punktquelle mit weißem Rauschen (TPF 1,5k Hz) ohne Persontracking . . . . .	39
5.5	Messung einer externer Punktquelle mit weißem Rauschen ohne Persontracking . . . . .	40
5.6	Messung einer fokussierter Punktquelle mit weißem Rauschen und Persontracking . . . . .	42
5.7	Messung mit fokussierter Punktquelle mit weißem Rauschen ohne Persontracking . . . . .	42
5.8	A-Format Aufnahme während der Durchschreitung einer fokussierten Punktquelle mit weißem Rauschen . . . . .	44
5.9	Anwendung der Koordinatentransformation nach dessen Identifizierung mit den ermittelten Schnittpunktwerten . . . . .	46
5.10	Plot des Kostenfunktional der vom MoWeC transformierten Punkte . . . . .	47
5.11	Ebene des minimalen quadratischen Fehlers . . . . .	48
5.12	Anwendung der Koordinatentransformation nach dessen Identifizierung mit Referenzwerten aus dem Kostenfunktionalbeispiel . . . . .	48

# Abkürzungen

**ASIO** Audio Stream Input/Output.

**DoF** Degrees of Freedom.

**FuMa** Furse-Malham.

**GUI** Graphical User Interface.

**KOS** Koordinatensystem.

**MoWeC** Motion Tracker-Wave Field Synthesis-Connector.

**OSC** Open Sound Control.

**RegEx** Regular Expressions.

**UDP** User Datagram Protocol.

**WFS** Wellenfeldsynthese.

# Symbolverzeichnis

- $\alpha$  Rotationswinkel der Koordinatentransformation.
- $\Theta, \vartheta$  Elevation zur Ambisonics-Schallfeldbeschreibung.
- $\mu, \theta, \phi$  Eulerwinkel zur Beschreibung der Ausrichtung von Targets im Tracking-System.
- $\gamma$  Ausgehend des Ambisonics-Mikrofons der Richtungswinkel zur Audioquelle.
- $\Phi, \varphi$  Azimut zur Ambisonics-Schallfeldbeschreibung.
- $p$  Schalldruck.
- $v$  Schallschnelle.

# 1 Einleitung

Die WFS-Anlage ist Teil des Labors für interaktive immersive Audiosysteme ( $i^2$ audiolab) [Foh] und dient zur Erzeugung von räumlichem Ton. Die Anlage besteht aus einem Lautsprecherrechteck von  $5 \times 6$  Metern. Innerhalb dieser Versuchsfläche ist das System in der Lage definierte Schallwellen zu rekonstruieren. Dadurch kann die Anlage Klangquellen erzeugen, die sich nicht in den Lautsprechern, sondern im Raum wiederfinden. Dabei können sich die Hörer:innen um die Klangquellen herumbewegen, als ob es sich um reale, im Raum befindliche Objekte handelt.

Ausgestattet ist die Anlage mit einem Positionserfassungssystem zur optischen Lokalisierung (Tracking-System der Firma A.R.T.). Mit diesem System können die Position und Orientierung von Objekten (Pose) erfasst werden. Anhand der Objektpose wird mit der WFS die Interaktion ermöglicht. Damit können zum Beispiel Schallquellen auf physikalische Objekte innerhalb der Anlage projiziert werden, dies simuliert die Klangquellen aus Richtung des Objekts.

Das Positionserfassungssystem verwendet sechs Infrarotkameras. Jede Kamera arbeitet dabei mit ihrem eigenen Koordinatensystem (KOS). Wenn die einzelnen Kameradaten in einem übergeordnetem Raum abgebildet und kombiniert werden, kann ein Objekt als solches im Aufnahmebereich erkannt werden. Die Koordinatentransformation dient dazu, die visuellen Koordinaten in dem übergeordnetem KOS abzubilden.

Eine Problematik besteht darin, dass die beiden Anlagen unterschiedliche Koordinatensysteme verwenden. Um die Interaktion miteinander zu ermöglichen, müssen die Daten vom Tracking-System in das KOS der WFS transformiert werden. Dadurch, dass es sich hierbei um ein akustisches und ein visuelles System handelt, ist die Durchführung des Koordinatenabgleichs kompliziert.

## 1.1 Motivation

Bisher basiert der Koordinatenabgleich der WFS-Anlage und des Tracking-Systems auf einer von Menschen mit Maßband und Gehör bestimmten Transformation. Diese ermöglicht es Daten vom Tracking-KOS ins WFS-KOS umzurechnen. Zukünftig soll die Koordinatentransformation durch einen mathematischen Ansatz präzisiert werden, wodurch die Immersion zwischen realer und virtueller Welt verbessert wird. Dafür sollen die Daten des Tracking-Systems sowie die Daten der WFS genauer aufeinander abgestimmt werden. Demnach ist ein Ziel dieser Arbeit die akustische Lokalisierung der durch die WFS-Anlage erzeugten Geräuschquelle im Tracking-System zu bestimmen. Dies geschieht mithilfe eines Soundfield-Mikrofons, welches optisch getrackt wird. Dabei wird von dem dreidimensionalen Tracking zur zweidimensionalen WFS reduziert.

Nach der Gewinnung von den visuellen und akustischen Daten, folgt die Kombination der beiden Datentypen. Die kombinierten Daten werden zur Berechnung der Koordinatentransformation zwischen der WFS-Anlage und dem Tracking-System verwendet. Audiosysteme besitzen ihre eigenen Koordinatensysteme, die von ihrer Funktionsweise abhängig sind. Bei dem Kombinieren von unterschiedlichen KOS entsteht immer das Problem der absoluten Orientierung. Zur Lösung der absoluten Orientierung wird die Technik der Triangulation und ihre Anwendbarkeit untersucht, da es sich um ein Standardverfahren zur Ortung handelt. Die Triangulation erfolgt mithilfe von Punkten im Raum, deren Position in den verschiedenen KOS bekannt sind.

## 1.2 Struktur der Arbeit

Als Erstes wird ein kurzer Einblick in die verwendeten Systeme gegeben. Die theoretischen Grundlagen der Schallquellenlokalisierung werden mithilfe der klassischen Triangulierung erläutert. Nachfolgend wird die homogene Transformation dargestellt. Hard- und Software für den verwendeten Versuchsaufbau werden beschrieben und durch eigene Software zur Datenauswertung ergänzt. Im darauf folgenden Kapitel werden die gemessenen Daten ausgewertet. Am Ende wird ein Fazit über die Ergebnisse dieser Arbeit gezogen und ein Ausblick in mögliche weiterführende Arbeiten gegeben.

## 2 Grundlagen

Das Tracking-System stellt ein Instrument zur Interaktion mit der WFS-Anlage dar. Beide Systeme, die WFS und das Tracking, arbeiten mit unterschiedlichen Koordinatensystemen im selben Raum. In Abbildung 2.1 ist der Lautsprecheraufbau des WFS-Labors mit den Tracking-Kameras zu sehen. Dabei kommunizieren die Systeme untereinander mithilfe des Motion Tracker-Wave Field Synthesis-Connectors, welcher zu einem späteren Zeitpunkt erörtert wird.

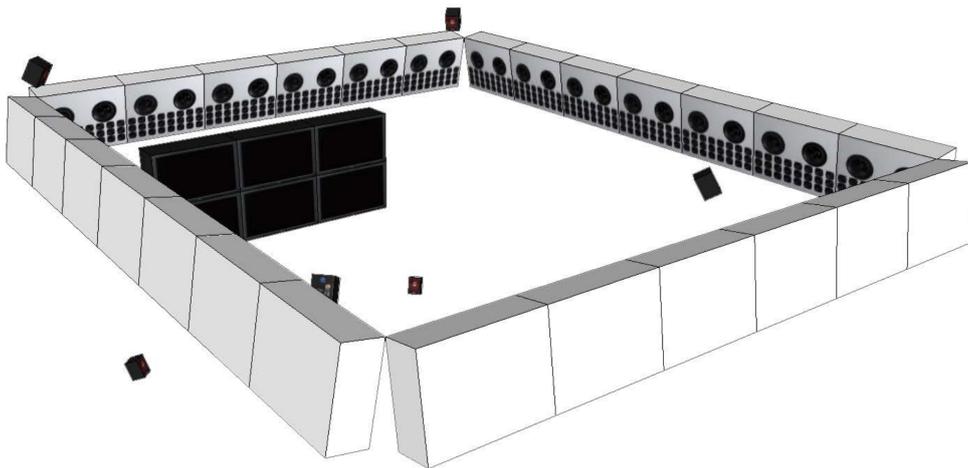


Abbildung 2.1: Skizze des WFS-Labors der HAW [Nog12]

In dieser Arbeit wird eine mathematische Bestimmung der Koordinatentransformation zwischen den beiden Systemen gesucht. Mit dieser können die Target-Punkte in das KOS der WFS-Anlage dargestellt werden. Für die Transformation werden Punkte benötigt die in beiden Systemen bekannt sind. Zur Bewältigung der Problematik wird mit dem Ambisonics-Soundfield-Mikrofon eine Kommunikationsschnittstelle geschaffen. Dieses Kapitel gibt einen Einblick in die unterschiedlichen Komponenten und wie sich deren Koordinatensysteme definieren.

Das Thema der WFS-Funktionsweise ([Ahr12], [Ber88] und [SRA08]) sowie das der optischen Lokalisierung [HZ03] ist zu umfangreich für diese Arbeit und kann in anderer Literatur ausführlicher nachgeschlagen werden. Für diese Arbeit ist es nicht zwingend erforderlich die Funktionsweisen im Detail zu verstehen, da diese bereits in der vorhandenen Software implementiert sind und hier nur Anwendung finden. Ein Grundverständnis der beiden Systeme ist jedoch Voraussetzung, um nachfolgend einen Bezug zur akustischen Lokalisierung herstellen zu können. Für die Lokalisierung hingegen ist es wichtig, das Ambisonics-Verfahren zu verstehen. Es liefert Daten, welche aufbereitet und verarbeitet werden müssen, bevor mit ihnen gearbeitet werden kann. Daher liegt der Schwerpunkt dieses Kapitels auf Ambisonics, während die anderen Themen nur kurz erörtert werden.

### 2.1 Wellenfeldsynthese

Zur Wiedergabe von Audiosignalen existieren unterschiedliche Möglichkeiten. Die klassischen Soundsysteme, wie bspw. 7.1, 5.1 und Stereo sind kanalbasiert. Bei kanalbasierten Systemen werden die Klänge auf verschiedenen Audiokanälen gemixt. Jeder Kanal repräsentiert dabei ein Audiosignal, welches auf einem Lautsprecher abgespielt wird. Durch die Wiedergabe eines Signals auf zwei oder mehr Kanälen kann eine räumliche Akustik erzeugt werden. Korrekt wahrnehmbar ist die entstandene Phantomschallquelle allerdings nur an einer beschränkten Position, dem Sweet Point.

Die WFS ist ein objektbasiertes Audiosystem [SRA08]. Der Unterschied zum Kanalbasierten ist, dass Klangquellen mit deren Positionsinformationen zur Laufzeit gerendert werden. Dadurch werden virtuelle Klangquellen synthetisiert, die sich ähnlich wie reale Tonquellen verhalten. Die akkurate räumliche Wahrnehmung der Klänge ist nur in dem Raum der WFS-Anlage möglich. Daher bezieht sich die vorliegende Arbeit auf den Bereich, der vom WFS-Lautsprecherrechteck begrenzt wird.

Im Labor (Abbildung 2.2) verwalten der Mac und die drei Rendering Cluster die virtuellen Schallquellen und ihre Daten, wie Position und Audioquellentyp. Der Rechner kommuniziert über das Netzwerk mit dem Open Sound Control (OSC)-Protokoll. Eine Graphical User Interface (GUI) visualisiert die Daten der Rendering Software Wonder ([BHSK07], [Win17]). Die WFS-Anlage wird mit OSC-Nachrichten gesteuert, diese ermöglichen das Erzeugen, Löschen und Bewegen von Audioquellen.

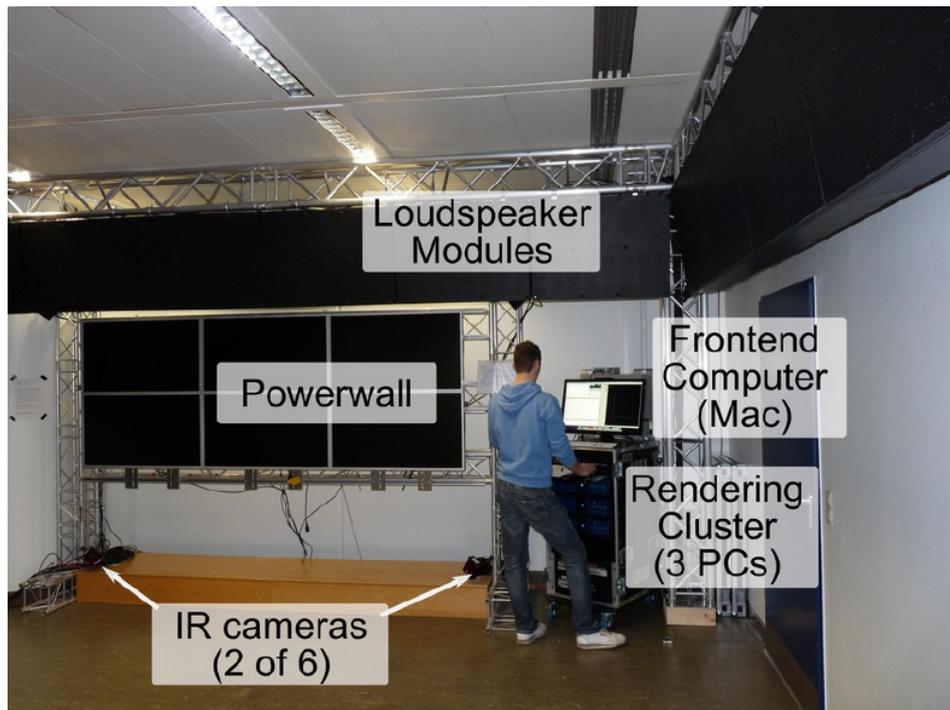


Abbildung 2.2: WFS Labor der HAW Hamburg [Foh]

### 2.1.1 Audioquellen

Die WFS funktioniert nach dem Huygensschen Prinzip [Rob60], welches Aussagen über die Ausbreitung von Wellen trifft. Es besagt, dass jeder Punkt einer Wellenfront ein Ausgangspunkt einer neuen Kreiswelle, bzw. Elementarwelle ist. Die neue Welle besitzt die gleiche Frequenz sowie Wellenlänge der ursprünglichen Kreiswelle. Generiert werden virtuelle Klangquellen von dicht aneinander liegenden, einzeln ansteuerbaren Lautsprechern. Für die Synthetisierung wird der Zeitpunkt berechnet, an dem die Elementarwelle der virtuellen Quelle auf die Lautsprecher trifft. Diese Verzögerung, die die Wellen der virtuellen Audioquelle brauchen, wird bei der Ausgabe auf den Lautsprechern mit rekonstruiert. Die Elementarwellen der einzelnen Lautsprecher überlagern sich und synthetisieren damit die Wellenfront der virtuellen Tonquelle. Erst durch diese Synthetisierung der Audioquellen wird die Lokalisierung der Position möglich.

In der WFS hat jede Audioquelle eine Position. Abhängig von der Position werden Audioquellen unterschieden in externe Punktquellen, lineare Quellen sowie fokussierte Punktquellen (Abbildung 2.3).

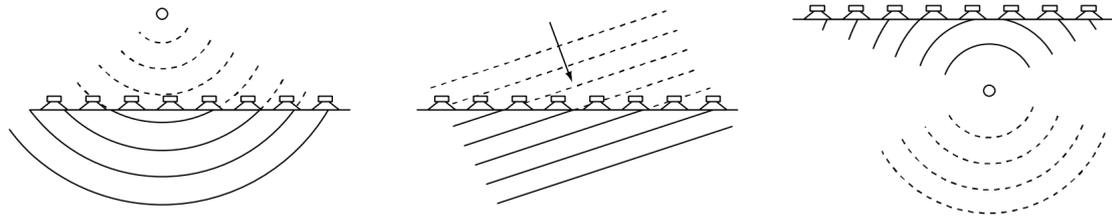


Abbildung 2.3: Schallwellenentwicklung unterschiedlicher virtueller Quellen  
(externe Punktquelle, lineare Quelle und fokussierte Punktquelle)  
[Mon11]

Bei **externen Punktquellen** breiten sich die Schallwellen kreisförmig aus. Der virtuelle Ursprung der Wellen befindet sich außerhalb des Raumes. Die linke Grafik zeigt die Ausbreitung der Schallwellen.

Mittig in Abbildung 2.3 ist die Ausbreitung einer **linearen Quelle** dargestellt. Die Wellen verlaufen in einer Linie und haben daher statt einer festen Quellenposition nur eine Richtung. Deren Positionslokalisierung ist folglich nicht möglich, sondern nur eine grobe Richtungsbestimmung.

Wenn sich die Quelle hingegen im Raum befindet, wird diese als **fokussierte Punktquelle** bezeichnet. Die Quelle wird anders gerendert als eine externe Quelle. Die Lautsprecher mit der geringsten Entfernung zur Quelle generieren die Schallwellen im Raum so, dass die Elementarwellen sich ab der Punktquelle konkav ausbreiten (Abbildung 2.3, rechts).

### 2.1.2 Probleme der WFS-Anlage

Bei fokussierten Quellen ist der Bereich für Hörer:innen, in dem die Quelle richtig wahrgenommen werden kann, deutlich geringer als bei externen Punktquellen. Denn zwischen Quelle und rendernden Lautsprechern ist diese nicht korrekt lokalisierbar, sondern wird aus der Lautsprecherrichtung wahrgenommen (Abbildung 2.4). Dies verhindert eine realistische Geräuschortung. Mittels Persontracking kann das Problem behoben werden. Dabei wird ein Target zur Lokalisierung der Personenposition verwendet. So können, abhängig der Position, entsprechende Lautsprecher zum Rendern der fokussierten Punktquelle angesteuert werden. [Chr14]

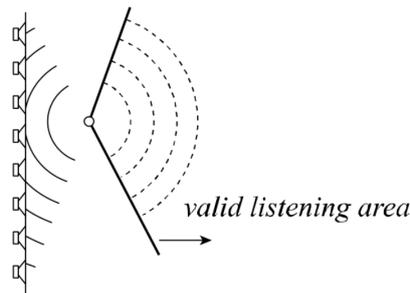


Abbildung 2.4: Korrekter Wahrnehmungsbereich einer fokussierten Punktquelle [Mon11]

Ein anderer Aspekt der zu beachten ist, ist dass die WFS-Anlage nicht dreidimensional ist. Sie beschallt nur eine Ebene in einem kartesischen Raum. Eine Einschätzung die Anlage um die Höhenebene zu erweitern, wurde bereits erarbeitet [Kü16]. Bisher wurde sich gegen eine dauerhafte Implementierung der Höhenkomponente entschieden. Folglich werden Klangquellen nur in einer Ebene erzeugt.

