

Realisierung einer Software für das 3D-Audio-Rendering im Kontext immersiver Umgebungen

Eugen Winter

Hochschule für Angewandte Wissenschaften Hamburg
20099 Hamburg

E-Mail: eugen.winter@haw-hamburg.de

Zusammenfassung—Im Rahmen dieses Berichtes sollen zuerst mehrere Verfahren für das Rendering von 3D-Audio betrachtet werden. Im Vordergrund steht dabei die Nutzung der Wellenfeldsynthese. Diese weist die Besonderheit auf, dass mit ihr, im Gegensatz zu herkömmlichen Audio-Wiedergabe-Verfahren, die Möglichkeit besteht sich als Zuhörer während der Wiedergabe frei im Zuhörer Raum und um sich darin befindliche virtuelle Klangquellen herum zu bewegen. Ferner soll kurz das Verfahren des Vector Base Amplitude Panning, welches eine richtungsgebundene 3D-Audiowiedergabe ermöglicht, als mögliche Alternative vorgestellt werden. Die Realisierung beider Verfahren birgt unterschiedliche Voraussetzungen. Auf diese soll, im Hinblick auf eine konkrete Implementierung, genauer eingegangen und mögliche Ansätze untersucht werden.

Die Schaffung einer immersiven Umgebung beschränkt sich jedoch nicht alleine auf die Erzeugung einer authentischen Klangkulisse. Vielmehr gehen die akustischen und visuellen Aspekte eines immersiven Erlebnisses Hand in Hand. Der zweite Teil des Berichtes widmet sich deshalb einem Überblick über die aktuellen Technologien aus dem Bereich der Augmented, Mixed und Virtual Reality. Dabei sollen die Vor- und Nachteile sowie mögliche Einsatzzwecke und Szenarien analysiert werden, bei denen sich die klanglichen und visuellen Darbietungen ergänzen.

1. Einleitung

Die Wellenfeldsynthese (WFS) tritt vornehmlich im Umfeld von Forschungseinrichtungen in Erscheinung. Neben der anhaltenden Forschung an den eigentlichen Audio-Rendering-Verfahren liegt die geringe Verbreitung von WFS-Anlagen darin begründet, dass für deren Betrieb eine Vielzahl von Lautsprechern und Audio-Equipment sowie genügend Rechenleistung benötigt wird. Besonders große WFS-Anlagen mit hunderten von Lautsprechern werden deshalb als Verteiltes System realisiert, das aus mehreren Rendering-Clustern besteht, welche untereinander die Rechenlast aufteilen. Dennoch hat WFS gegenüber herkömmlichen Audio-Rendering-Verfahren wie Stereo, 5.1 oder Vector Base Amplitude Panning (VBAP), die nur eine kanal- und damit richtungsgebundene Ortung des Schalls zulassen,

den großen Vorteil, dass hier der Schall einer tatsächlichen Klangquelle synthetisiert und damit eine objektgebundene Schallausbreitung erzeugt wird. Zudem lassen sich die virtuell erzeugten Klangquellen sowohl innerhalb, als auch außerhalb des Zuhörer Raumes positionieren und behalten, so wie man es von echten Klangquellen erwarten würde, ihre Position auch während der Bewegung eines Zuhörers im Zuhörer Raum bei. Dadurch werden sie begehbar.

Das Arrangieren und die Interaktion mit solchen virtuellen Klangquellen findet meist von der starren Position eines Autors aus statt, ohne dass sich dieser aktiv oder interaktiv an der Klangszene beteiligt. Die Szenerie wird von diesem vorab nach genauen Vorgaben zusammengebaut und danach abgespielt. Dabei benutzt der Autor für die Herstellung der Kompositionen überwiegend eine Digital-Audio-Workstation (DAW) wie Cubase¹ oder Ardour². Für die Ansteuerung kommen zusätzlich grafische Benutzerschnittstellen (GUI) zum Einsatz, die meistens Eigenentwicklungen sind und unabhängig von der Audio-Software eine beliebige Positionierung der virtuellen Klangquellen im zwei- oder dreidimensionalen Raum erlauben. Eine Erweiterung stellt die Möglichkeit dar einen Zuhörer innerhalb des Zuhörer Raumes mit einem Tracking-System zu tracken und abhängig davon weitere Aktionen durchzuführen oder den Klang anzupassen. Es bleibt jedoch festzuhalten, dass die Erzeugung und Interaktion mit einem immersiven Szenario bei der Wellenfeldsynthese eine eher lose Koppelung aufweist.

Auf dem Gebiet der Augmented, Mixed und Virtual Reality Hardware hat sich viel getan. Neben leistungsfähigen Geräten wie der HTC Vive³ oder der Oculus Rift⁴, setzen die Hersteller besonders große Hoffnungen auf mobile Lösungen. Aktuelle Smartphones sind bereits heute schon in der Lage Augmented und Virtual Reality darzustellen. Eine WFS-Anlage ergänzt den immersiven, audio-visuellen Gesamteindruck. Umgekehrt bilden die mobilen Geräte eine ideale Grundlage für die Erforschung neuartiger Bedien- und Interaktionskonzepte mit einer WFS-Anlage.

1. <https://www.steinberg.net/en/products/cubase/start.html>

2. <https://www.ardour.org/>

3. <https://www.vive.com>

4. <https://www.oculus.com/rift>

2. Grundlagen der Wellenfeldsynthese

Die physikalische Grundlage für die Wellenfeldsynthese bildet das Huygenssche Prinzip. Dieses besagt, dass jeder Punkt einer Wellenfront als Entstehungspunkt einer neuen Welle betrachtet werden kann. Die neue Welle wird als Elementarwelle bezeichnet. Überlagern sich all diese Elementarwellen, dann bilden sie erneut die anfängliche Wellenfront (siehe Abbildung 1).

