# Klangverfärbung in der Wellenfeldsynthese - Experimente und Modellierung

Hagen Wierstorf, Christoph Ende, Alexander Raake
*Email: hagen.wierstorf@tu-berlin.de*
*Assessment of IP-based Applications, Technische Universität Berlin, 10587 Berlin, Deutschland*

## Einleitung

In der Wellenfeldsynthese wird versucht ein vordefiniertes Schallfeld in einem ausgedehntem Zuhörerbereich zu synthetisieren. Dazu ist der Zuhörerbereich von einer Reihe von Lautsprechern umschlossen, die mit entsprechenden Signalen angesteuert werden. Auf Grund der endlichen Anzahl an Lautsprechern stellen diese eine räumliche Abtastung dar und das wiedergegebene Schallfeld ist nur bis zu einer bestimmten Frequenz mit dem vordefinierten identisch. Oberhalb dieser sogenannten Aliasfrequenz kommt es zu Aliasing im Schallfeld. Da die Abtastung im Gegensatz zum Zeitbereich in zwei bis drei Dimensionen anstatt in einer stattfindet, ist die Grenzfrequenz fürs Aliasing zudem von der Zuhörerposition abhängig und die Ausgeprägtheit des Aliasing kann sich im Zuhörerbereich ändern. Für die Aliasfrequenz kann eine untere Grenze abgeschätzt werden und zwar mit

$$F_{\text{alias}} = \frac{c}{2\Delta x_0}\,, \tag{1}$$

wobei $c$ die Lichtgeschwindigkeit und $\Delta x_0$ den Lautsprecherabstand angibt.

Neben der Zuhörerposition hängt die Aliasfrequenz von dem Abstand der einzelnen Lautsprecher zueinander ab. Im Folgenden wird eine zirkuläre Lautsprechergruppe mit einem Durchmesser von 3 m betrachtet. Die Lautsprechergruppe wird anschließend mit der Wellenfeldsynthese angesteuert, so dass eine Punktquelle synthetisiert wird, die sich am Punkt $x_\text{s} = (0, 2.5, 0)$ m befindet. Abbildung 1 zeigt den Frequenzgang der Wellenfeldsynthese-Systeme mit unterschiedlichen Lautsprecherabständen. Um eine systematische Untersuchung durchführen zu können wurden auch Lautsprecherabstände betrachtet, die sich zur Zeit nicht in der Realität realisieren ließen. Zusätzlich ist der Frequenzgang eines Stereophonie-Systems dargestellt, welches an der gleichen Position eine Punktquelle durch Panning versucht zu realisieren. Der Frequenzgang ist nicht an der Position $(0, 0, 0)$ m simuliert worden, sondern ungefähr an der Stelle des linken Ohres des Zuhörer bei $(-0.1, 0, 0)$ m.

Es ist zu erkennen, dass der Frequenzgang für die Wellenfeldsynthese ab der Aliasfrequenz kammfilterartige Strukturen aufweist, wobei die Aliasfrequenz umso höher ist, je kleiner der Lautsprecherabstand. Für niedrige Lautsprecherabstände von 0.5 cm und 0.3 cm sind für hohe Frequenzen immer noch Abweichungen ersichtlich, auch wenn die untere Grenze der Aliasfrequenz hier oberhalb von 20 kHz liegt – berechnet mit (1). Diese kleinen Abweichungen bei hohen Frequenzen sind durch die Verwendung von ganzzahliger zeitlicher Abtastung in der

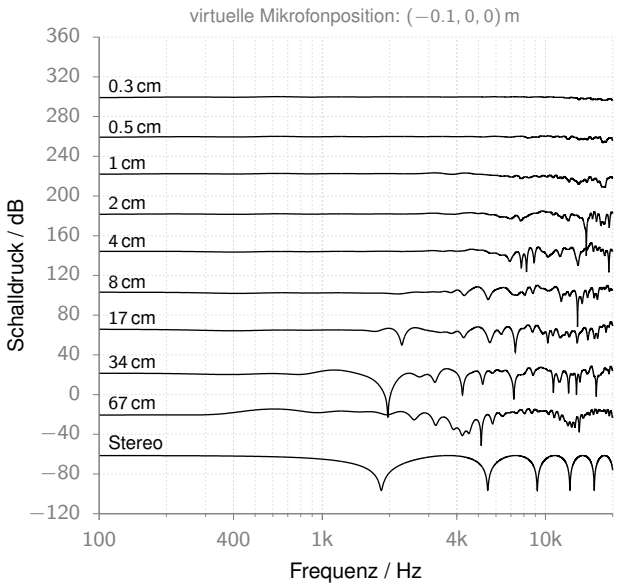


**Abbildung 1:** Frequenzgänge für Wellenfeldsynthese mit unterschiedlichen Lautsprecherabständen und für Stereophonie. Die jeweiligen Lautsprecherabstände sind über dem dazugehörigem Frequenzgang angegeben. Die Frequenzgänge wurden am Ort des linken Ohres des Zuhörers simuliert. Die Frequenzgänge sind in ihrem Absolutwert versetzt, um sie besser vergleichbar zu machen.