Weiterhin zu beachten ist, dass für das Rendern von Schallquellen eine Vielzahl von Lautsprecherkanälen berechnet und angesteuert werden. Bei den Schallwellen handelt es sich dabei aber nur um eine Approximation der realen Quellen. Deshalb kann es zu physikalischen Artefakten kommen. Ein weiteres physikalisches Problem hierbei ist der rechteckige Aufbau der WFS-Anlage. Bei der Erzeugung von Quellen in den Ecken kann die Ansteuerung der Lautsprecher eine verschwommene Rekonstruktion erzeugen. Dies hat zur Folge, dass die Lokalisierung einer Punktquelle im Bereich der Ecken ungenau wird.

## 2.2 Positionserfassungssystem

Um Objekte im dreidimensionalen Raum zu lokalisieren, sind sechs Freiheitsgrade (Degrees of Freedom (DoF)) zu berücksichtigen:

- Drei DoF für die Position des Objektes. Bei Verwendung von kartesischen Koordinaten sind dies die  $x$ -,  $y$ - und  $z$ - Werte des Schwerpunktes.
- Drei DoF für die Orientierung des Objektes, diese können z. B. mit den Eulerwinkeln oder Quaternionen beschrieben werden.

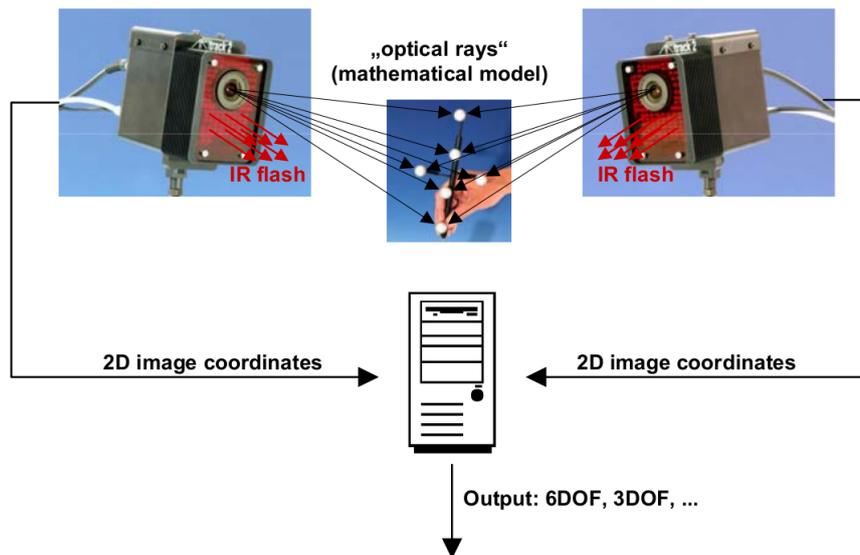


Abbildung 2.5: Optisches Tracking-Prinzip des A.R.T.-Systems [ART12]

In dem WFS-Labor wird das Positionserfassungssystem von A.R.T. (Advanced Realtime Tracking) verwendet. Es ist ein optisches System mit sechs Kameras sowie dazugehörigen Infrarot-Blitzgeräten, einer Steuerung und Auswerteeinheit. Die Kameras nehmen Objekte, genannt Targets, aus unterschiedlichen Positionen auf. Um ein Target mit Koordinaten in dem Aufnahmebereich zu erkennen, werden die 2D Bilder der einzelnen Kameras verwendet. Mit diesen Daten kann die Pose (Position und Orientierung) eines dreidimensionalen Objektes von einem Rechner ermittelt werden. Abbildung 2.5 zeigt exemplarisch die Funktionsweise anhand von zwei Infrarotkameras, die ein Target aufnehmen.

### 2.2.1 Targets

Ein Target besteht aus mehreren fest verbundenen Kugeln. Jede der Kugeln ist mit Reflexionsfolie überzogen und dient als passiver Marker. Die Marker werden durch Infrarotblitze angeleuchtet und mit der Kamera aufgenommen. Ein einzelner Marker besitzt die Angabe der Position und hat somit drei DoF. Zur Bestimmung der Pose eines Targets im Raum ist allerdings auch die Orientierung des Körpers erforderlich. Im System werden hierfür die unterschiedlichen Konstellationen der einzelnen Targets definiert. Dies ermöglicht die Lokalisierung im Raum.

Dabei wird die Position von jedem Target relativ zum eigenen Koordinatensystem angegeben. Der Ursprung des Target-Koordinatensystems kann auf einem der Kugeln liegen, im Mittelpunkt einer Kugel oder in einem speziellen Punkt des Targets liegen.

Die Target-Daten des A.R.T.-Systems bestehen aus den aufgelisteten Werten:

- ID
- X-,Y- und Z Koordinaten
- Euler Winkel
- Rotationsmatrix

Zur Objektdatenverarbeitung wird ein separater PC mit der A.R.T. Software verwendet. Dieser übergibt die Daten als String mit UDP Multicast in das lokale W-LAN. [ART06]

### 2.2.2 Tracking- und WFS-Koordinatensystem

Die kartesischen Koordinatensysteme (Abbildung 2.6) sind durch das 5 x 6 Meter Lautsprecherrechteck begrenzt. Der Ursprung des rechtshändigen, rechtwinkligen WFS-KOS besitzt den Koordinatenursprung mittig der Powerwall. Die Y-Achse geht positiv in Richtung des Versuchsraums.

Innerhalb der WFS-Anlage liegt auch der effektive Tracking-Bereich. Anders als das KOS der WFS ist das des Trackings linkshändig. Die X-Koordinaten sind vom Mittelpunkt zur Powerwall blickend nach rechts positiv, sowie die Y-Koordinaten positiv zur Powerwall hin steigen. Die Angaben der Software sind in Millimeter angegeben und bis auf einen halben Millimeter genau.

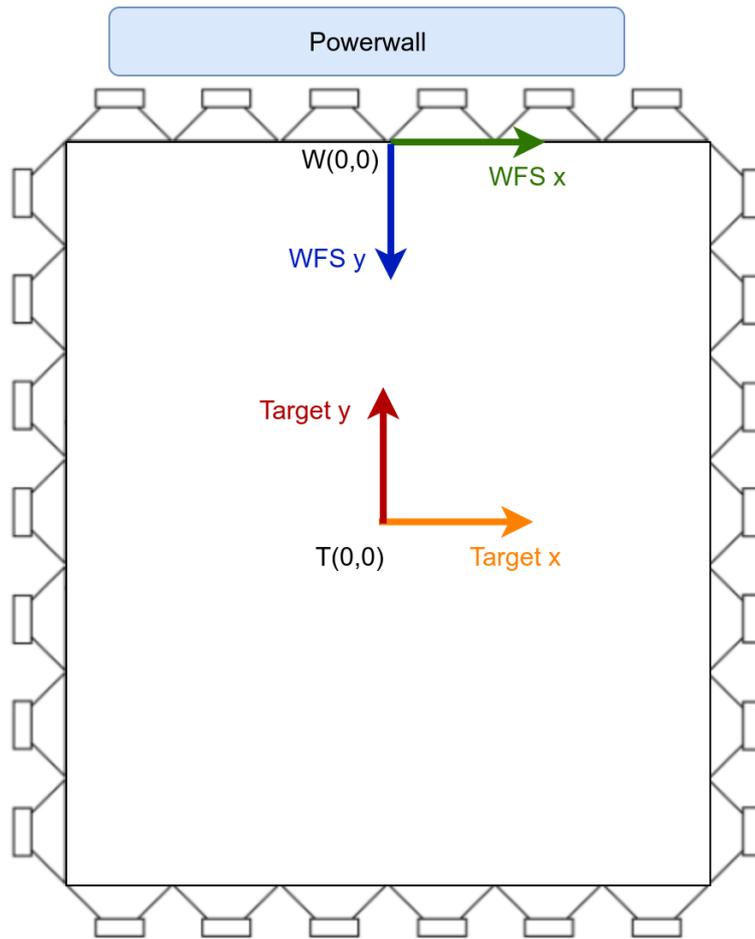


Abbildung 2.6: Ungefähre Lage der Koordinatensysteme im WFS-Labor

### 2.2.3 Motion Tracker-Wave Field Synthesis-Connector

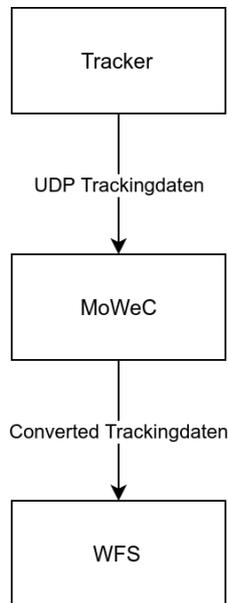


Abbildung 2.7: Tracking-Datenübertragung zur WFS

Bisher wurde die Transformation zwischen des WFS- und des Tracking-KOS manuell ermittelt. Hierfür wurde mit einem Marker der Nullpunkt des Tracking-KOS gesucht. Daraufhin wurde mit einem Zollstock die Entfernung bis zum Powerwall-Mittelpunkt (Abbildung 2.6) gemessen. Dieser Abstand ist mit dem WFS-Ursprung gleichzusetzen. Für ein normales Gehör ist dies ausreichend, aber für weiterführende Arbeiten zu ungenau. Daher wird eine mathematische Transformation zwischen WFS- und Tracking-System, mithilfe von gemeinsam bekannten Punkten, gesucht. Ein Ambisonics-Mikrofon dient dazu, die Klangquellen relativ zum Tracking-KOS zu lokalisieren.

Die Target-Daten werden von dem Tracker-PC über Multicast in das Netzwerk versendet. Diese Pakete enthalten Informationen aller erkannten Objekte des Trackers. Für die WFS werden die Daten auf dem WFS-Mac empfangen. Dieser Mac ist die Schnittstelle zwischen Benutzer und der WFS-Anlage.

Die Motion Tracker-Wave Field Synthesis-Connector (MoWeC)-Anwendung konvertiert die Target-Daten für die WFS, damit die Targets auf Audioquellen abgebildet werden können (Abbildung 2.7). Aus diesem Grund läuft der MoWeC auf dem Mac. Zur Koordinatentransformation im MoWeC werden einige Konverter genutzt. Sie unterteilen sich in Scaler, Shifter, Rotator und RotationMirror. Diese werden in der ConverterBox entsprechend zusammengefasst. [Nog12]

## 2.3 Ambisonics

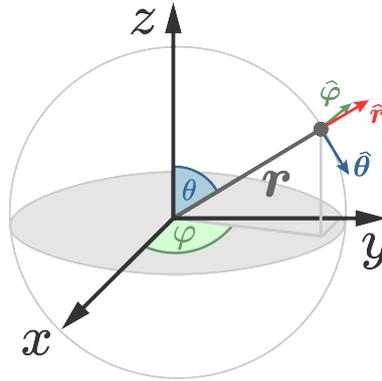
Ambisonics definiert eine schallfeldbasierte Audioaufnahme- und Wiedergabeart. Entstanden ist Ambisonics durch Michael Gerzon [Tho], der eine Möglichkeit suchte, räumlichen Sound aufzunehmen. Hierfür entwickelte er das Soundfield-Mikrofon, das in der Lage ist zusätzlich den Vertikalwinkel mit zu berücksichtigen. Das Format stellt die Beziehung der Richtungen im Raum, sowie die zeitliche Verzögerung der Signale dar. Somit kann die räumliche Aufnahme am Wiedergabeort synthetisiert werden, wie z.B. das Ambiente eines Konzertsaals. Daher wird der Name aus dem englischen *ambient* hergeleitet [Fel74].

Bei Ambisonics nimmt ein Mikrofon die eingehende Schalldruckrichtung von Objekten auf und beschreibt diese kugelförmig. Dieser Schalldruck wird zur Wiedergabe für die einzelnen Lautsprecher dekodiert. Verglichen mit der WFS, bei dem Position und Klang voneinander getrennt sind, ist bei Ambisonics die Richtung an den Klang gebunden. Außerdem besitzt Ambisonics einen Sweetspot an dem die Objekte ideal wahrgenommen werden können. Die WFS hingegen rekonstruiert Schallquellen auf einer Fläche.

### 2.3.1 Kugeloberflächenfunktionen

Mathematisch basiert Ambisonics auf der sphärischen Nahfeldholografie. Hier wird die Schallschnelle auf der Oberfläche einer Quelle bestimmt. Möglich ist dies mithilfe der Schallfeldrekonstruktion aus einem bekannten zweidimensionalen Schallfeld in der Umgebung der Quelle.[JP]

Für die sphärische Nahfeldholografie werden Kugelkoordinaten verwendet. Das Kugelkoordinatensystem wird mithilfe eines zusätzlich angegebenen kartesischen Koordinatensystems durch die Wahl des Ursprungs, der Polachse und die Äquatorial Ebene definiert. So wird die Position eines Punktes auf der Kugeloberfläche durch  $(r, \theta, \varphi)$  repräsentiert (Abbildung 2.8).

Abbildung 2.8: Sphärische Koordinaten <sup>1</sup>

Der Koordinatenursprung ist bei Ambisonics der Aufnahmeort eines virtuellen Mikrofons. Die Richtung der eingehenden Schallwellen wird mit den sphärischen Harmonischen beschrieben.

$$Y_n^m(\Theta, \phi) = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}} P_n^{|m|}(\cos\Theta) e^{im\phi} \quad (2.1)$$

Hierbei bezieht sich der Index  $n = 0, 1, 2, \dots$  auf die Ordnung und  $m = -n \dots 0 \dots n$  auf den Grad der Kugelflächenfunktion.  $\bar{P}_{m,n}$  stellt die Legendre-Funktionen dar [Wil99]. Mit  $Y_n^m(\Theta, \phi)$  kann jedoch nur eine Approximation des Schallfeldes durch die Überlagerung von Kugelfunktionen entwickelt werden. Aus diesen Überlagerungen kann der Schalldruck und seine Richtung beschrieben werden. Für eine vollständige mathematische Herleitung sollte weiterführende Literatur zurate gezogen werden ([Bam94], [Fel74], [FZ19] und [Zm2]).

Eine unendliche Anzahl an Funktionen kann in der Praxis aufgrund von Rechenzeit und Speicherplatz nicht durch Signale repräsentiert werden. Daher muss die Anzahl  $n$  der sphärischen Harmonischen begrenzt werden. Unter dem Aspekt des Informationsverlustes und dem Koordinatenursprung als Betrachtungspunkt kann das Ambisonics Schallfeld  $\vec{B}_{\Phi, \Theta}$  vereinfacht beschrieben werden:

$$\vec{B}_{\Phi, \Theta} = \vec{Y}_{\Phi, \Theta} * p_{\Phi, \Theta} \quad (2.2)$$

<sup>1</sup><http://de.universaldenker.org>

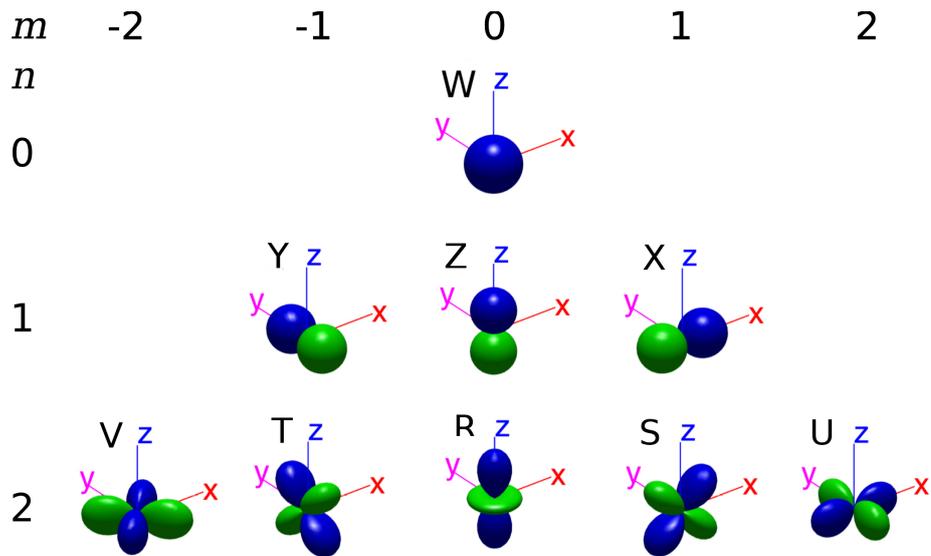


Abbildung 2.9: Ambisonics-Ordnung nach Furse Malham <sup>2</sup>

Ambisonics-Daten der 1. Ordnung werden im B-Format gespeichert. Eine Aufnahme besteht aus vier Kanälen. Dabei stellt W den omnidirektionalen Schalldruck  $p$  dar. X, Y, Z beschreiben die drei orthogonalen Komponenten der Schallschnelle  $v$  im Raum. Die Abhängigkeit im kartesischen KOS mit Azimut  $\varphi$  und Elevation  $\vartheta$  kann für die einzelnen Kanäle angegeben werden:

$$w = \frac{1}{\sqrt{2}} \tag{2.3}$$

$$x = \cos(\varphi)\cos(\vartheta) \tag{2.4}$$

$$y = \sin(\varphi)\cos(\vartheta) \tag{2.5}$$

$$z = \sin(\vartheta) \tag{2.6}$$

Bei höheren Ordnungen werden entsprechend weitere Kanäle und somit auch Mikrofone benötigt. Die Berechnungen werden dadurch zwar langsamer, aber es verbessert die Qualität der Schallfeldapproximation. Die Anzahl der Kanäle  $N$  kann mit

$$N = (M + 1)^2 \quad M \dots \text{Systemordnung} \tag{2.7}$$

berechnet werden [AW02].