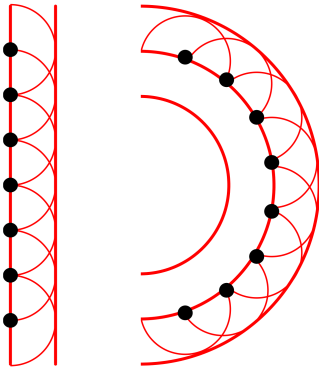


Abbildung 1. Huygensches Prinzip am Beispiel einer ebenen und kreisförmigen Wellenfront

Durch die Verwendung von Lautsprechern zur Erzeugung von Elementarwellen entwickelte Berkhout 1988 erstmals ein Verfahren für die Synthese des von einer Klangquelle erzeugten Schallfeldes [1]. Die formale Herleitung und Approximation bildet das Kirchhoff-Helmholtz-Integral sowie das Rayleigh-Integral. Die Geometrie ist in Abbildung 2 zu sehen und besagt, dass innerhalb eines Volumens, welches frei von Quellen ist, für jeden Punkt der Schalldruck ermittelt werden kann, wenn der Schalldruck und die Schallschnelle aller Oberflächenpunkte des Volumens bekannt sind [2]. Durch eine Approximation kann die Oberfläche als eine diskrete Anzahl von Lautsprechern beschrieben werden, die es ermöglichen für jeden Punkt eines Zuhörerraumes die korrekte Amplitude einer virtuellen Klangquelle zu erzeugen, die sich innerhalb oder außerhalb des Raumes befinden kann. Eine weitere Vereinfachung wird für die praxisnahe Aufstellung von Lautsprechern vorgenommen, indem die Oberfläche eines Volumens auf eine zweidimensionale, horizontale Ebene reduziert wird. Um dabei die kugelförmige Abstrahlcharakteristik eines Lautsprechers zu berücksichtigen, wurde der 2.5D-WFS-Operator eingeführt [2].

2.1. Linearquelle

Eine Linearquelle wird bestimmt durch einen Winkel. Dieser gibt den Ursprung der Klangquelle an. Dabei bleibt die Richtung des Ursprungs an jeder Position im Zuhörerraum identisch und eignet sich damit zur Abbildung sehr weit entfernter Klangquellen. Erreicht wird die Synthese des Schallfeldes durch die Vorgabe eines Winkels für den Ursprung der Linearquelle und einer korrekt aufeinander

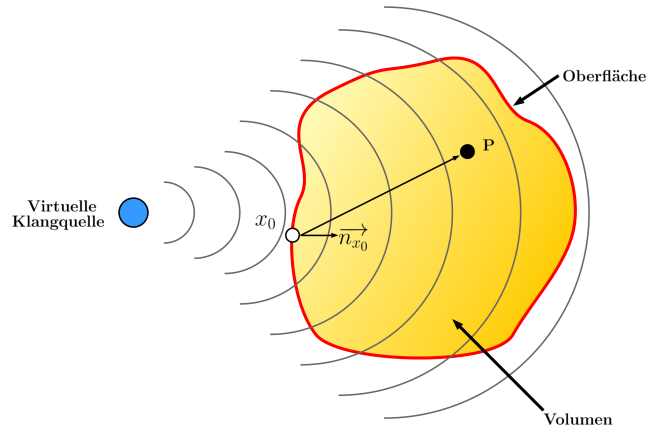


Abbildung 2. Geometrie des Kirchhoff-Helmholtz-Integrals

folgenden Verzögerung der einzelnen Lautsprecher (siehe Abbildung 3).

$$d_{2.5D_{pw}}(x_0, t) = \underbrace{2 \cdot g_0 \cdot a(t) \cdot h_{2.5D}(t) \cdot \langle n_k, n_{x_0} \rangle \cdot w(x_0)}_{\text{Amplituden-Gewichtung}} \cdot \underbrace{\delta\left(t - \frac{\langle n_k, n_{x_0} \rangle}{c}\right)}_{\text{Verzögerung}} \quad (1)$$

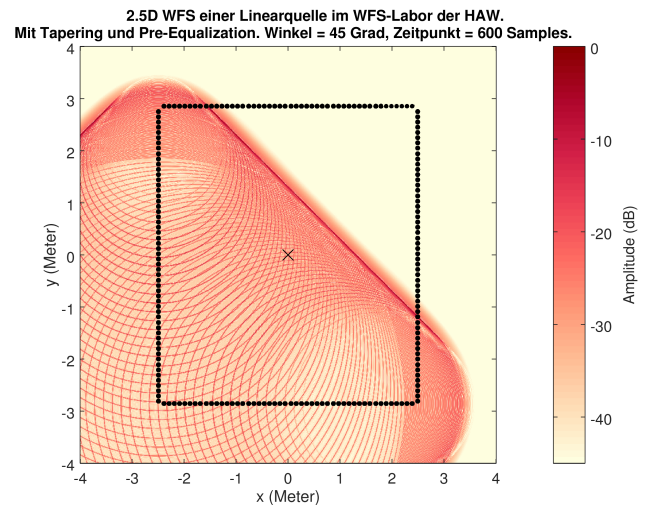


Abbildung 3. Simulation von 2.5D-WFS für eine Linearquelle in der Zeitdomäne

2.2. Punktquelle

Bei der Punktquelle handelt es sich um eine außerhalb des Zuhörerraumes befindliche virtuelle Klangquelle. Diese behält ihre Position auch dann bei, wenn sich der Zuhörer im Zuhörerraum bewegt. Damit ist sie umgehbar und

im Gegensatz zur Linearquelle genau lokalisierbar, da die Entfernung anhand ihrer Amplitude bestimmt werden kann (siehe Abbildung 4).

$$d_{2.5D_{ps}}(x_0, t) = \underbrace{\frac{1}{\sqrt{2\pi}} \cdot g_0 \cdot a(t) \cdot h_{2.5D}(t)}_{\text{Amplituden-Gewichtung}} \cdot \underbrace{\frac{\langle (x_0 - x_s), n_{x_0} \rangle}{\|x_0 - x_s\|^{\frac{3}{2}}} \cdot w(x_0)}_{\text{Amplituden-Gewichtung}} \cdot \underbrace{\delta\left(t - \frac{\|x_0 - x_s\|}{c}\right)}_{\text{Verzögerung}} \quad (2)$$

eine Basisverzögerung auf alle eingehenden Audiosignale angewendet, die sich aus dem maximalen Abstand zwischen einer fokussierten Punktquelle und einem Lautsprecher berechnet.