Wellenfeldsynthese zu erklären. In einer nachfolgenden Studie wird versucht diese Abweichungen durch die Verwendung von nicht-ganzzahliger zeitlicher Abtastung zu vermeiden. Neben den betrachteten Systemen für die Wellenfeldsynthese, weist auch das Stereophonie-System Abweichungen im Frequenzgang auf, da auch in diesem Fall eine räumliche Abtastung stattfindet, die nur direkt im Zentrum $(0, 0, 0)$ m einen glatten Frequenzgang ermöglicht.

Veränderungen des Frequenzgangs eines Systems führen in der Regel zur Wahrnehmung von einer Änderung in der Klangfarbe, was zum Beispiel für die Wellenfeldsynthese bereits von Helmut Wittek nachgewiesen wurde [1]. In seinem Versuch hat er jedoch nur eine geringe Anzahl an unterschiedlichen Lautsprecherabständen verwendet, was insbesondere im Hinblick auf die Entwicklung eines auditorischen Modelles für die Vorhersage der Klangverfärbung eine Einschränkung darstellt. Um dieses Problem zu umgehen und einen größeren Raum an unterschiedlichen Systemen aufzuspannen, wurde ein Experiment mit den in Abbildung 1 vorgestellten Systemen durchgeführt. Anschließend wurde versucht die Ergebnisse mit einem auditorischen Modell vorherzusagen.

## Methode

Das Experiment wurde mit Hilfe von Binauralsynthese und nicht-individuellen kopfbezogenen Übertragungsfunktionen (HRTFs) durchgeführt. Eine genaue Beschreibung des verwendeten Systems findet sich in Kapitel 4 in Wierstorf [2]. Die Verwendung von nicht-individuellen HRTFs hat die Limitierung, dass diese bereits zu einer Abweichung im Frequenzgang führen. Diese ist jedoch für alle Systeme gleich und sollte daher keinen Einfluss auf die Ergebnisse des Experimentes haben. Jedoch ist die Abweichung des Frequenzgangs abhängig von der Richtung der HRTF und ändert sich mit dieser. Würde der dynamische Teil der Binauralsynthese aktiv sein, würde sich die wahrgenommene Klangfarbe bei Kopfdrehung die ganze Zeit über ändern. Um dies zu verhindern wurde der dynamische Teil der Binauralsynthese für dieses Experiment ausgeschaltet und kein Headtracking der Personen durchgeführt.

## Stimuli

Um nach einer Änderung in der Klangfarbe zu fragen, der Klangverfärbung, wurde eine Punktquelle an der Position $(0, 2.5, 0)$ m als Referenz gewählt, welche durch eine einzelne HRTF realisiert wurde. Zusätzlich wurde dieselbe Punktquelle durch die oben beschriebenen Wellenfeldsynthese-Systeme und dem Stereophonie-System realisiert. Die angegebenen Lautsprecherabstände korrespondieren dabei bei der verwendeten runden Lautsprechergruppe mit einem Durchmesser von 3 m Anzahlen von 12, 28, 56, 112, 224, 448, 896, 1792 und 3584 Lautsprechern. Für das Stereophonie-System wurden zwei einzelne Lautsprecher an den Positionen $(\pm 1.4, 2, 5, 0)$ m realisiert. Alle Impulsantworten wurden auf das absolute Maximum normalisiert, bevor sie mit dem entsprechendem Audiomaterial gefaltet wurden.

Als Audiomaterial wurde 100 s langes gepulstes Rosa-Rauschen mit einer Länge der einzelnen Pulse von 800 ms und einer Pause von 500 ms zwischen den einzelnen Pulsen verwendet. Die einzelnen Pulse wiesen zudem eine Fensterung am beginn und Ende von 50 ms auf. Die einzelnen Pulse waren jeweils neue Realisationen von Rosa-Rauschen. Ein ähnlicher Stimulus wurde ebenfalls in der Studie von Wittek [1] verwendet. Zusätzlich wurde ein acht Sekunden langer Sprachausschnitt eines weiblichen Sprechers verwendet.