<sup>2</sup><http://pcfarina.eng.unipr.it/Aurora/SH.png>

Für die Aufnahme werden Mikrofone mit entsprechender Richtcharakteristik benötigt, damit Schallquellen lokalisiert werden können. Für die 1. Ordnung werden Mikrofone mit der Richtcharakteristik eines Dipols (Achtercharakteristik, Bidirektional) benötigt. Bei den höheren Ordnungen sind folglich komplexere Richtcharakteristiken erforderlich. Abbildung 2.9 veranschaulicht die einzelnen Kanäle bis zur 2. Ordnung nach Gerzons Schema.

### 2.3.2 Soundfield-Mikrofon und Berechnung des B-Formats

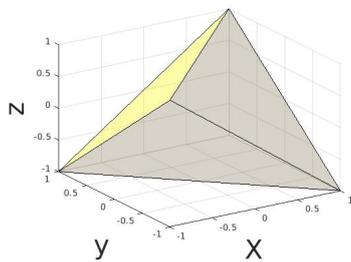


Abbildung 2.10: Tetraeder



Abbildung 2.11: Soundfield-Mikrofon SPS200 <sup>3</sup>

In dieser Arbeit wird ein Soundfield-Mikrofon erster Ordnung (Abbildung 2.11) verwendet. Mit 220g, einer Breite von 6 cm und einer Länge von 20 cm ist es ein recht kleines und leichtes Mikrofon. Beim SPS200 handelt es sich um ein Kondensatormikrofon, welches mit einer Phantomspannung von 48V gespeist werden muss. Die vier Mikrofon-Kapseln sind wie ein Tetraeder (Abbildung 2.10) angeordnet. Dadurch sind alle Kapseln gleichmäßig um einen Punkt verteilt. [Sou11]

Bei Aufnahmen mit dem Soundfield Mikrofon SPS200 entsteht das vier Kanal A-Format erster Ordnung. Jeder Kanal steht für eine der vier Nierenkapseln. Folgende Belegung der Kanäle 1 bis 4 trifft für die Mikrofonkapseln zu:

1. LF (Left Front)
2. RF (Right Front)
3. LB (Left Back)
4. RB (Right Back)

---

<sup>3</sup><http://www.soundfield.com/products/sps200>

Das A-Format wird anschließend noch in das von Gerzon entwickelte B-Format umgerechnet. Das klassische B-Format ist auch bekannt als Furse-Malham (FuMa)-Format. Zur Enkodierung in diese Darstellungsform spielt die Anordnung der Kanäle eine wichtige Rolle. Denn dabei müssen die einzelnen Kapselsignale umgerechnet werden:

$$W = \frac{LF + RF + LB + RB}{\sqrt{2}} \quad (2.8)$$

$$X = (LF + LB) - (RF + RB) \quad (2.9)$$

$$Y = (LF + RF) - (LB + RB) \quad (2.10)$$

$$Z = (LF + RB) - (RF + LB) \quad (2.11)$$

### 2.3.3 Speicherformate

Zum Speichern und Austauschen von Ambisonics-Daten existieren zwei geläufige Standardformate. Das FuMa-Format ist bereits lang etablierter Standard. Es ist bis zur 3. Ordnung Gerzons festgelegt und die einzelnen Kanäle sind in alphabetischer Anordnung. Beispielsweise ist das klassische B-Format (W, X, Y, Z) notiert. Die omnidirektionale Komponente W wird dabei  $\frac{1}{\sqrt{2}}$  gewichtet, während die anderen drei Signale den Wert von  $|1|$  nicht überschreiten. Diese Normalisierung ist auch bekannt als maxN. Der FuMa-Standard wird aufgrund des Alters in vielen Plugins verwendet. Jedoch besitzt der .amb Container eine Speicherlimitierung von 4GB und unterstützt nur bis zu 16 Kanäle. [Mal99]

Ein moderneres Format, welches gerade in der VR-Welt für Panoramavideos Anklang findet, ist AmbiX. Soziale Plattformen wie YouTube und Facebook 360° [Fac] nutzen dieses neuere Format. AmbiX unterscheidet sich zu FuMa vor allem in der Kanalreihenfolge. Diese geht nach der ACN Nummerierung (Gleichung 2.12), in der die einzelnen Kanäle zuerst der Ordnung  $n$ , nachfolgend dem Grad  $m$  sortiert werden. Beispielsweise das B-Format in AmbiX (W, Y, Z, X).

$$ACN = n^2 + n + m \quad (2.12)$$

Bei AmbiX beträgt der Gewichtungsfaktor des Schalldrucks in W  $f = \sqrt{\frac{1}{4\pi}}$ . Außerdem wird statt maxN in der sphärisch Harmonischen die Schmidt semi-normalisation (SN3D)

verwendet (Gleichung 2.13). Gespeichert werden die Daten in dem .caf Container. Von beiden Standards kann in das jeweilig andere konvertiert werden. [NZDS11]

$$N_n^m = \sqrt{(2 - \delta_{nm}) \frac{(n - |m|)!}{(n + |m|)!}} \quad (2.13)$$

In dieser Arbeit wird sich für die FuMa Anordnung entschieden, es ist nachvollziehbarer als AmbiX. Aufgrund des Alters ist die Literatur zum FuMa-Standard umfangreicher. Die Aufnahmen werden im .wav Container gespeichert und im Anschluss ins B-Format gewandelt.

### 2.3.4 Koordinatensystem des Soundfield-Mikrofons

Durch die Konvertierung vom Ambisonics A-Format ins B-Format können Geräusche, die das Soundfield-Mikrofon aufnimmt, den drei Achsen eines Koordinatensystems zugeordnet werden. Im B-Format wird die Richtung dargestellt, aus der die Audiogeräusche zu hören sind. Die vom Soundfield-Mikrofon verwendeten Achsen sind in Abbildung 2.12 dargestellt.



Abbildung 2.12: Soundfield-Mikrofon SPS200 mit den Achsen nach der Umrechnung ins B-Format

stellt. Bei frontaler Sicht wird die Höhe durch die Z-Achse definiert. Links und Rechts hingegen durch die Y-Achse. Vorne und hinten vom SPS200 wird mit dem X-Anteil beschrieben. Dabei werden die Signale wie folgt interpretiert: kommen die Geräusche von rechts, so hat das Y-Signal eine hohe Amplitude. Wenn sich die Tonquelle auf der linken Seite befindet, schlägt Y ebenfalls aus, jedoch invertiert zu W. Mit X und Z verhält es sich, der Richtung entsprechend, identisch.

### 3 Schallquellenlokalisierung und Koordinatentransformation

In den Grundlagen wurden die WFS-Anlage und das Tracking-System mit den jeweiligen Koordinatensystemen beschrieben. Dieses Kapitel widmet sich zuerst der Berechnung der Schallquellenposition im Tracking-System mithilfe des Soundfield-Mikrofons. Daraufhin wird die homogene Transformation zwischen dem WFS- und dem Tracking-KOS beschrieben.

Zur Berechnung der Koordinatentransformation müssen Punkte in beiden Systemen bekannt sein. Da die WFS zurzeit nur eine Ebene abdeckt, behandelt diese Arbeit nur zwei Dimensionen. Außerdem verfügt die WFS nicht über Sensoren, die eine Interaktion mit der Außenwelt ermöglichen würden. Sie kann lediglich Schallquellen ausgeben. Demnach werden die Schallquellen im Tracking-System lokalisiert und dargestellt.

Zur Bestimmung der Schallquellenposition wird der mathematische Ansatz der Triangulation verfolgt. Mittels dieser Methode kann die Position der Schallquelle in der Ebene mit mindestens zwei geeigneten Messungen ermittelt werden. Hierbei bildet das Ambisonics-Soundfield-Mikrofon die Schnittstelle.

Das Ambisonics-Mikrofon dient dazu, die Richtung vom Target zur Audioquelle zu ermitteln. Dafür sind Mikrofon und Target mit einem Stativ fixiert. Informationen zu dem Stativ erfolgen zu einem späteren Zeitpunkt. Durch die Fixierung können Targetkoordinaten zusammen mit den Ambisonics-Daten die Schallquellenrichtung relativ zum Stativ bestimmen.

### 3.1 Richtungsbestimmung mit Ambisonics

Die Lokalisierung von Schall wurde von der Robotik in den vergangenen Jahren erheblich vorangetrieben, indem traditionelle Techniken wie DOA-Schätzung (Single Direction-of-Arrival), lernbasierte Ansätze (wie neuronales Netzwerk und Machine Learning), strahlformbasierte Ansätze (Beamforming) und Subraummethode entwickelt und verfeinert wurden.

Die Richtungsbestimmung in Ambisonics von Schallquellen basiert auf der Schallintensität. In der Arbeit “Localization of the Sound Source with the Use of the First-order Ambisonic Microphone” [WMW13] werden drei Möglichkeiten zur Bestimmung der sphärischen Winkelkoordinaten von Schallquellen verglichen. Hierbei werden unterschiedliche Ansätze zur Lokalisierung von Schallquellen untersucht. Der Schalldruck  $p$  im B-Format ist die W-Komponente. X, Y, Z bilden die Schallschnelle  $v$ .

$$I = p \cdot v \quad (3.1)$$

Die drei Möglichkeiten zur Lokalisierung der Schallquelle mit dem Soundfield-Mikrofon wurden bereits mit der WFS-Anlage gezeigt [Feh18]. Daher wird zur Schallintensitätsberechnung die Summe aus den Produkten der relevanten Signale ([WMW13], Formel 7) verwendet.

$$I_x(n) = \sum_n X(n) \cdot W(n) \quad (3.2)$$

Über das komplexe Argument des Intensitätsvektors  $I = [I_x \ I_y]$  kann der Winkel  $\gamma$  berechnet werden. Da die WFS derzeit keine Höhenkomponente ansteuern kann, kann die Elevation des Mikrofons vernachlässigt werden. Infolgedessen wird nur der Winkel  $\gamma$  verwendet, um die Lage der Schallquelle relativ zur x-Achse anzugeben:

$$\gamma = \arg(I_x + iI_y) \quad (3.3)$$

Mit einer einzelnen Summe aus den Messwerten ist es nicht möglich, Messungenauigkeiten bzw. die daraus ermittelten Ergebnisse für die Richtung der Schallquelle  $\gamma$  abzuschätzen. Um die statistische Verteilung der Ergebnisse der Messungen abschätzen zu können, müssen daher mehrere gleichartige Versuche durchgeführt oder die Messdaten einer Messung in Messblöcke aufgeteilt werden. Vorausgesetzt, dass die Messungen innerhalb jedes Blockes lang genug sind, um die erforderliche Frequenzauflösung abzubilden.

### 3.2 Triangulation

Zusätzlich zum Winkel werden für eine Positionsbestimmung der Audioquelle durch Triangulation weitere Koordinaten benötigt. Der Begriff der Triangulation aus der Messtechnik bezeichnet die Bestimmung von Entfernungen mithilfe von optischen Messgeräten [HM06]. In dieser Arbeit wird Triangulation angewendet, um die Position der Schallquelle  $\vec{S}$  im Tracking-KOS  $\vec{P}$  zu ermitteln. Ein Ambisonics-Mikrofon unterstützt die optische Komponente des Tracking-Systems, indem es die akustische Richtung  $\vec{e}$  der Audioquelle misst.

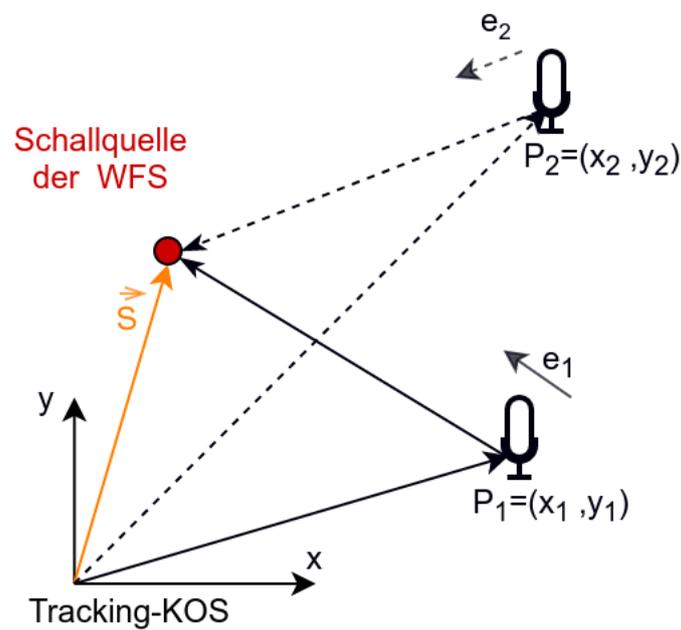


Abbildung 3.1: Messungen für die akustische Triangulation

Zur Bestimmung der Schallquelle (Abbildung 3.1) können trigonometrische Funktionen angewendet werden. Hierbei sind die Punkte  $\vec{P}_1$ ,  $\vec{P}_2$  des Tracking-Systems bekannt. Zusammen mit den entsprechenden Ambisonics-Richtungen  $\vec{e}_1$ ,  $\vec{e}_2$  bilden die Punkte eine Gerade ( $\vec{S}_n = \vec{P}_n + a_n \vec{e}_n$ ).

$$\begin{aligned} S_1 &= \begin{cases} x_1 + a \cdot \cos(\gamma_1) \\ y_1 + a \cdot \sin(\gamma_1) \end{cases} \\ S_2 &= \begin{cases} x_2 + b \cdot \cos(\gamma_2) \\ y_2 + b \cdot \sin(\gamma_2) \end{cases} \end{aligned} \quad (3.4)$$

Durch Gleichsetzen von  $\vec{S}_1 = \vec{S}_2$  können die zu Position der Schallquelle  $\vec{S}$  erforderlichen Abstände  $\vec{a}_1$  und  $\vec{a}_2$  berechnet werden. Bei dem entstandenen Gleichungssystem können die unbekanntenen Größen auf die linke Seite und die bekannten auf die rechte Seite umgestellt werden.

$$\begin{aligned} S_1 &= S_2 \\ x_1 + a \cdot \cos(\gamma_1) &= x_2 + b \cdot \cos(\gamma_2) \\ y_1 + a \cdot \sin(\gamma_1) &= y_2 + b \cdot \sin(\gamma_2) \\ a \cdot \cos(\gamma_1) - b \cdot \cos(\gamma_2) &= x_2 - x_1 \\ a \cdot \sin(\gamma_1) - b \cdot \sin(\gamma_2) &= y_2 - y_1 \end{aligned} \quad (3.5)$$

Daraufhin kann das Gleichungssystem in Matrixnotation aufgeschrieben werden.

$$A \cdot \vec{x} = \vec{b} \quad (3.6)$$

Hierbei enthält der Vektor  $\vec{x}$  die unbekanntenen Abstände und  $\vec{b}$  die bekannten Positionen.

$$\begin{aligned} A &= \begin{bmatrix} \cos(\gamma_1) & -\cos(\gamma_2) \\ \sin(\gamma_1) & -\sin(\gamma_2) \end{bmatrix} \\ \vec{b} &= \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \end{bmatrix} \\ \vec{x} &= \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \end{aligned} \quad (3.7)$$

Zunächst ist zu überprüfen, ob sich die Geraden schneiden und ob eine Lösung für das Gleichungssystem existiert. Dies kann mithilfe der Determinante  $A$  analytisch berechnet werden. Entspricht  $D = \det(A)$  dem Wert Null, kann es keinen Schnittpunkt geben und die zu invertierende Matrix ist singulär. Das ist der Fall bei zwei parallelen Geraden.

$$D = \det(A) = \cos(\gamma_2)\sin(\gamma_1) - \cos(\gamma_1)\sin(\gamma_2) \quad (3.8)$$

Ein Additionstheorem der Trigonometrie kann angewendet werden, um die Schreibweise zusammenzufassen.

$$D = \sin(\gamma_1 - \gamma_2) \quad (3.9)$$

Demnach kann bei zwei parallelen Geraden die inverse Matrix nicht berechnet werden.

$$\text{inv}(A) = \frac{1}{D} \cdot \begin{bmatrix} -\sin(\gamma_2) & \cos(\gamma_2) \\ -\sin(\gamma_1) & \cos(\gamma_1) \end{bmatrix} \quad (3.10)$$

Bei zwei sich schneidenden Geraden ( $D \neq 0$ ) gilt es die Lösung des Gleichungssystems zu bestimmen.

$$\begin{Bmatrix} a \\ b \end{Bmatrix} = \frac{1}{\cos(\gamma_2)\sin(\gamma_1) - \cos(\gamma_1)\sin(\gamma_2)} \begin{bmatrix} -\sin(\gamma_2) & \cos(\gamma_2) \\ -\sin(\gamma_1) & \cos(\gamma_1) \end{bmatrix} \begin{Bmatrix} x_2 - x_1 \\ y_2 - y_1 \end{Bmatrix} \quad (3.11)$$

So kommen folgende Formeln für die beiden Abstände  $a_1$  und  $a_2$  zustande:

$$\begin{aligned} a &= \frac{(y_2 - y_1)\cos(\gamma_2) - (x_2 - x_1)\sin(\gamma_2)}{\sin(\gamma_1 - \gamma_2)} \\ b &= \frac{(y_2 - y_1)\cos(\gamma_1) - (x_2 - x_1)\sin(\gamma_1)}{\sin(\gamma_1 - \gamma_2)} \end{aligned} \quad (3.12)$$

Die beiden Variablen müssen zur Berechnung der Schallquellenposition in Gleichung 3.4 eingesetzt werden. Sollte eine Lösung existieren, sind die beiden Punkte identisch.

Selbst wenn sich mathematisch die Geraden  $\vec{S}_1$  und  $\vec{S}_2$  schneiden, bleibt die Frage bestehen, ob sich der Schnittpunkt in einem relevanten Bereich befindet. Bei Messwerten ist es möglich, dass der Schnittpunkt an einer irrelevanten Position liegt und somit unbrauchbar ist. Entstehen können diese irrelevanten Punkte, indem der Schnittpunkt durch die Richtung deutlich entfernter der erzeugten Audioquelle berechnet wurde. Diese Schnittpunkte können durch eine Kontrolle der Werte  $a$  und  $b$  ausgeschlossen werden. Ist einer der beiden Werte negativ ist der Punkt außerhalb des zu betrachtenden Bereiches.