$$d_{2.5D_{fs}}(x_0, t) = \underbrace{\frac{1}{\sqrt{2\pi}} \cdot g_0 \cdot a(t) \cdot h_{2.5D}(t)}_{\text{Amplituden-Gewichtung}} \cdot \underbrace{\frac{\langle (x_0 - x_s), n_{x_0} \rangle}{\|x_0 - x_s\|^{\frac{3}{2}}} \cdot w(x_0)}_{\text{Amplituden-Gewichtung}} \cdot \underbrace{\delta\left(t + \frac{\|x_0 - x_s\|}{c}\right)}_{\text{Verzögerung}} \quad (3)$$

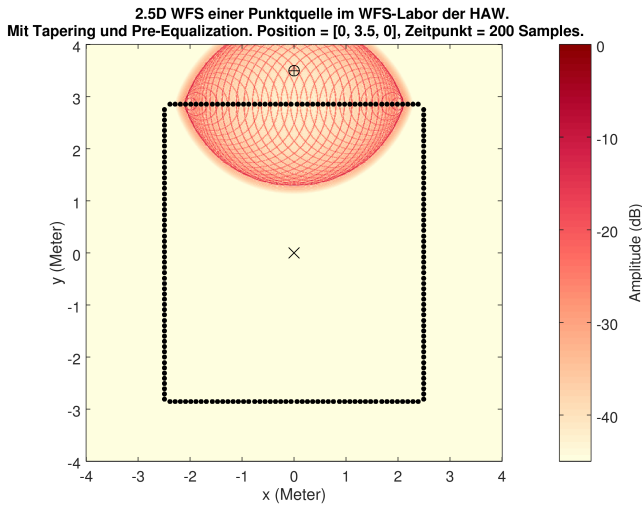


Abbildung 4. Simulation von 2.5D-WFS für eine Punktquelle in der Zeitdomäne

2.3. Fokussierte Punktquelle

Eine fokussierte Punktquelle ist eine virtuelle Klangquelle, die sich innerhalb des Zuhörerraumes befindet und vom Zuhörer ebenfalls zu einem gewissen Grad umgangen werden kann, solange sich dieser nicht zwischen die Position der fokussierten Klangquelle und die aktiven Lautsprecher stellt. Im Gegensatz zur herkömmlichen Punktquelle besitzt die fokussierte Punktquelle eine Richtung, in die sich das Schallfeld innerhalb des Zuhörerraumes ausbreiten soll. Zusätzlich werden die Lautsprecher in einer zeitlich invertierten Reihenfolge angesteuert. Das heißt, dass die Schallwellen sich zuerst konvergent in Richtung der virtuellen Klangquelle bewegen und nach Erreichen der Position wieder divergieren. Dadurch wird eine kreisförmige Ausbreitung der Schallwellen erreicht (siehe Abbildung 5). Aufgrund des invertierten Zeitkoeffizienten wird für die Fokussierung innerhalb des Zuhörerraumes eine negative Verzögerung des Audiosignals erwartet. Da dieses nicht möglich ist, wird

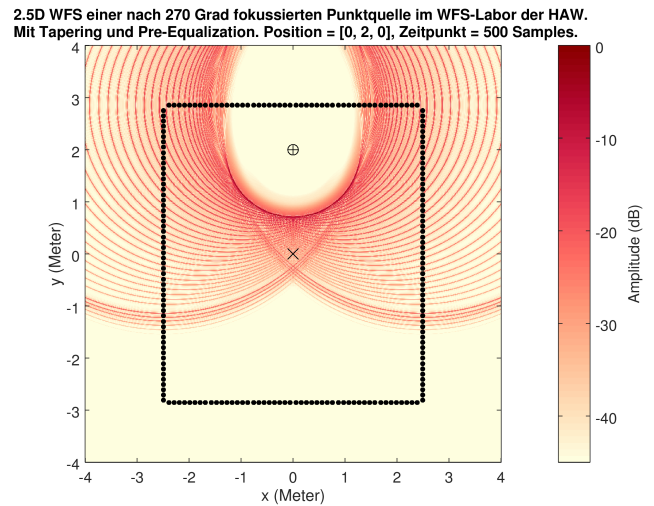


Abbildung 5. Simulation von 2.5D-WFS für eine fokussierte Punktquelle in der Zeitdomäne

3. Problemstellungen der Wellenfeldsynthese

Da die Wellenfeldsynthese nur ein Approximationsverfahren zur Rekonstruktion von echten Schallfeldern darstellt, kommt es durch die Diskretisierung und Vereinfachung bei der Herleitung der Ansteuerungsfunktionen zu Einschränkungen und Fehlern in der Darstellung des synthetisierten Schallfeldes.

3.1. Amplitudenfehler

Die Amplitude des synthetisierten Schallfeldes gilt bei der Ansteuerungsfunktionen für die 2.5D-WFS nur für einen bestimmten Punkt x_{ref} innerhalb des Zuhörerraumes [3]. Damit ein Zuhörer sich frei im Zuhörerraum bewegen kann und dabei gleichzeitig die korrekte Amplitude wahrnimmt, muss seine Position mit einem Tracking-System verfolgt werden. Der Referenzpunkt x_{ref} entspricht dabei stets der

aktuellen Position des Zuhörers, für die eine korrekte Amplitude berechnet wird [4].

3.2. Truncation-Effekt

Der Truncation-Effekt ist ein Beugungseffekt, der an den Enden von Lautsprecher-Arrays auftritt. Begründet liegt er darin, dass bei den WFS-Berechnungen nicht von einer endlichen Anzahl an Lautsprechern ausgegangen wird. Die fehlenden Lautsprecher an den Enden sorgen dafür, dass die Endlautsprecher Schattenwellen erzeugen, die mit den Schallwellen benachbarter Lautsprecher interferieren [5] (siehe Abbildung 6). Um diesem Verhalten entgegenzuwirken wird Tapering eingesetzt (siehe Abbildung 7). Dabei handelt es sich um die Nutzung einer Fensterfunktion, die den Amplitudenverlauf des Anfangs und des Endes eines Lautsprecher-Arrays dämpft (siehe Abbildung 8).

2.5D WFS einer nach 90 Grad fokussierten Punktquelle mit einem 32 Lautsprecher Array. Ohne Tapering. Position = [0, 1, 0], Frequenz = 800Hz.