### 3.3 Homogene Transformation

Kartesische KOS können sich in Ursprung, Rotation, Skalierung, Händigkeit und Anzahl der Dimensionen unterscheiden. Mit einer Transformation können Targets im WFS-KOS und Schallquellen im Tracking-KOS beschrieben werden. In dieser Arbeit dient die homogene Transformation zur Ermittlung der Translation sowie der Orientierung zwischen Tracking- und WFS-System.

Es gibt unterschiedliche Darstellungen für die Rotation zwischen zwei KOS. Unter anderem Quaternionen. Sie erweitern den reellen Zahlenbereich mit drei komplexen Zahlen. Dabei tritt das Kommutativ-Gesetz außer Kraft und die Reihenfolge der Elemente spielt eine Rolle. Quaternionen sind im Raum zur Beschreibung von Orientierungen geläufig. Bei der Verwendung des Zahlenbereichs lassen sich Rotationen elegant formulieren. [Hor87]

Im Folgenden wird jedoch der Ansatz mittels Rotationsmatrix verfolgt. Dieser ist vorteilhaft, da ihre Komplexität im Vergleich zu Quaternionen geringer ist. Ebenfalls lässt sich eine Rotationsmatrix sowie eine Translation zu einer Transformationsmatrix zusammenfassen. Besagtes ist insofern sinnvoll, da das Tracking-System Rotationsmatrizen übermittelt. Um die unterschiedlichen Punkte in einem projektiven Raum vereinheitlicht darzustellen, werden diese als homogene Koordinaten definiert und mittels Matrizen beschrieben.

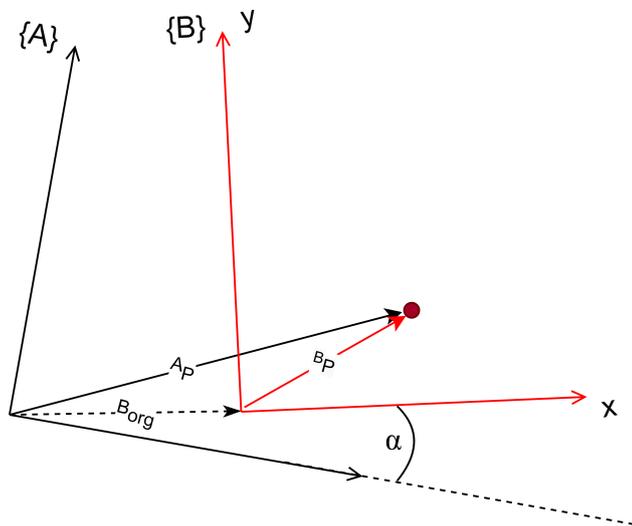


Abbildung 3.2: Punkt P des Koordinatensystems B dargestellt in A

Veranschaulicht wird die Herangehensweise der homogenen Transformation in Abbildung 3.2. Hier existieren zwei kartesische KOS,  $\{A\}$  und  $\{B\}$ . Beide Ebenen werden von ihrem eigenen Ursprung mit X- und Y- Achsen aufgespannt und haben eine Orientierung  $\alpha$  zueinander. Die Translation von  $\{A\}$  nach  $\{B\}$  wird mit dem Vektor  $\vec{B}_{org}$  dargestellt. Der Punkt  $P$  ist in  $\{B\}$  bekannt und kann wie folgt beschrieben werden:

$${}^B P = \begin{Bmatrix} {}^B P_x \\ {}^B P_y \end{Bmatrix} \quad (3.13)$$

Die Transformation dient dazu Vektoren aus dem KOS  $\{B\}$  in dem anderen KOS  $\{A\}$  darzustellen. Dabei wird für  ${}^A P$  die Transformationsmatrix mit dem Vektor  ${}^B P$  multipliziert.

$${}^A P = {}^A_B T \cdot {}^B P \quad (3.14)$$

Besteht  $\{B\}$  nur aus einer Drehung  $\alpha$  von  $\{A\}$  kann dies in Spaltenvektoren einer Rotationsmatrix ausgedrückt werden. Da in dieser Arbeit das WFS-KOS aus der XY-Ebene besteht, findet die Rotation um die Z-Achse statt. Daher ergibt sich folgende 3x3 Rotationsmatrix:

$${}^A_B R(\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.15)$$

Eine komplette Transformation setzt sich aus der Translation und Rotation zusammen. Im kartesischen Raum kann  ${}^A_B T$  mit einer 4x4 Matrix beschrieben werden.

$${}^A_B T = \begin{bmatrix} {}^A_B R(\alpha) & B_{org} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 & B_{org_x} \\ \sin(\alpha) & \cos(\alpha) & 0 & B_{org_y} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.16)$$

Dann kann mithilfe von homogenen Koordinaten der Punkt  ${}^A P$  berechnet werden. Hierbei findet die Berechnung nur auf der Ebene statt, weshalb die dritte Dimension vernachlässigt werden kann. Zur Vermeidung von unnötigen Multiplikationen mit Nullen kann die Gleichung vereinfacht geschrieben werden:

$$\begin{Bmatrix} {}^A P_x \\ {}^A P_y \end{Bmatrix} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & B_{org_x} \\ \sin(\alpha) & \cos(\alpha) & B_{org_y} \end{bmatrix} \begin{Bmatrix} {}^B P_x \\ {}^B P_y \end{Bmatrix} \quad (3.17)$$

Die Werte für  ${}^A\vec{P}$  und  ${}^B\vec{P}$  in der Gleichung sind bekannt. Im Idealfall ist das Gleichungssystem mit mindestens zwei Messpunkten lösbar. Gelöst bedeutet in dem Sinne, die Rotation sowie die Translation zu bestimmen.

$$err_i = {}^A\vec{P}_i - {}^A_B R(\alpha) \cdot {}^B\vec{P}_i - \vec{B}_{org} \quad (3.18)$$

Die Funktion enthält die gesuchten Variablen  $\alpha$  sowie  $x, y$  der Ursprungstranslation. Ein analytischer Ansatz zur Lösungsbestimmung bei Messwerten birgt Probleme. Durch das entsprechende Aufnahmegerät der Messung entstehen Abweichungen ( $err$ ). Somit kann bei Aufstellung des Gleichungssystems mit den ungenauen Werten keine exakte Lösung gefunden werden. Deshalb wird numerisch nach einer Annäherung gesucht, bei der die Unbekannten optimal sind.

Es wird als Kostenfunktional aus der Lösung ein Skalar  $err' \cdot err$  als Fehlerquadrat gebildet. Mit dem Levenberg-Marquardt-Algorithmus [Lev44] wird iterativ das Minimum in dem Fehlerquadrat gesucht.

Der Fehler kann als ein Gebirge mit mehreren Tälern vorgestellt werden. Gefragt ist hierbei das tiefste Tal. Demnach gilt es eine nicht lineare multivariable Funktion zu lösen. Dabei wird immer die größte Steigung abwärts (Gefälle) der angrenzenden Quadrate ermittelt. So werden auf dem schnellsten Weg Täler, also die Variablenwerte mit dem minimalen Fehler, gefunden. Eine Einschränkung vom Suchraum des minimalen Fehlers (conditional optimization) beschleunigt die Rechenzeit erheblich.

## 4 Aufbau und Durchführung der Messungen

Für die Berechnungen aus dem vorherigen Kapitel müssen Daten via Messung gewonnen werden. Diese sollen ausgewertet werden, um Koordinatentransformation zwischen der WFS-Anlage und dem Tracking-System zu bestimmen.

Ein Datenset bezeichnet folgend die Gruppierung von Target-Koordinaten und den dazugehörigen Ambisonics-Audiospuren. Aus diesem Datenset entsteht die Gerade mit Ursprung (Target) und Richtung (Ambisonics). Ein Messdurchlauf besteht demnach aus mindestens drei relevanten Datensets. Relevant in dem Sinne, dass sich die Datensets ungefähr an einem bestimmbareren Punkt schneiden.

In diesem Kapitel wird zuerst auf die einzelnen Komponenten der Messungen eingegangen. Mit diesem Verständnis zur Grundlage kann schließlich die Durchführung eines Messdurchlaufs aufgeführt werden.

### 4.1 Hardware-Komponenten

Um die einzelnen Bestandteile eines Messdurchlaufs und deren Zusammenhang zu erläutern, werden diese grafisch in Abbildung 4.1 dargestellt. Die Hardware unterteilt sich in Stativ, welches Target und Soundfield-Mikrofon verbindet, dem Audio-Interface, Tracking-PC sowie der WFS-Anlage.

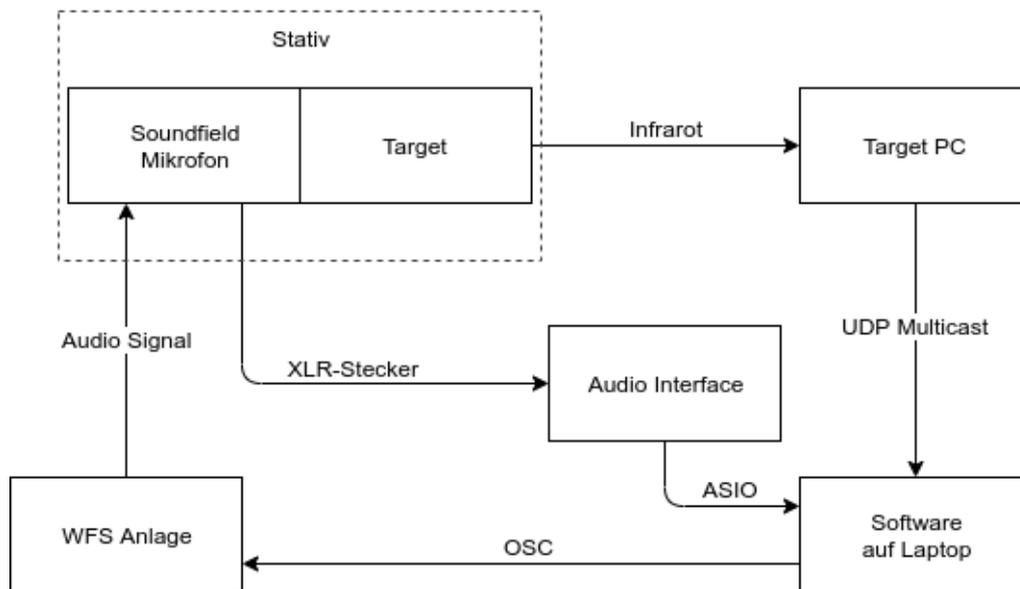


Abbildung 4.1: Komponenten-Diagramm der Hardware

### Stativ

Für diese Arbeit wurde ein Stativ entwickelt. Es fixiert das Soundfield-Mikrofon SPS200 mit dem Target TT5 des A.R.T.-Systems in einer bestimmten Position (Abbildung 4.2).

Der Prototyp besteht aus einem Meter langen Rohr und einem Aluminiumkopf. Befestigt wird die Stange an einem Dreibein, wodurch es einem Mikrofonstativ ähnelt und eine maximale Höhe von 1,45 Meter ohne das Target hat.

Eine M5 Bohrung mit Feingewinde befindet sich auf der Unterseite des verwendeten Targets. Als Gegenstück wurde dafür nach DIN 13 ein metrisches M5 ISO-Gewinde an dem Aluminiumkopf befestigt. Das Target TT5 definiert sich mit dem Ursprung in einem vom Gewinde abstehenden Marker. Dieser ist über den vier Mikrofonkapseln des SPS200 angeordnet worden. Dadurch kann die Distanz zwischen Target und Mikrofon für anstehende Berechnungen vernachlässigt werden.

Das Mikrofon wird durch das in den Aluminiumkopf gefräste Loch geschoben und mit einem kleinen Plastikbolzen festgeklemmt. Mit dieser Anordnung ist der Abstand und die Orientierung von Mikrofon und Target fest definiert.

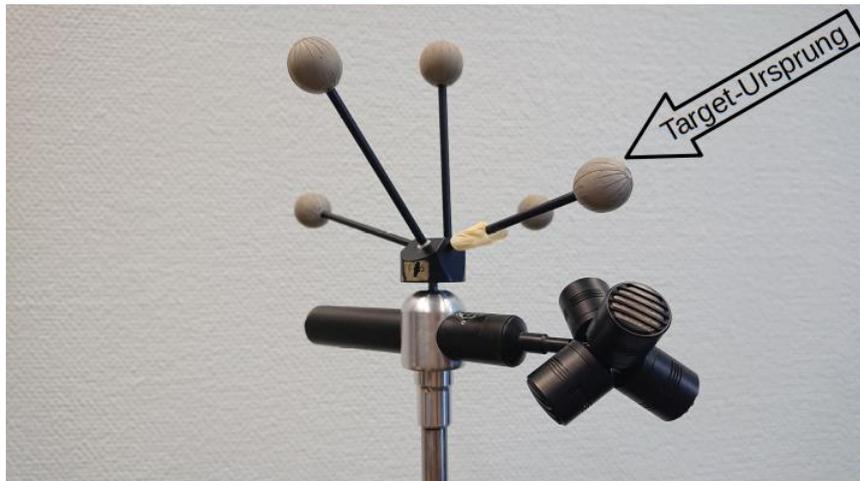


Abbildung 4.2: Halterung mit Target TT5 und SPS200 Soundfield-Mikrofon

### Soundfield-Mikrofon SPS200

Mit dem Mikrofon werden die eingehenden Richtungen der Schallwellen gemessen. Verwendet wird hierbei das Soundfield-Mikrofon SPS200 (Abbildung 4.2). Die vier synchronen Signale werden von den einzelnen Kapseln im A-Format aufgenommen. Die Dekodierung in das B-Format findet nicht in der vom Hersteller gestellten Software, sondern in Matlab statt.

### Audio-Interface

Als externes Audio-Interface wurde ein Focusrite 18i8 1st Generation verwendet. Dieses verfügt über 18 Ein- sowie 8 Ausgänge. Die Verbindung zwischen dem Interface und einem PC findet über USB 2.0 statt. Beim Anschließen der vier XLR-Steckern des Soundfield-Mikrofons muss die Reihenfolge beachtet werden. Für jeden XLR-Eingang besitzt das Interface einen eigenen, stufenlosen Audioverstärkungsregler.

Die Übertragung zwischen dem Windows 7 Laptop und dem Audio-Interface findet mittels der Bibliothek portAudio [Por] statt. Es verwendet dabei das Audio Stream Input/Output (ASIO)-Protokoll, um die 4 Kanäle auszulesen. Ein Protokoll welches primär das Betriebssystem Windows unterstützt.

### **Tracking-PC**

Der Tracking-PC ist ein separater Computer auf dem die Software der Firma A.R.T. läuft. Mit den Informationen der Infrarotkameras wird in der Software die Position und die Rotation der erfassten Targets berechnet. Diese Werte werden von dem PC mit Multicast in das Netzwerk verteilt.

### **WFS-Anlage**

Die WFS dient dazu Audioquellen zu rendern, welche dann vom Mikrofon aufgenommen und lokalisiert werden. Hier sei nochmal erwähnt, dass die Anlage OSC-Nachrichten über das Netzwerk empfangen kann. OSC dient hier dazu, Audioquellen zu steuern. Die Transformation zwischen den beiden KOS, der WFS-Anlage und des Tracking-Systems, soll mathematisch kontrolliert und ggf. angepasst werden.

## **4.2 Datenaufnahme und -verarbeitung**

Mikrofon- sowie Target-Daten müssen zusammen aufgenommen werden. Unter diesem Aspekt wird von der Verwendung von Musikproduktionssoftware für die Aufnahme abgesehen. Stattdessen wird für den Laborlaptop eine Software in C++ entworfen, welche die benötigten Daten erfasst und speichert. Auswertung und Berechnung der Daten erfolgen in einem separaten Programm.

### **4.2.1 Erfassen und Speichern der Daten**

Die Berechnungen finden in Matlab statt, da diese damit elegant zu lösen sind. Ebenfalls ist die Ansteuerung vom Netzwerk möglich. Allerdings stellt sich hiermit die Ansteuerung des Audio-Interfaces und somit die Audioaufnahmen als schwierig heraus. Dementsprechend erfolgt die zusammenhängende Aufnahme von Mikrofon- und Target-Daten in einem separaten Programm.

Es wurde die Programmiersprache C++ gewählt. Hierfür existiert die Audio-Bibliothek portAudio mit dem das Audio-Interface Focusrite 18i8 angesteuert werden kann. Des

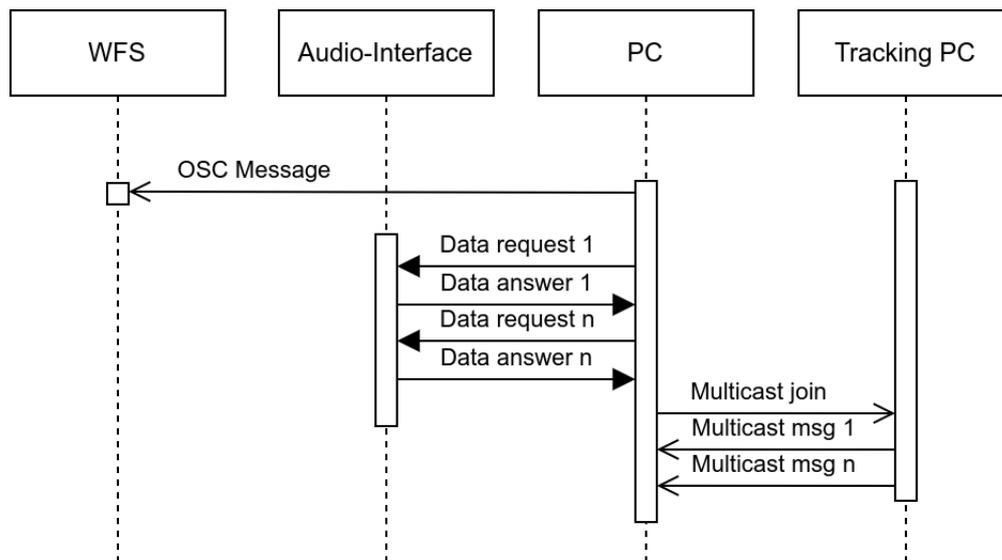


Abbildung 4.3: Sequenzdiagramm der Datenerfassung für ein Datenset

Weiteren wurde ab C++17 die Regular Expressions (Regex) mit als Funktion aufgenommen. Zu einem späteren Zeitpunkt stellte sich hingegen die Nutzung von Regex zur Datenfilterung in Matlab als vorteilhafter heraus.