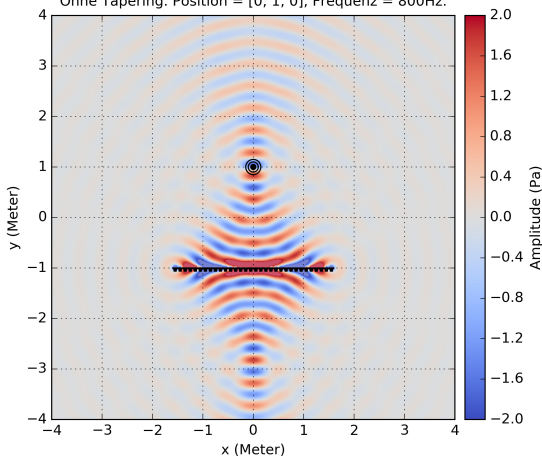


Abbildung 6. 2.5D-WFS eines Arrays und einer fokussierten Punktquelle ohne Tapering

3.3. Spatial-Aliasing

Das Spatial-Aliasing wird durch die Diskretisierung auf eine endliche Anzahl von Lautsprechern hervorgerufen, die stets einen gewissen Mindestabstand aufweisen. Dadurch können bei der Rekonstruktion von Schallfeldern nicht alle Frequenzen fehlerfrei synthetisiert werden [3] (siehe Abbildung 9 und Abbildung 10). Die Frequenz, ab welcher es bei einer WFS-Anlage zu Sampling-Fehlern kommt, berechnet sich mittels Gleichung 4.

$$f_{alias} = \frac{c}{2\Delta x_0} \quad (4)$$

3.4. Implementierungsabhängige Fehlerquellen

Bei der Umsetzung einer Software-Implementierung für die 2.5D-Wellenfeldsynthese müssen bezüglich der Ansteuerfunktionen einige Dinge beachtet werden:

2.5D WFS einer nach 90 Grad fokussierten Punktquelle mit einem 32 Lautsprecher Array. Mit Tapering (Tukey $\alpha=0.3$). Position = [0, 1, 0], Frequenz = 800Hz.

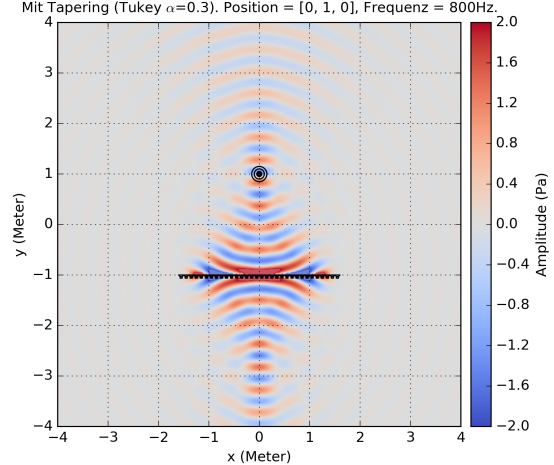


Abbildung 7. 2.5D-WFS eines Arrays und einer fokussierten Punktquelle mit Tapering

Fenster-Funktionen für das Tapering eines Arrays bestehend aus 32 Lautsprechern.

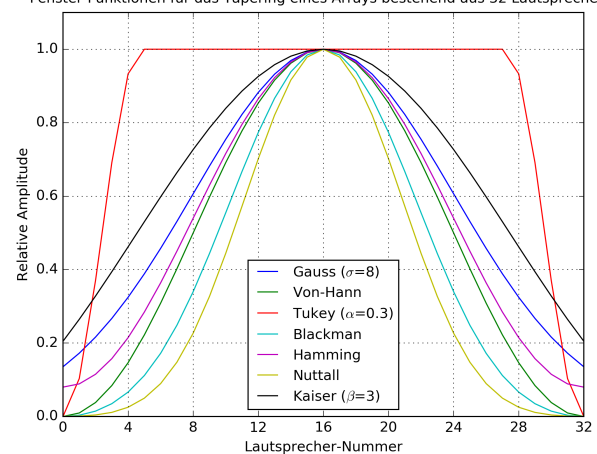


Abbildung 8. Mögliche Fensterfunktionen für den Einsatz von Tapering

- Anzahl Lautsprecher \times Anzahl Klangquellen = Maximale Anzahl von Berechnungen
- Eine Division durch Null ist möglich, weil Brüche involviert sind
- Beim Durchtritt einer normalen Punktquelle durch die Lautsprecher wird diese zu einer fokussierten Punktquelle und umgekehrt. In diesem Moment muss sich auch die jeweilige Ansteuerungsfunktion ändern
- Kommt eine virtuelle Punktquelle der Position eines Lautsprechers sehr nahe, dann verändert sich die Amplitude sehr stark. Eine Möglichkeit besteht darin einen Bereich vor und hinter dem Lautsprecher zu definieren, ab dem eine variable Fensterfunktion die Amplitude approximiert oder dämpft, um dadurch starke Amplitudenschwankungen bei den Übergängen zu vermeiden [6, Kapitel 3.4.1]

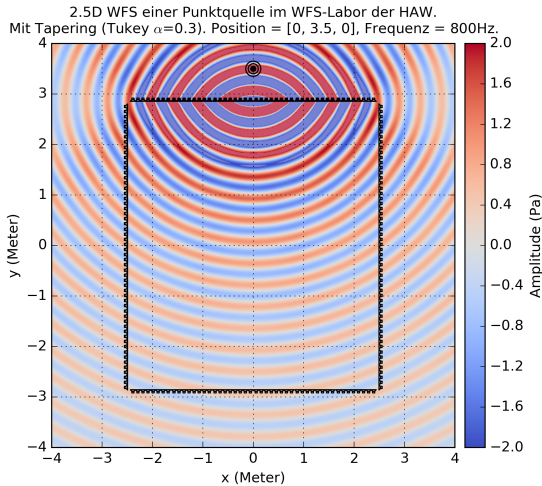


Abbildung 9. Kein Spatial-Aliasing unterhalb $f_{alias} = 1715\text{Hz}$ und $x_0 = 0.1\text{m}$

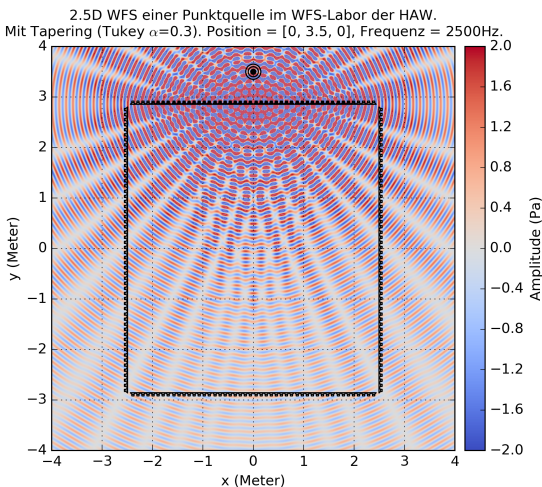


Abbildung 10. Deutliches Spatial-Aliasing oberhalb $f_{alias} = 1715\text{Hz}$ und $x_0 = 0.1\text{m}$

- Sollen fokussierte Klangquellen innerhalb des Zuhörerraumes berechnet werden können, müssen alle Audiosignale mit Hilfe einer Delay Line künstlich um die Dauer der negativen Latenz vorverzögert werden, da die Ansteuerungsfunktion für fokussierte Punktquellen einen negativen Zeitkoeffizienten aufweist (siehe Gleichung 3 und Abbildung 11)

3.5. Berechnungsaufwand

Die Berechnungen für WFS werden mit einer zunehmenden Zahl von virtuellen Klangquellen und Lautsprecherkanälen immer aufwendiger. Gleichzeitig sollen diese aber möglichst in Echtzeit abgearbeitet werden. Bei der Implementierung einer Software für WFS-Berechnungen kann

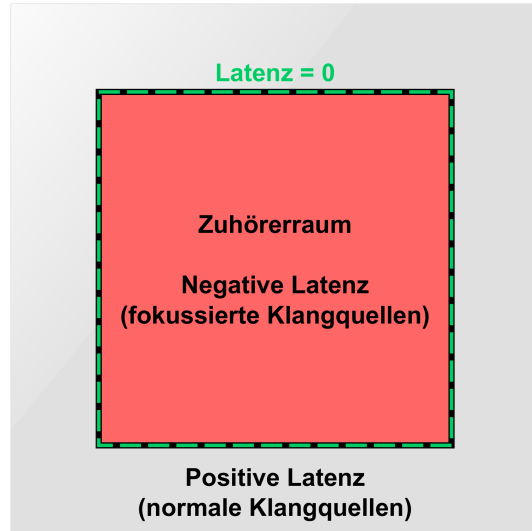


Abbildung 11. Latenz unterschiedlicher Klangquellen im Zuhörerraum

auf unterschiedliche Art und Weise eine Beschleunigung erreicht werden:

- Programmierung mit Multithreading, Vectorizing oder der Einsatz von Bibliotheken wie OpenMP⁵
- Programmierung unter Nutzung der GPU [7], welche aus mehreren Tausend Cores bestehen kann
- Hardwarenahe Programmierung eines DSPs und Nutzung von FPGAs [8]

4. Funktionsprinzip von VBAP

Das Vector Base Amplitude Panning (VBAP) bildet eine alternative Möglichkeit mit welcher 3D-Audio wiedergegeben werden kann. Dabei kommt es mit weniger Lautsprechern aus und hat einen geringeren Rechenaufwand als die Wellenfeldsynthese. Jedoch synthetisiert VBAP keine virtuellen Klangquellen, sondern gibt die Klänge richtungsgebunden am Ort des nächstliegenden Lautsprechers aus (siehe Abbildung 12). Damit macht es sich das Phänomen der Phantomschallquelle zunutze, welches auch beim Stereopanning zum Einsatz kommt. Dabei wird zwischen zwei oder mehr Lautsprechern die Lautstärke so angepasst, dass beim Zuhörer der Eindruck erweckt wird, der Klang würde zwischen den Lautsprechern wiedergegeben.

Trotz des geringeren Rechenaufwandes von VBAP kann dieser weiter minimiert werden. Ein interessanter Ansatz, der aus dem Bereich der Computergrafik stammt, ist die Nutzung einer konvexen Hülle aus Dreiecken, deren Eckpunkte Lautsprecherpositionen repräsentieren. Mit dieser lässt sich die Position der Quelle als Kollisionspunkt zwischen einem Strahl (*englisch Ray*), der der Klangquelle entspringt und dem davon getroffenen Dreieck darstellen (siehe Abbildung 13⁶).

5. <http://www.openmp.org>

6. <http://research.spa.aalto.fi/projects/vbap-lib/vbap.html>

Die Lautstärken der drei benachbarten Lautsprecher werden entsprechend den baryzentrischen Koordinaten des getroffenen Dreiecks berechnet. Bei der Kollisionsdetektierung in der Computergrafik werden solche Algorithmen mit Raytracing oder Raycasting bezeichnet.

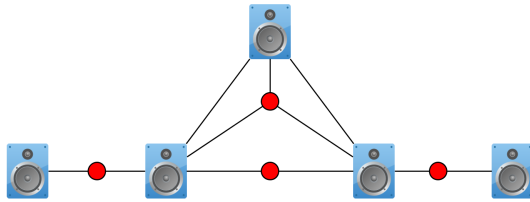


Abbildung 12. VBAP Quellenpositionen in 2D und 3D

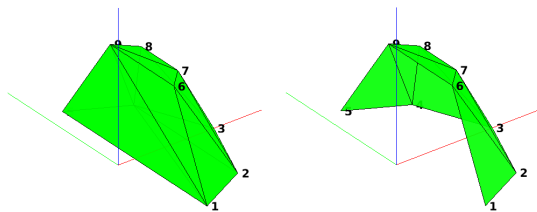


Abbildung 13. Konvexe Hülle des 3D VBAP Lautsprecher-Setups

5. Bestehende Software-Lösungen

Im Bereich der Open Source Software gibt es mehrere Programme, die in der Lage sind WFS und VBAP zu rendern. Dazu zählen unter anderem WONDER [6], das seit 2011 bei der WFS-Anlage im Labor für interaktive immersive Audiosysteme (I²AudioLab)⁷ der Hochschule für Angewandte Wissenschaften Hamburg (HAW) eingesetzt wird, SuperCollider⁸, welches auch im Creative Space for Technical Innovations (CSTI)⁹ der HAW benutzt wird sowie SoundScape Renderer¹⁰ und wfs-designer [9]. Zusätzlich wird an der HAW an CoRGII [10], einem Nachfolger für WONDER gearbeitet.