Der Messablauf eines Datensets wird in Abbildung 4.3 visualisiert. Gemessen werden Punktquellen, welche von der WFS-Anlage erzeugt werden. Der Wiedergabeort wird mittels OSC-Befehl über das Netzwerk gesendet. Sobald dies geschehen ist, wird die Audioaufnahme vom Mikrofon gestartet. Dafür wird auf das Audio-Interface mit dem ASIO-Protokoll zugegriffen. Die einzelnen Kanäle des Focusrite werden für eine gewisse Zeit aufgenommen. Im Anschluss wird von der Multicast-Gruppe des Tracking-PCs ein Mitschnitt der Tracking-Daten gespeichert.

Einstellungen der Aufnahme, wie bspw. Netzwerkinformationen oder die Datenset-Anzahl eines Messdurchlaufs, sind in einer simplen Config-Datei vorzunehmen. Die Config-Datei ist im Anhang (A.1) beschrieben.

Ein Messdurchlauf besteht aus mindestens drei Aufnahmen von Datensets mit der gleichen Punktquelle. Nach der Messung eines Datensets wird das Stativ mit dem Mikrofon und Target an eine neue Position bewegt und bestätigt, um eine neue Aufnahme zu starten.

### Ansteuerung der WFS

Die Steuerung der Punktquellen erfolgt in der WFS-Anlage mit OSC-Nachrichten. Die zwei Befehle *activate* und *position* sind vonnöten, um die Punktquellen einmalig auf eine entsprechende Audioquelle zu initialisieren und dann auf eine Position im WFS-KOS zu bewegen:

```
/WONDER/source/activate,i  
/WONDER/source/position,iff
```

OSC-Nachrichten sind übersichtlich aufgebaut. Es beginnt mit einer Art URL-Adresse und dem Befehl. Nach einem Komma werden Typ-Tags der darauffolgenden Argumente aufgelistet. Das *i* steht hierbei für den Identifikator der Audioquelle als Integer. Um die Position zu setzen, muss eine X- sowie Y-Koordinate in Metern angegeben werden. Für die beiden Floats stehen die zwei Typ-Tags *ff*. Da in dieser Arbeit die Anzahl der verwendeten Befehle sehr gering ist, wurde sich gegen die Verwendung einer Bibliothek entschieden.

### Audioaufnahmen

Für Audioaufnahmen wird die Bibliothek portAudio [Por] verwendet. Sie dient dazu, das verwendete Audio-Interface auszuwählen und mit dem ASIO-Protokoll anzusteuern. Vor Beginn der Aufnahme wird mit der Kanalanzahl und dem portAudio-Device ein Stream zu dem Interface geöffnet. Mittels dieses Streams werden die 4 Kanäle empfangen und über einen Zeitraum von 10 Sekunden aufgenommen. Gespeichert wird das Ganze in einer WAVE-Datei. Die Daten liegen als A-Format vor. Der Dateiname setzt sich aus der im Config-File angegebenen Endung und der entsprechenden Iteration zusammen.

### Tracking-Daten

Im Anschluss an die Mikrofonaufnahme müssen Tracking-Daten gewonnen werden. Diese Daten werden von dem Tracking-PC an eine Multicast-Gruppe im Netzwerk gesendet. Demnach muss ein User Datagram Protocol (UDP)-Socket auf das lokale Netzwerk-Interface gebunden werden, um sich dann an die Multicast-Gruppe anzumelden. Darüber können die Pakete des Tracking-Systems empfangen werden.

Ein Paket besteht aus einer Vielzahl an Informationen als ASCII-String. Unter anderem alle erkannten Targets. Ein Target wird dabei in drei hintereinander liegenden Klammern beschrieben:

$$[id, qu][x, y, z, \mu, \theta, \phi][Rotationmatrix]$$

Ein Target könnte demnach wie folgt aussehen:

```
[14 1.000] [-2.181 727.632 1455.677 1.6850 1.7945 -81.9603]
[0.139791 -0.989614 -0.033494 0.989686 0.140711 -0.026882 0.031316
-0.029390 0.999077]
```

Die ersten Klammern beherbergen die Targe-ID und die Qualität. Der zweite Teil der Informationen besteht aus den kartesischen Koordinaten  $(x, y, z)$  und den Eulerwinkeln  $\mu, \theta, \phi$  in Floats. Die letzten Klammern enthalten die Rotationsmatrix. Die kartesischen Werte sind in Millimetern angegeben, die Winkel hingegen im Bogenmaß. Die entsprechenden Informationen des Target TT5 ( $id = 14$ ) werden mithilfe von RegEx gefiltert.

### 4.2.2 Verarbeitung der Daten

Eine Herausforderung bestand darin, die optimale Messdatenauswertung zu entwerfen. Im Laufe der Arbeit wurde dieser Aspekt stetig ausgebaut, sodass Abbildung 4.4 dieses in Matlab realisierte Verfahren darstellt.

In der ersten Phase werden die gemessenen Daten eingelesen. Darauf folgt die Kontrolle, welches Rendering-Verfahren der Punktquellen verwendet wurde. Falls mit dem MoWeC eine Audioquelle auf ein Target gelegt wurde, wird dieses ebenfalls aus den Tracking-Daten gefiltert und geplottet. Ansonsten erfolgt dieser Referenzpunkt durch manuelle Eingabe. Der Punkt soll visuell helfen die Relevanz der Messung abzuschätzen.

Danach werden die eigentlichen Target-Daten bearbeitet. Hier werden die benötigten Daten  $(x, y$  sowie  $\phi)$  aus dem Tracking-Daten gefiltert. Die Höhenkomponente  $Z$  wird in den Berechnungen vernachlässigt. Der Winkel  $\phi$  hingegen dient zur Rotationskorrektur. Damit kann sichergestellt werden, dass der menschliche Faktor beim Positionieren des Stativs keine entscheidende Rolle spielt. Dafür wird der Unterschied von  $\phi$  bezüglich eines Referenzwertes bei der Ambisonics-Winkelberechnung mit einbezogen.

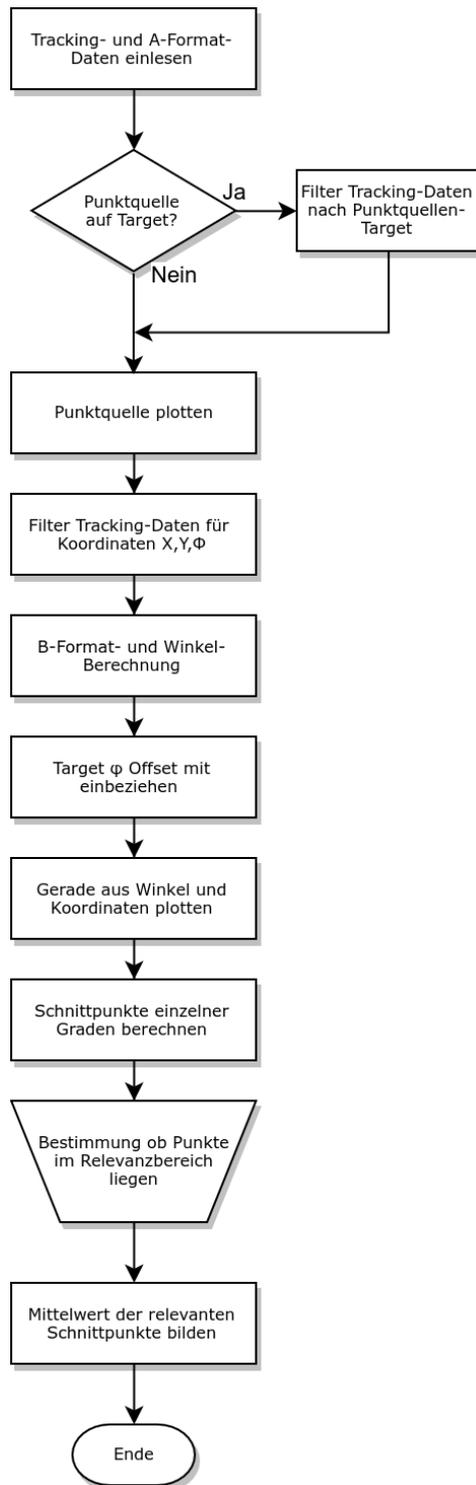


Abbildung 4.4: Ablauf der Datenverarbeitung von einem Messdurchlauf

Um den Ambisonics-Winkel zu berechnen, müssen zuerst die Daten vom A- ins B-Format konvertiert werden. Erst in dieser Form kann, sowohl der Winkel, als auch die Standardabweichung bestimmt werden. Im Anschluss wird noch der Rotationsfaktor des Stativs mit in den Winkel ein berechnet.

Dieser Winkel bildet zusammen mit den Target-Koordinaten eine Gerade. Demnach entstehen in einem Messdurchlauf mehrere Geraden. Von den einzelnen Geraden können Schnittpunkte bestimmt werden (Abschnitt 3.2). Zu beachten ist, dass die Schnittpunkte aussagekräftig sind und in einem relevanten Bereich liegen. Um dies einschätzen zu können, werden die Geraden visuell dargestellt.

Aus den relevanten Schnittpunkten eines Messdurchlaufs wird der Mittelwert gebildet. Dieser agiert als Bezugspunkt für die Bestimmung der Koordinatentransformation, bei welcher die Transformationsmatrix zwischen WFS und Tracking ermittelt werden soll.

### 4.3 Durchführung der Messung

Dieser Abschnitt fasst noch einmal den gesamten Ablauf in der Praxis zusammen. Unter anderem, auf welche Besonderheiten bei den Messungen geachtet oder bestimmte Vorbereitungen getroffen werden müssen. Zweck ist es, Daten eines Messdurchlaufs auswerten und bewerten zu können.

Für die Messungen sind die einzelnen Komponenten wie in Abbildung 4.1 angeschlossen. Bestmögliche Laborbedingungen sind geschaffen, indem Vorhänge und Tür geschlossen sind. Von in Betrieb genommener WFS-Anlage, sowie Tracking-Software wird ebenfalls ausgegangen.

Bei der Verwendung des SPS200 ist es zusätzlich zu der Phantomspannung relevant, die Audioverstärkung (Gain) aller Kanäle am Audio-Interface identisch einzustellen. Die vier Kanäle bilden zusammen das A-Format. Bei unterschiedlicher Verstärkung einzelner Kanäle kann es zur fehlerhaften Dekodierung ins B-Format kommen. In diesem Aufbau wurden die Regler auf ihr Maximum gestellt. Grund hierfür ist die fehlende Expertise zur Abmischung vom Mikrofon.

Sobald diese Vorbereitungen getroffen sind, geht es an den eigentlichen Messdurchlauf mithilfe des Aufnahmeprogramms. Das Stativ befindet sich innerhalb des Raumes, während die WFS-Anlage eine Punktquelle erzeugt und die entsprechende Audiodatei darauf ausgibt. Eine Datenset-Aufnahme dauert in etwa 15 Sekunden. Die Dauer eines Messdurchlaufs wird deutlich von der Platzierung des Stativs an neue Positionen und deren Bestätigung in der Software beeinflusst. Bei der Positionierung sollten die Messpunkte nicht auf einer Geraden zur Punktquelle liegen. Die Geraden könnten parallel sein und sich nicht oder in einem irrelevanten Bereich schneiden.

Eine Liste für häufige Fehlerquellen der Messungen befindet sich im Anhang (A.2). Nach der erfolgreichen Durchführung eines Messdurchlaufs werden die gespeicherten Daten wie zuvor verarbeitet und die entstanden Geraden auf deren Relevanz geprüft, um folgend ausgewertet zu werden.

# 5 Auswertung der Messungen

Dieses Kapitel widmet sich der Auswertung der unterschiedlichen Messungen. Um eine Kreuzpeilung mittels Target und Mikrofon-Daten vornehmen zu können, soll der Schnittpunkt mehrerer Datensets einer Messung ermittelt werden. Hierfür wurden verschiedene Varianten getestet. Im Folgenden werden die unterschiedlichen Versuche erörtert und die Resultate ausgewertet. Im Anschluss wird die Anwendung der Koordinatentransformation beschrieben und die Anwendbarkeit des Verfahrens als Ganzes bewertet.

## 5.1 Messtechniken

Im Laufe der Arbeit wurden mehrere Varianten an Audioaufnahmen ausprobiert. Dabei ging es darum, die Ungenauigkeit des Soundfield-Mikrofons sowie Artefakte der WFS-Anlage zu reduzieren. Dabei sind die externen Einflüsse des Labors nicht veränderbar. Dementsprechend wurde das Experiment im Bezug auf die Audioaufnahme bzw. -wiedergabe angepasst. Die Varianten werden in unterschiedliche Punktquellen, alternative Audiosignale und Persontracking-Aufnahmen gruppiert beschrieben.

### 5.1.1 Messgenauigkeit

Im Raum herrschen keine idealen Aufnahmebedingungen. Es fehlen gedämmte Wände. Eine Fensterfront und eine offene Seite sorgen für zu viel Reflexion. Aufgrund dessen und auch durch das Aufnahmegerät selbst wird von suboptimalen Messwerten ausgegangen. Um dieses zu überprüfen, wurde bei einer Messung (Abbildung 5.1) sich linienartig von einer fokussierten Punktquelle entfernt. Dabei wird die Messung in dem Tracking-KOS betrachtet, dieses arbeitet mit Millimeterangaben. Target-Koordinaten werden mit  $Tx$  nummeriert und als Viereck dargestellt. Das Audiosignal beschreibt den Winkel der aufgenommenen Quelle und wird mit einem Pfeil visualisiert.

Wahrnehmbar ist, dass die einzelnen Datensätze trotz der geradlinigen Entfernung zur Punktquelle im Winkel streuen.  $T_0$ , der mit der am nächsten befindlichen Position, besitzt die größte Ungenauigkeit. Als Ungenauigkeit wird die Abweichung von dem erwarteten Winkelwert bezeichnet. Diese Winkelungenauigkeit besteht auch bei identischer Messung. Es ist demnach nicht möglich durch Wiederholung derselben Messung den Fehler zu minimieren.

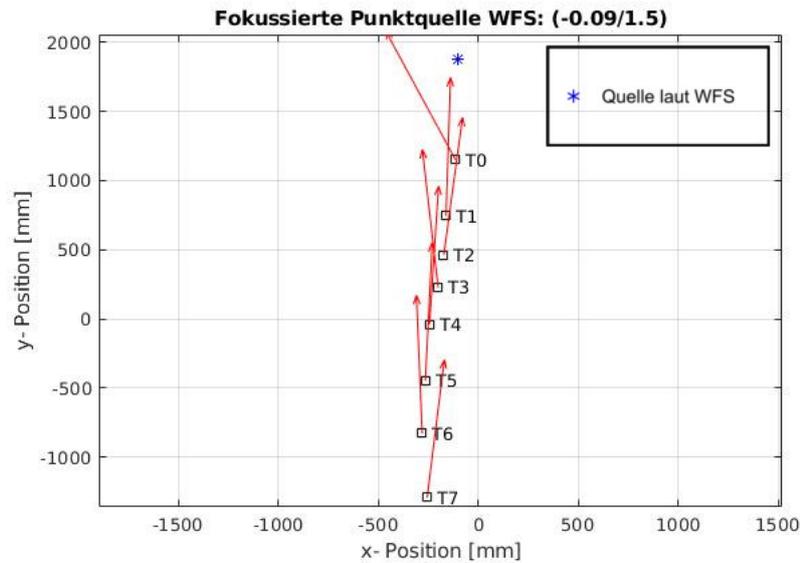


Abbildung 5.1: Messung einer fokussierten Punktquelle mit weißem Rauschen auf einer geraden Linie zur Quelle

Um die Genauigkeit der Messung zu beurteilen, ist die Standardabweichung ein weiterer Bezugspunkt. Unter Einbezug der Standardabweichung des Winkels kann ein Konfidenzintervall ermittelt werden. Der Mittelwert, folgend  $mean(\gamma)$ , wird über die Gesamtheit der Intervalle berechnet. Für die Standardabweichung  $std(\gamma)$  wird das Audiosignal unterteilt, damit für jeden Abschnitt der Winkel berechnet werden kann.

Konfidenzintervalle haben ein Konfidenzniveau, welches angibt, wie groß die Wahrscheinlichkeit ist, dass eine Messung in diesen Bereich fällt. Dabei ist die Nummerierung mit der Genauigkeit des Ergebnisses in dem Bereich aufsteigend. So handelt es sich bei dem Konfidenzintervall  $\gamma = mean(\gamma) + / - std(\gamma)$  um ein Konfidenzniveau von 68.2%.

Aus einer Messung mit fokussierter Punktquelle und weißem Rauschen wurden exemplarisch drei Datensätze,  $T_5$ ,  $T_9$  und  $T_{13}$  (Abbildung 5.2), ausgewählt. Hierbei handelt es sich nur um die beispielhafte Anwendung des Konfidenzintervalls. Ein Rechteck stellt

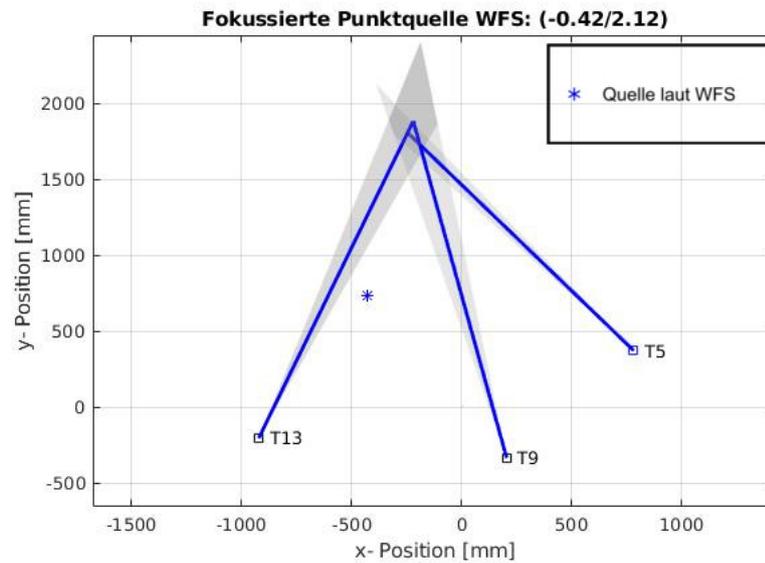


Abbildung 5.2: Exemplarischer Konfidenzintervall von drei sich schneidenden Geraden

die Target-Koordinaten dar. Eine Linie markiert den entsprechenden Winkel. Mit einem Stern wird die ungefähre Audioquelle gekennzeichnet. Diese Position der Quelle wird durch die von Hand eingestellte Umrechnung des MoWeC bestimmt.