Fast alle Programme haben eines gemeinsam: Die Trennung zwischen dem Rendering und der Ansteuerung. Dadurch kann das aufwendige Berechnen der Audiosignale auf einen Computer oder Cluster ausgelagert werden, während die Steuerung der Klangquellen über einen anderen Computer erfolgen kann. Dabei wird der Nachrichtenaustausch zum großen Teil über das Open Sound Control Protokoll

(OSC)¹¹ abgewickelt. Diese Trennung ist wichtig, denn dadurch kann die Steuerung eines WFS oder VBAP Soundsystems in die Hand eines oder mehrerer Teilnehmer/Zuhörer gelegt werden, die zum Beispiel eine Applikation auf ihrem Smartphone nutzen können, um damit OSC-Nachrichten mit Aktionen oder ihrer aktuellen Position zu verschicken. Auf diese Weise entsteht eine Interaktion zwischen dem Benutzer und dem Soundsystem.

6. Immersive Systeme

Immersive Systeme können in mehrere Kategorien unterteilt werden. Dabei werden folgende Unterscheidungen getroffen:

- **Augmented Reality:** Systeme, welche die echte Welt benutzen, um auf dieser einige zusätzliche Informationen einzublenden
- **Mixed Reality:** Systeme, welche nur Teile der echten Welt benutzen, um darauf eine Menge an zusätzlichen Informationen einzublenden
- **Virtual Reality:** Systeme, welche nur noch eine virtuelle Welt darstellen und den Nutzer, mehr oder weniger, von der echten Welt abschotten

Zusätzlich kann eine Unterscheidung zwischen den Tracking-Methoden vorgenommen werden. So sind die meisten Virtual Reality Systeme abhängig von externer Tracking-Peripherie, wie die HTC Vive von den Lighthouse genannten Trackern, oder die Oculus Rift von zwei oder mehr Kameras. Die Tracker weisen im von ihnen erfassten Bereich eine relativ hohe Positionsgenauigkeit auf.

Die Microsoft HoloLens¹² stellt ein Mixed Reality Headset dar, welches mobil ist und keinen zusätzlichen Computer benötigt. Sie besitzt einen Visor, durch welchen man hindurchschauen und die echte Welt sehen kann. Gleichzeitig dient er als Projektionsfläche für Informationen. Zusätzlich zu einem Gyroskop und Magnetometer besitzt die HoloLens auch einen Tiefensensor, ähnlich der Microsoft Kinect¹³. Damit lassen sich Tiefeninformationen von Räumen gewinnen, welche kartografiert und später auch wiedererkannt werden können. Aus diesem Grund benötigt sie keine externen Tracking-Mechanismen und beherrscht somit das *Inside-Out Tracking*. Angemerkt sei, dass Microsoft auch geschlossene Headsets als Mixed Reality Headsets anbietet. Diese weisen zwar einen Sensor ähnlich dem der Kinect auf und beherrschen damit das Inside-Out Tracking ohne externe Tracking-Peripherie, zählen aufgrund der geschlossenen Bauweise aber eigentlich zur Familie der Virtual Reality Headsets.

Augmented Reality Systeme wie die Google Glass¹⁴ blenden in Form eines Brillenzusatzes nur einige wenige Informationen ein. Integrierte Sensoren, Kamera und

7. <https://i2audiolab.de>
 8. <https://supercollider.github.io>
 9. <https://csti.haw-hamburg.de>
 10. <http://spatialaudio.net/ssr>

11. <http://opensoundcontrol.org>
 12. <https://www.microsoft.com/de-de/hololens>
 13. <https://developer.microsoft.com/de-de/windows/kinect>
 14. <https://www.x.company/glass>

Spracheingabe sorgen auch hier für ein, wenn auch eingeschränktes, Inside-Out Tracking, wodurch kein externes Tracking-System am Einsatzort notwendig ist.

Im Bereich der Bilderkennung gibt es seit längerer Zeit mit Hilfe von Kameras und QR-Codes die Möglichkeit Augmented Reality einzusetzen. Aktuelle Frameworks zur Bilderkennung können aber noch viel mehr und sind bereits in der Lage spezifische Flächen, Lichteinfall, Abstände, Personen, Orte und alltägliche Gegenstände zu erkennen und sind auch auf Smartphones oder Geräten des Internet-of-Things (IoT) lauffähig.

7. Bedienkonzepte

Bei der Interaktion mit einem Soundsystem bieten die Headsets aller Kategorien einen Mehrwert, weil sie die Wahrnehmung und den Informationsaustausch um den Sehsinn erweitern. Erste Versuche zur Bedienung einer WFS-Anlage mittels Augmented Reality gab es bereits vor über 10 Jahren [11]. Im I²AudioLab der HAW wurde anfangs ein kamerabasiertes Tracking-System mit Tracking-Targets für eine Gestensteuerung benutzt [12]. Später wurde mit Hilfe des Vuforia Frameworks zur Bilderkennung¹⁵ eine Applikation für ein Android-Tablet entwickelt, welche eine Navigation mittels Inside-Out Tracking im Labor und die Verwaltung von virtuellen Klangquellen beherrschte [13]. Aktuell bietet sowohl das I²AudioLab, als auch das CSTI neben der Benutzung des sehr präzisen, auf sechs Kameras basierten ART-Tracking-Systems¹⁶ auch die Nutzung einer Microsoft Kinect und das Tracking der HTC Vive, der Oculus Rift und der Microsoft HoloLens an. Mittels Leap Motion Controller¹⁷ kann das Tracking um eine Hand- und Fingererkennung erweitert werden. Damit decken die Labore der HAW alle Kategorien immersiver Systeme ab und ermöglichen das Konzipieren und Entwickeln zahlreicher Anwendungen.

8. Mobile Systeme

Der größte Wandel im Bereich immersiver Systeme zeichnet sich im Moment bei den Smartphones ab. Zwar gibt es tragbare und auch leistungsstarke Virtual und Mixed Reality Peripherie wie die Microsoft HoloLens oder spezielle Notebook-Rucksäcke mit großen Akkus für die kabellose Nutzung einer HTC Vive oder Oculus Rift, dennoch bleibt nichts davon so mobil und vielseitig einsetzbar wie ein Smartphone. Das ist darauf zurückzuführen, dass Smartphones, sowohl was die Pixeldichte und Größe der Displays, als auch was die Prozessorleistung betrifft, in den letzten Jahren stetig verbessert wurden. Aktuelle mobile Prozessoren wie der Qualcomm Snapdragon 835¹⁸ besitzen bereits alle Sensoren an Bord, um Inside-Out Tracking zu ermöglichen und als Standalone-Headset für Virtual Reality benutzt zu

werden. Zudem macht das Vorhandensein einer Kamera schon lange Augmented Reality mittels Bilderkennung möglich, zum Beispiel in Form von QR-Codes. Gleichzeitig können viele Smartphones mittels speziellen Halterungen zu Virtual Reality Brillen umfunktioniert werden. Mit einem Stück Pappe angefangen wurde aus Google Cardboard¹⁹ das heutige Google Daydream²⁰. Das mehrjährige Google *Project Tango*²¹, welches dazu geschaffen wurde Bilderkennung und Tiefenkameras in Smartphones zu etablieren und damit sehr genaues Inside-Out Tracking und bessere Augmented Reality Erlebnisse zu bieten, führte letztlich nur zu zwei relativ teuren Smartphone-Modellen, die ein Nischendasein fristen. Die Mehrkosten für eine Tiefenkamera und der höhere Akkuverbrauch waren für die meisten Hersteller ein Grund nicht in diese Technologie zu investieren.

Bewegung kommt nun jedoch durch Google Konkurrent Apple ins Spiel. In seiner kommenden iPhone-Generation und der damit erscheinenden Version 11 des mobilen Betriebssystems iOS, will Apple das ARKit²² einführen und zu einem Bestandteil des Betriebssystems machen. Wenige Tage vor dem Erscheinen hat Google nun sein Pendant ARCore²³ vorgestellt. Basierend auf den Erkenntnissen von Project Tango, welches damit sehr wahrscheinlich nicht mehr weiterentwickelt wird, hat Google eine neue Schnittstelle für Augmented Reality Inhalte zur Verfügung gestellt. Der einzige Unterschied zu Project Tango ist, dass es bereits mit den integrierten Sensoren und der Kamera eines Smartphones funktioniert und nicht auf einen Tiefensensor angewiesen ist. Damit läuft Googles ARCore nicht nur auf zwei Modellen, wie Project Tango, sondern auf tausenden von Android basierten Smartphone-Modellen, mit dem selbst gesteckten Ziel bis zum Ende des Jahres 2017 auf 100 Millionen Smartphones verfügbar zu sein.

Zwar mag es zuerst wie ein Wettlauf erscheinen, bei dem es darum geht welcher Konzern als erstes die beste Augmented Reality Applikation veröffentlicht. Hinter den Kulissen sind aber bereits jetzt Anbindungen für die beliebtesten Grafik-Engines Unity3D²⁴ und Unreal²⁵ vorhanden. Das heißt, dass nicht nur ein großes Potential hinter der Technologie steckt, sondern auch große Anstrengungen unternommen werden und enorm viel Geld investiert wird, damit alles so schnell wie möglich bei den Kunden landet und von diesen angenommen wird. Ein Vorgeschmack könnte der Überraschungserfolg Pokémon GO²⁶ sein, der mit seinem Augmented Reality Spielprinzip zu einer regelrechten Massenbewegung wurde.

15. <https://www.vuforia.com>

16. www.ar-tracking.com

17. <https://www.leapmotion.com>

18. <https://www.qualcomm.com/products/snapdragon/processors/835>

19. <https://vr.google.com/cardboard>

20. <https://vr.google.com/daydream>

21. <https://get.google.com/tango>

22. <https://developer.apple.com/arkit>

23. <https://developers.google.com/ar>

24. <https://unity3d.com>

25. <https://www.unrealengine.com>

26. <http://pokemongo.nianticlabs.com>

9. Fazit

Mit der zunehmenden Verbreitung von Augmented, Mixed und Virtual Reality Geräten macht es Sinn sich für einen immersiven, audio-visuellen Gesamteindruck auch näher mit dem Thema 3D-Audio zu beschäftigen. Die Möglichkeit der Wellenfeldsynthese virtuelle Klangquellen nicht nur außerhalb, sondern auch innerhalb von Räumen zu synthetisieren, ermöglicht viele neue Szenarien der Interaktion zwischen dem Benutzer, den Räumlichkeiten, dem Gesehenen und dem Sound. Der Preis dafür ist die kostenintensive Installation sehr vieler Lautsprecher, die eine Grundvoraussetzung der Wellenfeldsynthese darstellen. Bei der Realisierung spielen zudem noch weitere Faktoren wie die Raumakustik und der Rechenaufwand eine Rolle. Letzterer kann bei der Implementierung mit Bedacht verringert werden, indem die eingesetzte Software die Rechenlast effektiv in einem Verteilten System auslagert, gut parallelisiert oder auf spezifischen Prozessoren wie der Grafikkarte oder auf einem FPGA ausgeführt wird.

Als Alternative wurde das VBAP 3D-Audio Rendering-Verfahren vorgestellt. Dieses kommt mit weniger Lautsprechern aus und benötigt nicht soviel Rechenleistung wie die Wellenfeldsynthese. Eine mögliche Optimierung der Berechnungen wurde anhand des Raytracings vorgestellt, das aus dem Bereich der Computergrafik stammt. Als Konsequenz der geringeren Anforderungen erlaubt dieses Rendering-Verfahren jedoch nur eine richtungsgebundene 3D-Audio-Wiedergabe.

Beim Aufbau des Großteils der verfügbaren Software fällt auf, dass eine Unterstützung für den Nachrichtenaustausch, meist via OSC-Protokoll, von großer Wichtigkeit ist. Da das Protokoll plattformunabhängig ist, kann es relativ leicht auf allen möglichen Geräten eingesetzt werden. Ein weiterer Vorteil ist, dass auf diese Weise bidirektionale Interaktionsmöglichkeiten zwischen dem Audiosystem und einem oder mehreren Zuhörern entstehen. Das lineare Komponieren und Abspielen eines Stückes kann dadurch aufgebrochen werden. Die Grenzen zwischen der Rolle des Zuhörers und des Komponisten weichen auf, da jeder Beteiligte durch seine Interaktionen an der Entstehung des Stückes teilhaben und diese beeinflussen kann.