Die drei Geraden schneiden sich zwar nicht an demselben Punkt, jedoch in geringem Abstand zueinander. Um die Geraden befindet sich ein grauer Bereich, eingegrenzt von der Standardabweichung. Dort wo sich die Geraden schneiden, ist der Grauton dunkler, da die Bereiche übereinander liegen. Dies ist der Vertrauensbereich, in dem sich das Ergebnis befindet. Nicht zu vernachlässigen ist dennoch, dass die Quelle der WFS sich nicht in diesem Bereich befindet.

### 5.1.2 Sinussignal vs weißes Rauschen

Ein weitere entscheidender Faktor ist die Art des Audiosignals für die Aufnahme. Die WFS-Anlage gibt Signale mit einer Frequenz von 48kHz wieder. Bei einem 220 Hz Sinussignal werden 220 Samples benötigt, um eine gesamte Periode abzudecken. Mit einem Sinussignal sind Phasenverschiebungen gut zu veranschaulichen. Jedoch kann es zu einem Phänomen von stehenden Wellen kommen. Dies passiert, wenn die Sinuswelle reflektiert wird und in gleicher Frequenz und Amplitude die einfallende Welle überlagert.

Alternativ kann ein weißes Rauschsignal verwendet werden, um die stehenden Wellen zu umgehen. Es werden hierbei unterschiedliche Frequenzen überlagert. Dadurch existiert eine höhere Intensität in dem Signal. Gewählt wurde ein weißes Rauschen mit Tiefpassfilter von 1,5k Hz. Die Frequenz für den Filter ist auf Aliasing-Effekte zurückzuführen. Der Grund dafür ist, dass sich ab einer Grenzfrequenz Schallfelder in der WFS-Anlage nicht mehr richtig rekonstruieren lassen, wodurch es zu Sampling-Fehlern kommt. Diese Frequenz ist von dem Abstand der Lautsprecher zueinander abhängig. [Ver97]

$$f_{max} = \frac{c}{2\Delta x} \quad (5.1)$$

In diesem Fall, mit einem Lautsprecherabstand von  $x = 0,1 m$  und der Schallgeschwindigkeit  $c = 343 \frac{m}{s}$ , beträgt die Aliasing-Frequenz 1,7k Hz. Demnach wird mit dem gewählten Filter sichergestellt, nicht in den Frequenzbereich der Artefakte zu gelangen.

Beide Signalarten haben ihre Stärken und Schwächen. Aus diesem Grund wurden Sinus (Abbildung 5.3) und weißes Rauschen (Abbildung 5.4) als fokussierte Punktquelle in der WFS-Anlage getestet. Zu lesen sind die beiden Grafiken wie zuvor, Target-Koordinaten als Rechtecke, die dazugehörigen Winkel als Linien und in Grau dessen Standardabweichung. Hinzu kommt die berechnete Audioquelle als durchschnittlicher Schnittpunkt relevanter Geraden. Anzumerken sei hier, dass die WFS-Position der Quelle verloren gegangen ist und die Messungen demnach nicht zur Berechnung der Transformation verwendet werden kann.

In der Sinussignalmessung (Abbildung 5.3) zeigen die Pfeile in unterschiedliche Richtungen. Zu erwarten wäre, dass sich die Geraden an einem Punkt schneiden; dem Punkt der fokussierten Punktquelle. Auch ist die errechnete Standardabweichung nicht sichtbar, da sie Minimal ausfällt. Das Sinussignal weist auch bei wiederholten Messungen deutliche Ungenauigkeit auf. Die berechneten Aufnahmewinkel waren teilweise sehr fraglich.

Die Geraden der Messung mit weißem Rauschen als Punktquelle (Abbildung 5.4) hingegen haben teilweise deutlich größere Abweichungen. Dennoch existiert mit den Datensätzen  $T2$  bis  $T4$  ein valider Schnittpunkt.  $T0$ ,  $T1$  und  $T5$  sind allerdings ungenau. Sie liegen außerhalb des korrekten Renderingbereichs für fokussierte Punktquellen, weswegen die Winkel auf die signalerzeugenden Lautsprecher zeigen. Folglich ist deren Ungenauigkeit ein Resultat des Synthetisierens der Quelle. Es sei an diesem Punkt anzumerken, dass es sich bei der Messung um einen schwierig reproduzierbaren Idealfall, der schneidenden Geraden, handelt.

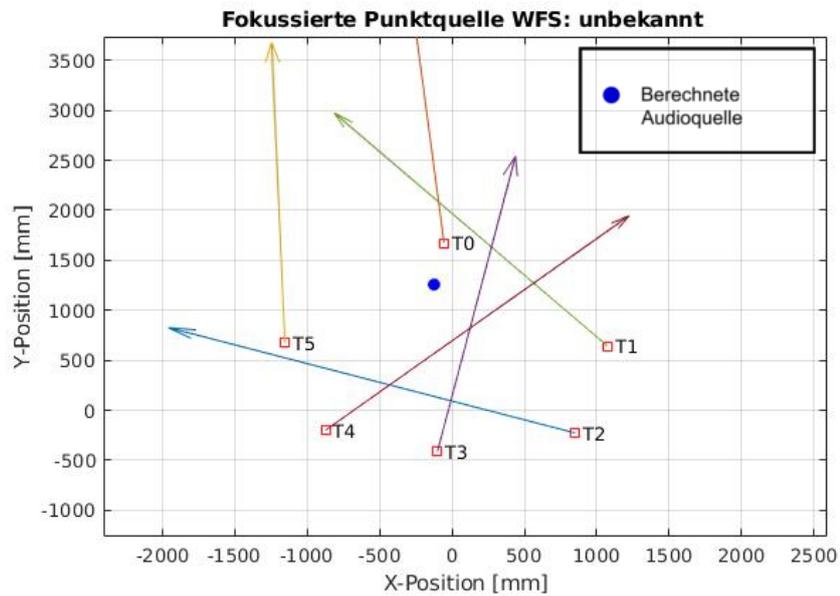


Abbildung 5.3: Messung einer fokussierten Punktquelle mit einem Sinussignal (220 Hz) ohne Persontracking

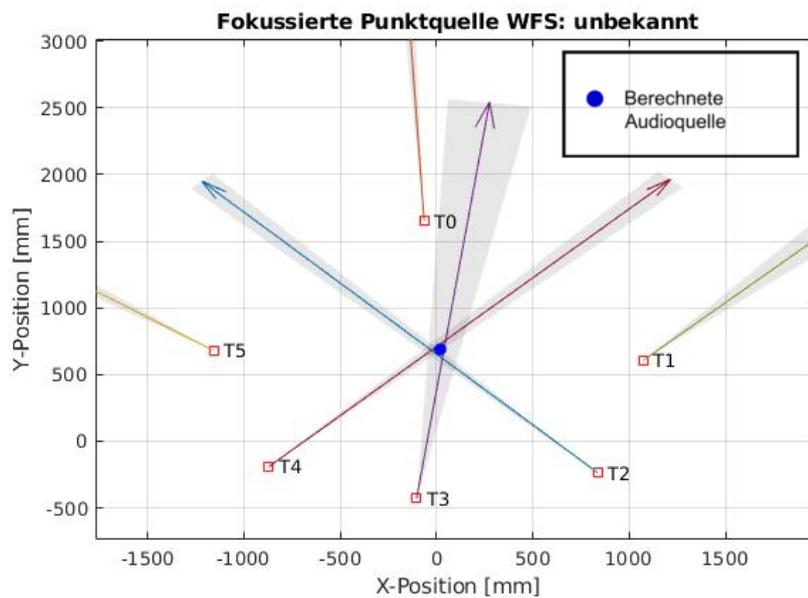


Abbildung 5.4: Messung einer fokussierten Punktquelle mit weißem Rauschen (TPF 1,5k Hz) ohne Persontracking

Unter Einbezug aller Geraden ist die Abweichung zu stark, um einen sinnvollen Schnittpunkt zu bestimmen. Einige Datensätze müssen zur Berechnung aussortiert werden. Ansonsten führen diese zu Diskrepanzen. Dies wird an den driftenden Geraden im oberen Bereich der Messung (Abbildung 5.4) deutlich, die wie schon erwähnt außerhalb des korrekt darstellbaren Bereichs liegen. Mit händischen Aufwand kann die Genauigkeit der Schnittpunkte für die Koordinatentransformation nicht ausreichend optimiert werden. Dennoch kann das weiße Rauschen unter entsprechenden Umständen für eine Schnittpunktberechnung verwendet werden. Infolgedessen wurde sich für das Rauschen mit Tiefpassfilter entschieden.

### 5.1.3 Unterschiedliche Punktquellen

Es kommen zwei Arten von Punktquellen für die Audiowiedergabe in Betracht; externe und fokussierte Punktquellen. Externe Quellen liegen außerhalb des WFS-Bereichs. Daher ist der Abstand zum Mikrophon größer, als zu einer im Bereich liegenden Audioquelle. Aufgrund der Distanz führen Messungenauigkeiten zu größeren Abweichungen.

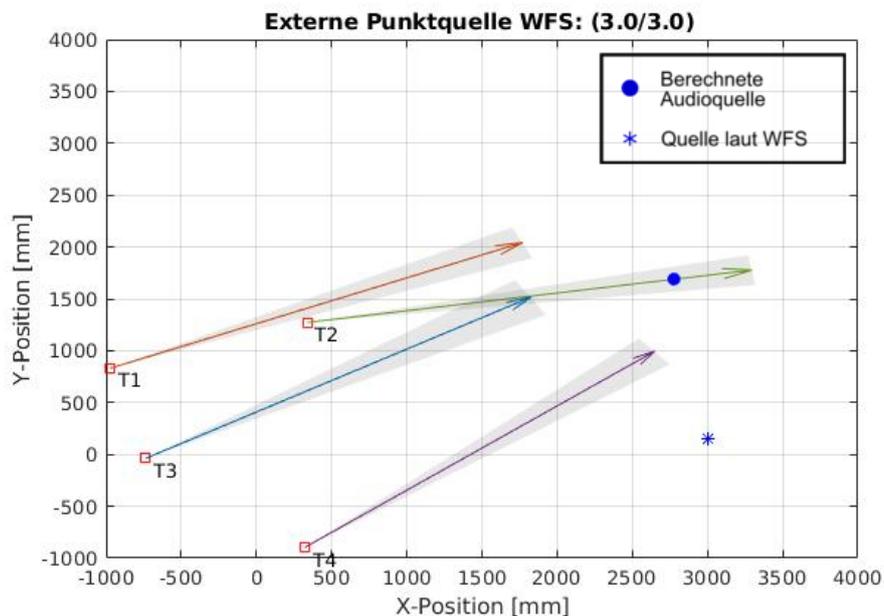


Abbildung 5.5: Messung einer externen Punktquelle mit weißem Rauschen ohne Person-tracking

Bei der externen Quellenmessung (Abbildung 5.5) werden die Target-Koordinaten und Winkel wie zuvor dargestellt. Auffällig sind die Geraden, welche sich zwar schneiden, jedoch liegen teilweise Meter zwischen den einzelnen Schnittpunkten. Bspw.  $T1$  und  $T3$  (5171/3545), sowie  $T2$  und  $T4$  (3695/1846) haben einen Versatz von 2.25 m. In Abbildung 5.4, der fokussierten Quelle, hingegen schneiden sich die einzelnen Geraden  $T2-T4$  deutlich näher an einem Punkt. Auch wenn die Richtungen der restlichen Geraden strittig sind. Des Weiteren ist bemerkbar, dass die Position der Quelle laut WFS einen deutlichen Versatz zu der triangulierten Quelle aufweist.

Bei der externen Punktquelle zeigen alle Datensätze in eine ähnliche Richtung. Die oberhalb der fokussierten Punktquelle zeigen hingegen in keine relevante Richtung. Dies ist auf die unterschiedlichen Verfahren beim Rendering zurückzuführen. Der Vorteil der Fokussierten ist dennoch, dass ein vollständiges umrunden der Quelle möglich ist, wenn auch mit strittigen Werten. Eine mögliche Verbesserung des validen Wahrnehmungsbereichs wäre demnach mittels Persontracking des MoWeC.

### 5.1.4 MoWeC Persontracking

Persontracking [Chr14] ist eine Erweiterung des MoWeC. Diese Funktion soll das Problem des Sweets Points für fokussierte Punktquellen im Raum beheben. Allerdings verwendet es die von Hand eingemessene Transformation. Dennoch wurde getestet, ob hiermit verwendbare Daten gesammelt werden können. Dabei wurde das Persontracking auf dem Stativ-Target TT5 angewendet.

Die Messung (Abbildung 5.6) ist wie der vorige Punktquellentest zu lesen. Übersichtshalber wurde in dieser Grafik die Standardabweichung ausgeblendet. Des Weiteren ist die ungefähre Audioquelle als Stern eingezeichnet. Sie besitzt die WFS-Koordinaten  $(-0.42/2.12)$ . In der Messung mit Persontracking (Abbildung 5.6) zeigen die Winkel vermehrt in Richtung der Punktquelle. Dabei liegt die ungefähre Position der Audioquelle laut glswfs deutlich vor dem durchschnittlichen Schnittpunkt der Geraden. Auch existiert ein Bereich, in dem einige der Winkel ungenau sind. Bei diesem Bereich handelt es sich um die Region, in dem der MoWeC das Standard-Rendering anfängt.

Bei Wiederholung der Messung ohne die Erweiterung (Abbildung 5.7) ist sichtbar, dass die Winkel oberhalb der Quelle deutlich ungenauer sind. Sie zeigen, wie zu erwarten, in Richtung der Lautsprecher, welche zur Erzeugung der Quelle angesteuert werden. Die Datensätze in dem Sweet Point schneiden sich jedoch auch oberhalb des erwarteten Punktes.

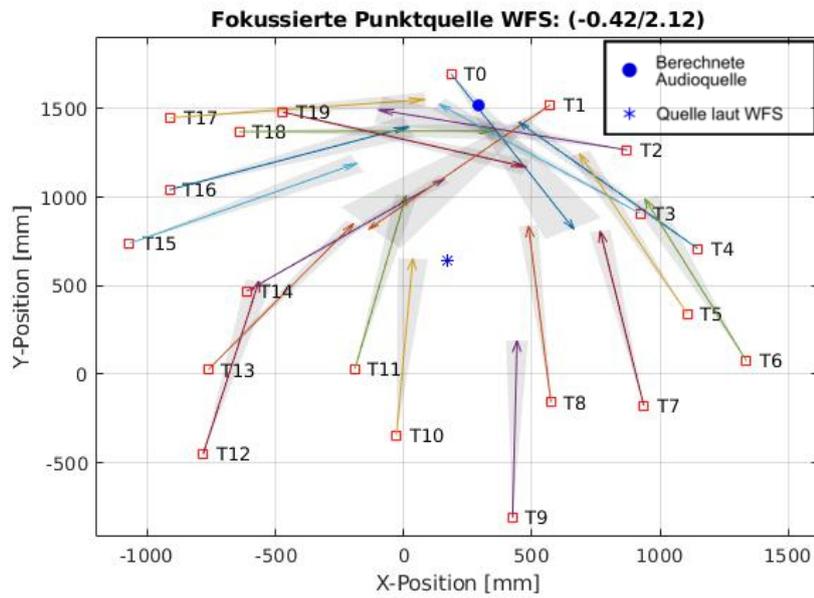


Abbildung 5.6: Messung einer fokussierter Punktquelle mit weißem Rauschen und Persontracking

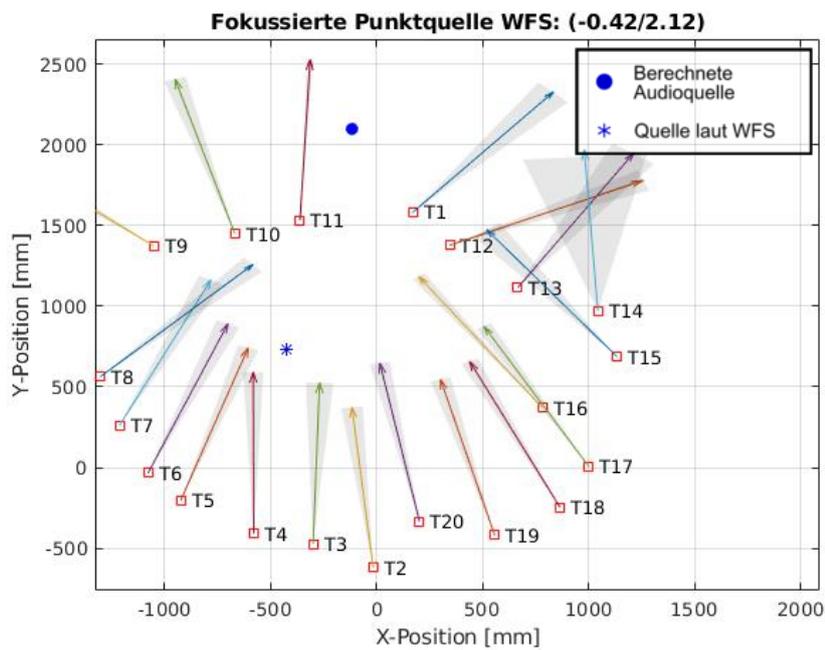


Abbildung 5.7: Messung mit fokussierter Punktquelle mit weißem Rauschen ohne Persontracking

Die relevanten Geraden schneiden sich deutlich versetzt ( $-200/1500$ ) zur angenommenen Audioquelle ( $-421/727$ ). Demnach besteht über einen Meter Versatz in der berechneten und der WFS erzeugten Audioquelle.

Mit der Persontracking-Erweiterung zeigen die Datensets eher in Richtung der Audioquelle, als ohne. Dennoch ist eine bestehende Ungenauigkeit in den Winkeln zu erkennen. Dies beeinträchtigt eine genaue Lokalisierung der Punktquelle. Grund ist der zu grobe Bereich, in dem sich die Datensätze schneiden. Außerdem wird die Punktquelle an einer bereits transformierten Position erzeugt. Deshalb stellt dies nur bedingt eine Alternative zur fokussierten Quelle ohne das Persontracking dar.

### 5.2 Durchschreitung der fokussierten Punktquelle

Ein anderer Ansatz, anstelle der Triangulation, ist ebenfalls mithilfe von Persontracking. Hierbei wird getestet, ob es eine auswertbare Veränderung bei dem Durchwandern einer fokussierten Audioquelle gibt. Dabei befindet Mikrofon und Target im Persontracking-Modus. Das Signal der Aufnahme wurde im rohen A-Format ausgewertet.

Die vier Kanäle  $LF$ ,  $RF$ ,  $LB$ ,  $RB$  in Abbildung 5.8 stellen die Audioaufnahme dar. Auffallend ist, dass die Signale  $RB$  und  $RF$  lauter werden, während  $LB$  hingegen leiser wird. Der genaue Moment des Umschwungs ist jedoch nicht zu erkennen. Ein Indiz könnte das  $LB$  Signal sein. Bei diesem ist zur dritten Sekunde ein kleiner Einbruch der Amplitude sichtbar. Grund hierfür kann aber auch das Wiedergabesignal sein, welches aus zufälligen Frequenzen erzeugt wurde. Eindeutige Einbrüche, die auf ein Zeitpunkt eines besonderen Ereignisses schließen lassen, sind weder im A-Format noch im B-Format sichtbar. Daher liefert diese Technik keine verwendbaren Ergebnisse für die Koordinatentransformation.

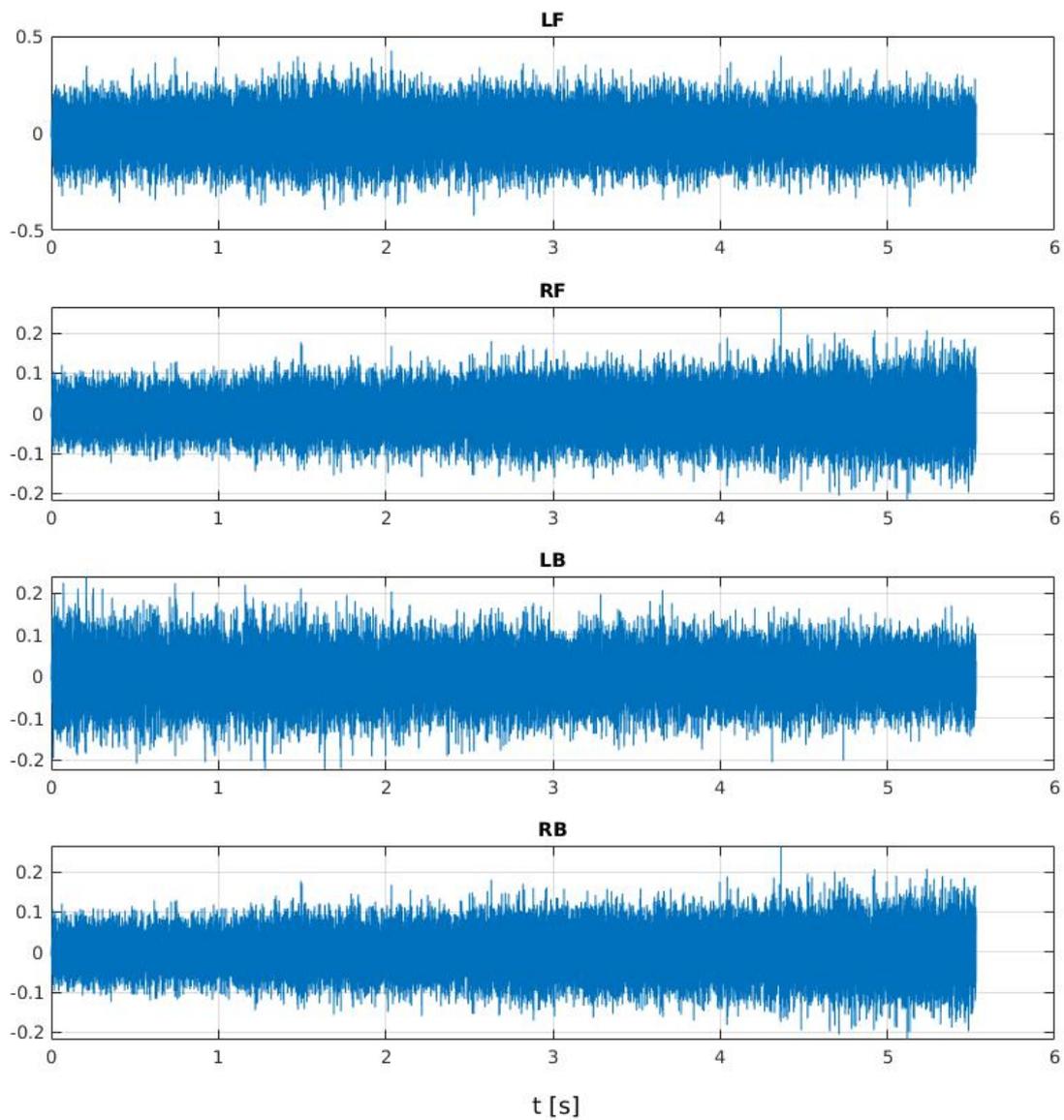


Abbildung 5.8: A-Format Aufnahme während der Durchschreitung einer fokussierten Punktquelle mit weißem Rauschen

### 5.3 Bewertung der unterschiedlichen Messungen

Aus den Messungen geht hervor, dass die besten Winkelergebnisse für das Triangulationsverfahren mit einer fokussierten Punktquelle und weißem Rauschen erzeugt werden. Zuvor wird das Wiedergabesignal mit einem Tiefpassfilter 1,5k Hz für die Zwecke verbessert, damit es zu keinen Aliasing-Artefakten kommen kann [Win17]. Damit liegt die Aufnahme noch in den mittleren Frequenzen, welche unter den Laborbedingungen mit einem Soundfield-Mikrofon gut zu messen sind [WMW13].

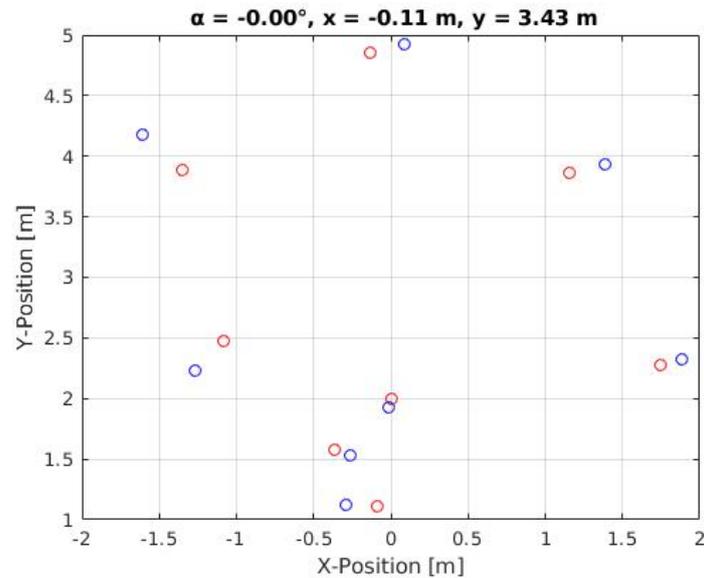
In der WFS-Anlage herrschen jedoch keine idealen Laborbedingungen, daher sind die Winkel der Datensets teilweise unbrauchbar für die Triangulation. Eine weitere Fehlerquelle könnte die Ambisonics Ordnung sein. Forschungen bestätigen, dass erst ab Ambisonics vierter Ordnung die Aufnahmen signifikant genauer zur Lokalisierung werden. Bei den Untersuchungen wurden explizit eine Aufnahme richtung  $\gamma \leq 120^\circ$  begutachtet. Dabei wiesen Aufnahmen erster Ordnung relevante Abweichungen in Bezug zum Winkel. Die Wahrnehmung der lateralen Quellen zeigten eine Abweichung von bis zu  $180^\circ$  Azimut. ([BDP<sup>+</sup>07], [BF11])

### 5.4 Berechnung der Transformation

Die Schnittpunkte aus den Messungen ergaben eine Datenwolke von 8 Werten (Tabelle 5.1). Verwendet werden diese Punkte zur Identifizierung der Translation ( $x_T, y_T$ ) sowie des Rotationswinkels  $\alpha$  vom Tracking- in das WFS-System. Hierfür wird ein Kostenfunktional berechnet.

WFS Punktquelle		Ermittelte Quelle	
x	y	x	y
-0.36	1.58	-150	1900
0	2	100	1500
-1.08	2.48	-1154	1200
-1.35	3.89	-1500	-750
-0.13	4.86	200	-1500
1.16	3.86	1500	-500
1.75	2.28	2000	1100
-0.09	1.11	-180	2300

Tabelle 5.1: Ermittelte Schnittpunktkoordinaten der Messungen und deren WFS-Koordinaten



htb

Abbildung 5.9: Anwendung der Koordinatentransformation nach dessen Identifizierung mit den ermittelten Schnittpunktswerten

In Matlab kann für die Lösung einer nicht linearen multivariablen Funktion, wie das Kostenfunktional, von *fmincon* Gebrauch gemacht werden. Die Funktion verwendet constrained optimization. *fmincon* erwartet für jede Variable eine entsprechende obere und untere Grenze. Diese müssen grob geschätzt werden. Bei der Anwendung darf die Händigkeit nicht außer Acht gelassen werden. In diesem Fall muss die y-Richtung negiert werden, damit beide KOS dieselbe Händigkeit erhalten.

Unter Verwendung der Matlab-Funktion wird mit den ermittelten Schnittpunkten die Koordinatentransformation bestimmt. Dabei wurde für die Rotation  $\alpha = 0^\circ$  sowie die Translation  $(-0.11/3.34)$  berechnet. Mit diesen Werten kann die Transformation von den Target-Koordinaten in das WFS-KOS erfolgen (Abbildung 5.9). In Rot sind die WFS- sowie in Blau die transformierten Target-Koordinaten dargestellt. Zu erwarten wäre, dass die jeweiligen Punkte übereinander liegen. Allerdings liegt fast ein halber Meter Versatz zwischen der WFS-Position und der aus dem Tracking-KOS transformierten Koordinate.

Diese Differenz in der Transformation könnte aufgrund von Ungenauigkeiten bei der Schnittpunktbestimmung entstanden sein. Daher wurde zur Beurteilung der Anwendbarkeit dieser Methode ein separater Datensatz an Punkten gesammelt. Hierfür wurde eine Audioquelle auf ein Target mittels MoWeC gesetzt. So konnten jeweils WFS- sowie

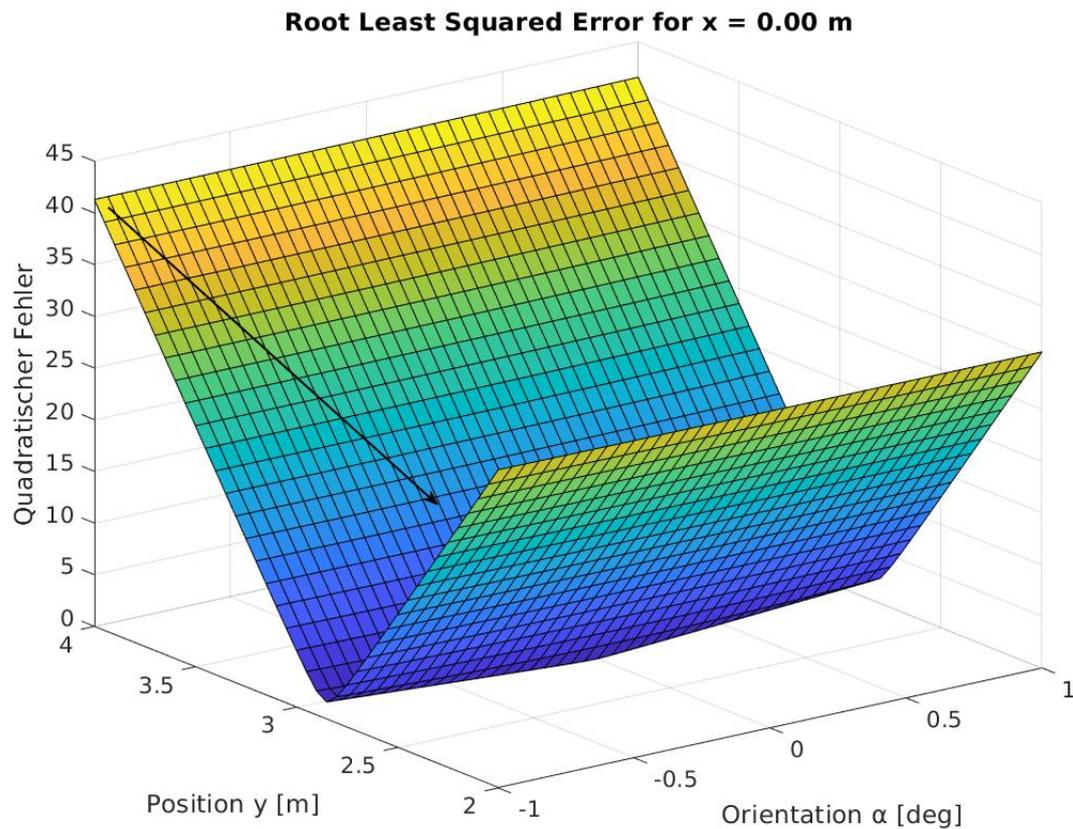


Abbildung 5.10: Plot des Kostenfunktionalis der vom MoWeC transformierten Punkte

Target-Koordinaten im Tracking-KOS bestimmt werden. Die Punkte enthalten allerdings bereits die Koordinatentransformation des MoWeC, deshalb werden sie nur zum Testen verwendet. Als Ergebnis dieser Methodik kann höchstens die angewendete Transformation bestimmt werden.

Die Fläche des Lösungsraums ist in Abbildung 5.10 visualisiert. Um diese im dreidimensionalen Raum abzubilden, muss eine Variable fix sein. In diesem Fall handelt es sich um die Translation auf der  $X$ -Achse. Zu erkennen ist eine kleine Wölbung an der Fläche (Abbildung 5.10) die das Minimum des Kostenfunktionalis abbildet. Um diese Spitze genauer zu betrachten, wird in Abbildung 5.11 die Ebene dargestellt. Farblich gekennzeichnet sind die Höhenlinien, Gelb markiert die höchste Stelle. Das Minimum liegt am Schnittpunkt der beiden Geraden.

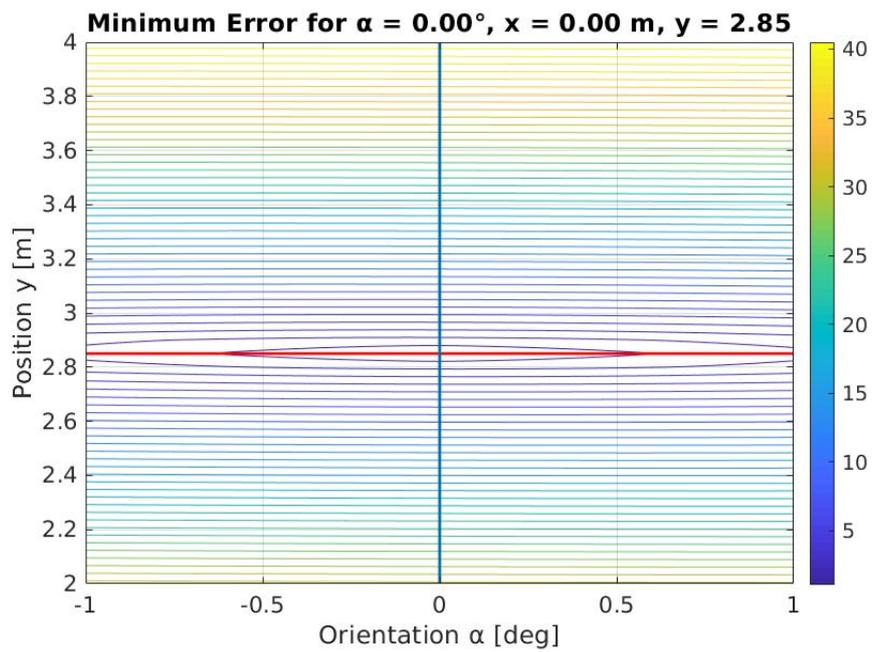


Abbildung 5.11: Ebene des minimalen quadratischen Fehlers

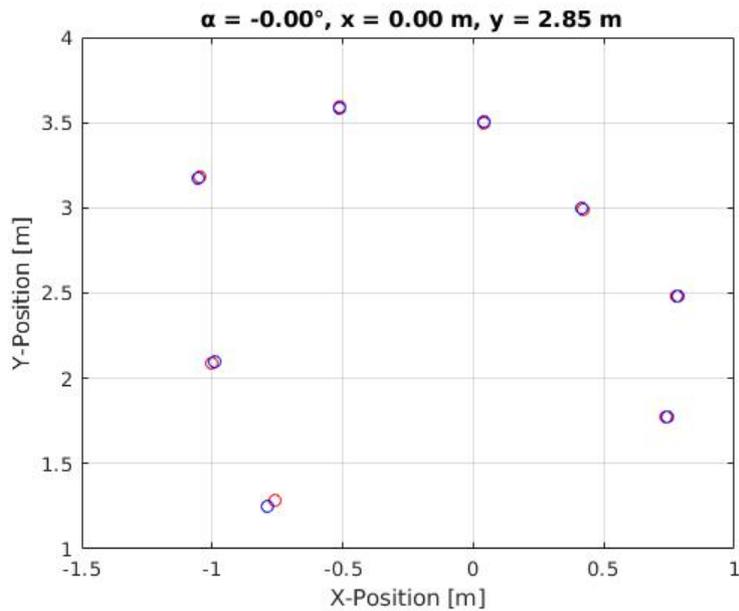


Abbildung 5.12: Anwendung der Koordinatentransformation nach dessen Identifizierung mit Referenzwerten aus dem Kostenfunktionalbeispiel

Die Transformation im MoWeC verwendet eine Translation von (0.0/2.85) mit der Invertierung der Y-Koordinaten und eine Rotation von 0°. Das entspricht den berechneten Werten (Abbildung 5.11). Die verwendeten Daten für diese Berechnung sind die bereits im MoWeC transformierten Werte. Die angewendete Transformation (Abbildung 5.12) bestätigt die Genauigkeit der Koordinatentransformation. In Rot wieder die WFS- sowie in Blau die transformierten Werte. Wie zu erwarten liegen die jeweiligen Punkte nah aufeinander.