Zuletzt wurde die rasante Entwicklung und der Stand von aktuellen Smartphones betrachtet. Deutlich zu erkennen ist dabei, dass die Hersteller von Software im Moment große Hoffnungen auf den Bereich Augmented Reality legen und viel dafür tun, damit dieser vorangetrieben wird. Gleichzeitig liefern die Hersteller von Hardware immer leistungsstärkere mobile Prozessoren ab und erweitern diese um eine Vielzahl von Sensoren, so dass Features wie das Inside-Out Tracking bereits möglich sind und ein aktuelles Smartphone mit dem nötigen Zubehör sowohl Augmented, als auch Virtual Reality fähig ist. Aufgrund der Tatsache, dass ein Smartphone die Menschen heutzutage auf Schritt und Tritt begleitet, kann man davon ausgehen, dass die Interaktionsmöglichkeiten mit der Umgebung mittels Augmented Reality zukünftig weiter zunehmen und vielleicht auch neue Formen annehmen werden.

10. Ausblick

Damit sich in Zukunft ein Audio-Verfahren wie die Wellenfeldsynthese durchsetzen kann, muss sich ein allgemeines Verständnis über die Möglichkeiten dahinter durchsetzen. Das geht nur, wenn es auch genug Unterstützer gibt. Vorrangig Hersteller von Lautsprechern, aber auch Klanginstallationen, die der Allgemeinheit zur Verfügung stehen und den Besuchern das Verfahren näher bringen. Aktuell gibt es nur eine handvoll Hersteller von WFS-Systemen, z.B. FourAudio²⁷ und Holoplot²⁸ aus Berlin. Zu den hohen Preisen der Hardware kommt noch die Tatsache dazu, dass es keine wirkliche Tonspur, Kodierung oder Medien für die Wellenfeldsynthese gibt. Aktuelle Kodierungen basieren auf einzelnen Kanälen, während die Wellenfeldsynthese Objektdaten, also Koordinaten der Klangquellen benötigt. Ein Anfang zum Speichern und Wiedergeben von Projekten ist das Format *SpatDIF*²⁹.

Mit Blick auf den Einsatz von Klanginstallationen unter Verwendung von Audio-Rendering-Software, die einen Nachrichtenaustausch unterstützt, zum Beispiel über OSC, lässt sich aufgrund der hohen Verfügbarkeit von Smartphones und Applikationen sicher ein Großteil der Besucher dazu animieren das System zu benutzen oder gemeinschaftlich bei Installationen oder Events mitzumachen. Ausstellungen und Museen bekämen durch das freie Positionieren mehrerer Klangquellen im Raum die Möglichkeit bestimmte Ereignisse akustisch nachzubauen und zusätzlich begehbar zu machen.

Für den Einsatz in Augmented oder Virtual Reality Umgebungen wäre die Wellenfeldsynthese eine ideale klangliche Repräsentation der visuell wahrgenommenen Informationen.

Besonders spannend bleibt die Zukunft der Nutzung des Smartphones als Augmented oder Virtual Reality Gerät. Konzerne wie Apple und Google müssen überzeugende, aber auch sinnvolle Anwendungsmöglichkeiten und immersive Erlebnisse mit ihren Systemen liefern, damit das ganze nicht wie bisher eher als eine bloße Spielerei im Sande verläuft.

Literatur

- [1] A. J. Berkhout, „A holographic approach to acoustic control,” *Journal of the Audio Engineering Society*, Vol. 36, Nr. 12, S. 977–995, 1988.
- [2] E. W. Start, „Direct Sound Enhancement by Wave Field Synthesis,” PhD-Thesis, Technische Universiteit Delft, 1997.
- [3] E. Verheijen, „Sound Reproduction by Wave Field Synthesis,” PhD-Thesis, Technische Universiteit Delft, 1997.
- [4] F. Melchior, C. Sladeczek, D. de Vries, und B. Fröhlich, „User-Dependent Optimization of Wave Field Synthesis Reproduction for Directive Sound Fields,” in *124th Convention of the Audio Engineering Society*, 2008.
- [5] P. Vogel, „Application of Wave Field Synthesis in room acoustics,” PhD-Thesis, Technische Universiteit Delft, 1993.

27. <http://www.fouraudio.com>

28. <http://www.holoplot.com>

29. <http://www.spatdif.org>

- [6] M. A. J. Baalman, *On Wave Field Synthesis and electro-acoustic music, with a particular focus on the reproduction of arbitrarily shaped sound sources.* VDM, 2008.
- [7] J. A. Belloch, A. Gonzalez, E. S. Quintana-Ortí, M. Ferrer, und V. Välimäki, „GPU-Based Dynamic Wave Field Synthesis Using Fractional Delay Filters and Room Compensation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 25, Nr. 2, S. 435–447, Feb 2017.
- [8] D. Theodoropoulos, G. Kuzmanov, und G. Gaydadjiev, „Multi-Core Platforms for Beamforming and Wave Field Synthesis,” *IEEE Transactions on Multimedia*, Vol. 13, Nr. 2, S. 235–245, April 2011.
- [9] M. N. Montag, „Wave Field Synthesis In Three Dimensions By Multiple Line Arrays,” Master-Thesis, University of Miami, 2011.
- [10] E. Winter, „Analyse und Reimplementierung einer plattformunabhängigen Software für das 3D-Audio-Rendering,” Bachelor-Thesis, Hochschule für Angewandte Wissenschaften Hamburg, 2016.
- [11] F. Melchior, T. Laubach, und D. de Vries, „Authoring and User Interaction for the Production of Wave Field Synthesis Content in an Augmented Reality System,” in *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, Serie ISMAR '05. Washington, DC, USA: IEEE Computer Society, 2005, S. 48–51. [Online]. Verfügbar: <http://dx.doi.org/10.1109/ISMAR.2005.20>
- [12] M. Nogalski, „Gestengesteuerte Positionierung von Klangquellen einer Wellenfeldsynthese-Anlage mit Hilfe eines kamerabasierten 3D-Tracking-Systems,” Bachelor-Thesis, Hochschule für Angewandte Wissenschaften Hamburg, 2012.
- [13] S. Jandt, „Visualisierung und Steuerung von Wellenfeldsynthese-Quellen mit Android-Geräten,” Bachelor-Thesis, Hochschule für Angewandte Wissenschaften Hamburg, 2014.