	Verwendete Transformation im MoWeC	Berechnete Transformation mit MoWeC-Werten	Berechnung anhand von ermittelten Schnittpunkten (Tabelle 5.1)
$\alpha$ [°]	0	$-1.35e^{-7}$	$-4.76e^{-9}$
x [m]	0	0.0012	-0.1121
y [m]	2.85	2.8517	3.428

Tabelle 5.2: Vergleich der Koordinatentransformation mit unterschiedlichen Werten

Bei Betrachtung der Transformationen mit den unterschiedlichen Werten miteinander (Tabelle 5.2) wird ersichtlich, dass die berechneten Werte mit der bereits transformierten Audioquelle fast identisch mit den Werten im MoWeC sind. Die Abweichungen sind minimal. Demnach kann die Bestimmung der Transformation als funktionsfähig angesehen werden. Bei den ermittelten Schnittpunkten hingegen fällt eine Diskrepanz der Werte auf. Die Y-Koordinate hat über einen halben Meter mehr Versatz als die MoWeC-Transformation. Auch der Winkel weist einen Unterschied auf. Die Unterschiede könnten auf unzureichende Einschränkung des Lösungsbereichs zurückzuführen sein. Eine visuelle Kontrolle des Kostenfunktional ergab jedoch dieselben Transformationswerte. Die Hauptfehlerquelle ist die Ungenauigkeit bei der Bestimmung der Schnittpunkte.

Die Messungenauigkeiten entstehen teilweise aufgrund der Laborverhältnisse sowie durch das Soundfield-Mikrofon. Anlässlich dieser Erkenntnis lässt sich die Aussage treffen, dass die Hardware und der Raum ungeeignet sind, die Theorie der Triangulation in die Praxis umzusetzen. Infolgedessen ist der Testaufbau zum Zwecke der Bestimmung der Koordinatentransformation zwischen den KOS nicht nutzbar.

## 6 Fazit und Ausblick

Das Ziel dieser Arbeit war es, die Koordinatentransformation zwischen der WFS-Anlage und dem Tracking System zu bestimmen. Bisher wurde sie händisch ermittelt und sollte mithilfe von Sensorik mathematisch präzisiert werden. Hierfür wurde ein Aufbau zur Datenerfassung sowie eine Auswertungskette erarbeitet. Dabei wurde erfolgreich das optische Tracking-System mit dem Ambisonics-Mikrofon verknüpft. Nach Gewinnung der Messdaten verläuft die Auswertungskette bis auf den Wechsel von der Aufnahmeanwendung zu Matlab, sowie der händischen Auswahl der relevanten Datensätze durchgängig automatisch. In dieser Arbeit wurde untersucht und bewiesen, dass mittels einfacher trigonometrischer Berechnungen und Annäherungsverfahren die Transformation in der Theorie möglich ist.

Bei den Berechnungen galt es einen Überblick über die einzelnen Koordinatensysteme zu behalten. Grund hierfür war, dass sie sich nicht nur durch die Rotation und Translation voneinander unterscheiden, sondern auch in ihrer Händigkeit. Die Berechnung der Transformation konnte mittels Testdaten aus dem MoWeC auf Funktionalität bestätigt werden. Anhand dieser konnte die Transformation die im MoWeC verwendet wird erfolgreich bestimmt werden.

Das bestehende Problem lag in der Triangulation mit den erfassten Daten. Im Besonderen in den Aufnahmen mit dem Ambisonics-Mikrofon, welche zu ungenau sind, um die Triangulation präzise zu berechnen. Außerdem sorgt der Raum durch die vielen technischen Geräte und einer Klimaanlage, die ein Grundrauschen erzeugt, für Störgeräusche. Hinzu kommt auch, dass der Raum selbst nicht optimal geschaffen ist und Hall dadurch zu einem nicht vernachlässigbaren Aspekt wird. Daher wurden zur Verbesserung der Messwerte unterschiedliche Arten von Audioaufnahmen und -wiedergaben untersucht. Diese Werte sind ausschlaggebend für die Berechnung der Koordinatentransformation auf Basis der triangulierten Werte. Dabei stellte sich heraus, dass Audioaufnahmen von fokussierten Punkzellen mit weißem Rauschen, welches mit einem Tiefpassfilter von 1,5k Hz gefiltert wurde, die besten Ergebnisse ergaben.

Forschungen besagen, dass Ambisonics ab der vierten Ordnung für Lokalisierungsmessungen genauere Ergebnisse liefert ([BDP<sup>+</sup>07], [BF11]). Diese Erkenntnis könnte in weiterführenden Arbeiten untersucht werden. Da die externen Einflüsse des Raums gegeben sind, könnte die Kalibrierung des Mikrofons eine zusätzliche Verbesserungsmöglichkeit sein [HB12].

Ein gänzlich anderer Ansatz zur Audiolokalisierung wäre die Verwendung eines Mikrofon-Arrays [VMRE04]. Das Software-Framework ManyEars [GLF<sup>+</sup>13] ortet in Echtzeit mit acht Mikrofonen bis zu vier Audioquellen. Aber auch neue Lokalisierungstechniken wie GEVD MUSIC (Generalized EigenValue Decomposition Multiple Signal Classification) und GSVD-MUSIC (Generalized Singular Value Decomposition Multiple Signal Classification) könnten in Erwägung gezogen werden, um bessere Ergebnisse zu erzielen [NNI12].

# Literaturverzeichnis

- [Ahr12] AHRENS, Jens: *Analytic Methods of Sound Field Synthesis*. 2012
- [ART06] ART GMBH: *DTrack ABCman Manual*. Am Oeferl 6, 82362 Weilheim i.OB, Germany, 2006
- [ART12] ART GMBH: *DTrack2 User Manual V2.12.0 ART DTrack 2.12*. Am Oeferl 6, 82362 Weilheim i.OB, Germany, 2012
- [AW02] ABHAYAPALA, T. D. ; WARD, D. B.: Theory and Design of High Order Sound Field Microphones using Spherical Microphone Array. In: *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing* Bd. 2, 2002
- [Bam94] BAMFORD, J.S.: Ambisonic sound for the masses. In: *Canadian Acoustics* 22 (1994), September, S. 63 – 64
- [BDP<sup>+</sup>07] BERTET, Stéphanie ; DANIEL, Jerome ; PARIZET, Etienne ; GROS, Laetitia ; WARUSFEL, Olivier: Investigation of the Perceived Spatial Resolution of Higher Order Ambisonics Sound Fields: A Subjective Evaluation Involving Virtual and Real 3D Microphones. (2007), März, S. 15 – 17
- [Ber88] BERKHOUT, A. J.: A Holographic Approach to Acoustic Control. In: *Journal of The Audio Engineering Society* 36 (1988), S. 977 – 995
- [BF11] BRAUN, Sebastian ; FRANK, Matthias: Localization of 3D Ambisonic Recordings and Ambisonic Virtual Sources, 2011
- [BHSK07] BAALMAN, Marije ; HOHN, Torben ; SCHAMPIJER, Simon ; KOCH, Thilo: Renewed Architecture of the WONDER Software for Wave Field Synthesis on large Scale Systems, 2007
- [Chr14] CHRISTOFFEL, Carola: *Modifikation der Software einer Wellenfeldsyntheseanlage zur Wiedergabe fokussierter Quellen in Abhängigkeit der Zuhörerposition*. 2014

- [Fac] FACEBOOK: *How do I upload a 360 video on Facebook?* <https://www.facebook.com/help/828417127257368>. – Abruf: 20.09.2020
- [Feh18] FEHLING, Matthias: *Implementierung von Ambisonics Mikrofonsignalen in die Wellenfeldsyntheseanlage der HAW-Hamburg*. 2018
- [Fel74] FELLGETT, Peter: Ambisonic Reproduction of Directionality in Surround-Sound Systems. In: *Nature* 252 (1974), S. 534 – 538
- [Foh] FOHL, Wolfgang: *i2audiolab*. <https://i2audiolab.de/>. – Abruf: 16.10.2020
- [FZ19] FRANZ ZOTTER, Matthias F.: Ambisonics. In: *A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer International Publishing, 2019, S. 243 – 255
- [GLF<sup>+</sup>13] GRONDIN, François ; LÉTOURNEAU, Dominic ; FERLAND, François ; ROUSSEAU, Vincent ; MICHAUD, François: The ManyEars open framework. In: *Autonomous Robots* 34 (2013), April
- [HB12] HELLER, Aaron ; BENJAMIN, Eric M.: Calibration of Soundfield Microphones Using the Diffuse-Field Response. In: *Journal of the Audio Engineering Society* (2012), Oktober
- [HM06] HERING, Ekbert ; MARTIN, Rolf: Optische Messtechnik. In: *Photonik: Grundlagen, Technologie und Anwendung*. Springer Berlin Heidelberg, 2006. – ISBN 9783540297086, S. 233 – 281
- [Hor87] HORN, Berthold K. P.: Closed-Form Solution of Absolute Orientation Using Unit Quaternions. In: *Journal of the Optical Society of America* (1987), April, S. 629 – 642
- [HZ03] HARTLEY, Richard ; ZISSERMAN, Andrew: *Multiple View Geometry in Computer Vision*. 2. USA : Cambridge University Press, 2003. – ISBN 0521540518
- [JP] JENS PRAGER, Björn A.T. P. Uwe Buchholz B. Uwe Buchholz: Analyse komplex-geformter Schallquellen mit Hilfe der sphärischen Nahfeldholographie. In: *DAGA 2005 München*
- [Kü16] KÜPELIKILINC, Rutkay: *Untersuchung der Höhenwahrnehmung an einer um Einzellautsprecher erweiterten WFS-Anlage*. 2016

- [Lev44] LEVENBERG, Kenneth: A Method for the Solution of Certain Non-Linear Problems in Least Squares. In: *Quarterly of Applied Mathematics* Bd. 2 (1944), S. 164 – 168. – ISSN 0033569X
- [Mal99] MALHAM, David G.: Higher Order Ambisonic Systems for the Spatialisation of Sound. In: *ICMC*, 1999
- [Mon11] MONTAG, Matthew N.: *Wave Field Synthesis in Three Dimensions by Multiple Line Arrays*, University Of Miami, Master Thesis, 2011
- [NNI12] NAKAMURA, Keisuke ; NAKADAI, Kazuhiro ; INCE, Gökhan: Real-Time Super-Resolution Sound Source Localization for Robots, 2012. – ISBN 9781467317375, S. 694 – 699
- [Nog12] NOGALSKI, Malte: *Gestengesteuerte Positionierung von Klangquellen einer Wellenfeldsynthese-Anlage mit Hilfe eines kamerabasierten 3D-Tracking-Systems*, HAW Hamburg, Bachelor Thesis, 2012
- [NZDS11] NACHAR, Christian ; ZOTTER, Franz ; DELEFLIE, Etienne ; SONTACCHI, Alois: Ambix - Suggesting an Ambisonics Format, 2011
- [Por] PORTAUDIO: *PortAudio Tutorials*. [http://portaudio.com/docs/v19-doxydocs/tutorial\\_start.html](http://portaudio.com/docs/v19-doxydocs/tutorial_start.html). – Abruf: 17.08.2020
- [Rob60] ROBERT, Siebeck: Die Wellentheorie des Lichtes. Das Huygenssche Prinzip. In: *Optik des Menschlichen Auges*. Springer Berlin Heidelberg, 1960. – ISBN 9783642877193, S. 29 – 53
- [Sou11] SOUNDFIELD: *SPS200 Software Controlled Microphone User Guide*. 2011
- [SRA08] SPORS, Sascha ; RABENSTEIN, Rudolf ; AHRENS, Jens: The Theory of Wave Field Synthesis Revisited. (2008), Januar
- [Tho] THORNTON, Stephen: *Ambisonics*. <https://www.michaelgerzonphotos.org.uk/ambisonics.html>. – Abruf: 22.10.2020
- [Ver97] VERHEIJEN, Edwin: *Sound Reproduction by Wave Field Synthesis*, Technische Universiteit Delft, PhD. Thesis, 1997
- [VMRE04] VALIN, Jean-Marc ; MICHAUD, François ; ROUAT, Jean ; ETOURNEAU, Dominic: Robust Sound Source Localization Using a Microphone Array on a Mobile Robot. In: *IEEE International Conference on Intelligent Robots and Systems* 2 (2004), April

- [Wil99] WILLIAMS, Earl G.: Chapter 6 - Spherical Waves. In: *Fourier Acoustics*. London : Academic Press, 1999. – ISBN 9780127539607, S. 183 – 234
- [Win17] WINTER, Eugen: *Analyse und Reimplementierung einer plattformunabhängigen Software für das 3D-Audio-Rendering*, HAW Hamburg, Bachelor Thesis, 2017
- [WMW13] WIERZBICKI, Jacek ; MALECKI, Pawel ; WICIAK, J.: Localization of the Sound Source with the Use of the First-order Ambisconic Microphone. In: *Acta Physica Polonica A* 123 (2013), Mai, S. 1114 – 1117
- [Zm2] ZMÖLNIG, Johannes: *Entwurf und Implementierung einer Mehrkanal-Beschallungsanlage*, Erzherzog-Johann-Universität Graz und Kunstuniversität Graz, Diplomarbeit, 2002

# A Anhang

## A.1 Config-File

Hier zu sehen ist die *config.txt* Datei:

```
number-repeats:
4
pa-device:
12
wfs-ip:
192.168.14.10
wfs-port:
58100
target-group:
230.230.230.230
target-port:
5230
target-interface:
192.168.14.104
target-file:
target.csv
rec-file:
record.wav
```

Die Datei ist ein Hilfsmittel für das Datenaufnahmeprogramm, um gebündelt die Netzwerkinformationen zu überprüfen und zu ändern. Es ist immer aufgeteilt in eine Zeile Variablen Bezeichnung, gefolgt von einer zweiten Zeile mit den entsprechenden Daten.

**number-repeats** sind die Anzahl an Messungen. D.h. wie oft das Stativ bewegt wird.

Nach jedem Positionswechsel muss bestätigt werden, für die nächste Aufnahme an Messwerten.

**pa-device** steht in dieser Arbeit für das verwendete Audio-Interface, an dem das Mikrofon angeschlossen ist.

**wfs-port** ist für die Kommunikation mit der WFS notwendig.

**target-group, -port und -interface** werden benötigt, um sich an der Multicast Gruppe des Tracking-PCs anzumelden, damit die Target Informationen empfangen werden können. Das Interface ist der lokale Netzwerk Adapter, der mit dem benötigten Netzwerk verbunden ist.

**target- und rec-file** sind die Dateinamenendungen der zu speichernden Daten. Vor diesen wird noch der Messdurchlauf als Zahl angefügt.

## A.2 Häufige Fehler bei den Messungen

Der Fehlerguide dient der Unterstützung bei dem Ausführen des Aufnahmeprogramms der Messungen. Fehler werden meistens in dem Programm ausgegeben. In Ausnahmefällen kann die Anwendung sich direkt beenden, dies tritt im Falle des Hibernate-Modus auf (siehe. Tabelle A.1).

Fehler	Mögliche Ursachen	Fehlersuche und Behebung
Audio-Interface kann nicht initialisiert werden	Andere Programme blockieren das Audio-Interface  Falsches Audio-Interface	- Andere Audioprogramme schließen, neu versuchen - Print.AudioDevices ausführen, Focusrite USB 2.0 AUDIO Driver, ASIO max Outputs:8 auswählen  - Richtiges Interface in der Config.txt angeben  - evtl. LAN deaktivieren
Target wird nicht gefunden	Tracking-System hat kein Target-Objekt gefunden	- Sicherstellen, dass Tracking-PC richtiges Target im Raum erkennt
Socket kann nicht initialisiert werden	Multicast Problem	- Sichergehen, dass die Locale IP im Config.txt unter target-interface steht
Initialisierungen schlagen fehl	Windows 7 Hibernate Modus	- Reboot PC

Tabelle A.1: Häufige Fehler beim Ausführen des Aufnahmeprogramms

## Erklärung zur selbstständigen Bearbeitung einer Abschlussarbeit

Gemäß der Allgemeinen Prüfungs- und Studienordnung ist zusammen mit der Abschlussarbeit eine schriftliche Erklärung abzugeben, in der der Studierende bestätigt, dass die Abschlussarbeit „– bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit [(§ 18 Abs. 1 APSO-TI-BM bzw. § 21 Abs. 1 APSO-INGI)] – ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt wurden. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich zu machen.“

*Quelle: § 16 Abs. 5 APSO-TI-BM bzw. § 15 Abs. 6 APSO-INGI*

## Erklärung zur selbstständigen Bearbeitung der Arbeit

Hiermit versichere ich,

Name: \_\_\_\_\_

Vorname: \_\_\_\_\_

dass ich die vorliegende Bachelorarbeit – bzw. bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit – mit dem Thema:

### **Halbautomatischer Koordinatenabgleich zwischen einem optischen Positionserfassungssystem und einer Wellenfeldsynthese-Anlage**

ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

\_\_\_\_\_  
Ort

\_\_\_\_\_  
Datum

\_\_\_\_\_  
Unterschrift im Original