## Durchführung

Die Versuchspersonen wurden gebeten die Klangverfärbung bezogen auf die Punktquelle zu bewerten, auf einer Skala mit den Endpunkten *keine* und *stark*. Dies wurde in einem MUSHRA Design durchgeführt, mit versteckter Referenz und einem Anker. Für den Anker wurde nicht die MUSHRA Vorgabe verwendet, sondern eine 5 kHz Hochpass gefilterte (Butterworth, 2. Ordnung) Version der Referenz erstellt. Die Versuchspersonen wurden instruiert die Klangverfärbung zu bewerten und dabei mögliche Unterschiede in der Lautheit und der Externalisation zu ignorieren. Vor dem eigentlichem Experiment, gab es jeweils einen Trainingsdurchlauf.

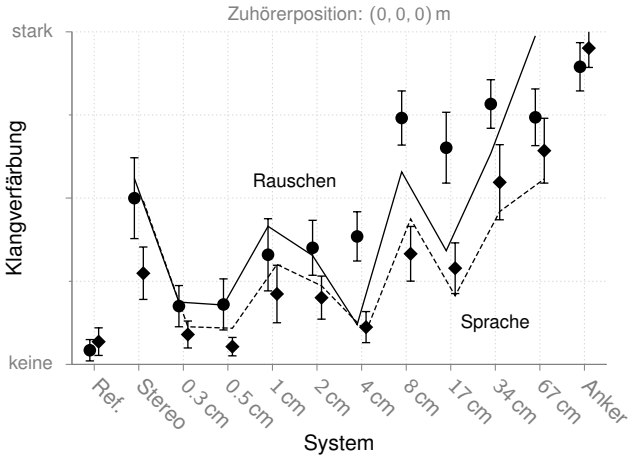


**Abbildung 2:** Die Punkte repräsentieren den Mittelwert der wahrgenommenen Klangverfärbung über alle Versuchspersonen mit Konfidenzinterval. Die Linien zeigen die Vorhersage der Modellierung. Kreise und die durchgezogene Linie gehören zu den Rausch-, die Diamanten und gestrichelte Linie zu den Sprachstimuli.

Während eines Durchganges mussten die Versuchspersonen alle 10 Konditionen, die versteckte Referenz und den Anker bewerten, jeweils für Rauschen und für die Sprache. Die Stimuli wurden dabei endlos wiederholt und zwischen den einzelnen Konditionen konnte instantan hin- und hergeschaltet werden. Die Messung pro Audiomaterial wurde jeweils zwei Mal durchgeführt.

## Versuchspersonen

15 Zuhörer nahmen an dem Experiment teil, sie waren zwischen 23 und 29 Jahren alt. Alle Teilnehmer waren nach eigener Aussage normalhörend. Sie wurden finanziell für ihre Teilnahme entschädigt.

## Ergebnisse

Abbildung 2 fasst die Ergebnisse für das Rauschen und die Sprache zusammen. Es ist die Stärke der Klangverfärbung in Abhängigkeit des verwendeten Systems dargestellt. Die Klangverfärbung ist dabei als Mittelwert über alle Versuchspersonen zusammen mit dem Konfidenzinterval aufgetragen. Es ist zu erkennen, dass die versteckte Referenz als nicht klangverfärbt und der Anker als stark klangverfärbt bewertet worden sind. Die einzelnen Wiedergabesysteme reihen sich dazwischen ein, wobei für die Sprachstimuli die beiden Wellenfeldsynthese-Systeme mit dem geringstem Lautsprecherabstand sich nicht signifikant von der Bewertung der versteckten Referenz unterscheiden. Die Sprache wurde generell als weniger klangverfärbt als das Rauschen bewertet, was sich durch ihre geringere Energie bei höheren Frequenzen erklären lässt. Die Wellenfeldsynthese-Systeme werden besonders ab einem Lautsprecherabstand von 8 cm als stark klangverfärbt bewertet. Dies ist interessant, da bei 8 cm ungefähr die Grenze liegt für die sich die Lautsprechergruppen noch durch Hardware realisieren lassen.

Die Ergebnisse für Stereophonie zeigen, dass auch hier die Punktquelle als klangverfärbt bezogen auf die Referenz wahrgenommen wird, jedoch nicht so stark wie bei dem Wellenfeldsynthese-System mit einem Lautsprecherabstand von 8 cm.

Weiterhin ist zu beachten, dass es sich bei der Klangfarbe um einen multi-dimensionalen Raum handelt. Gleiche Bewertungen in der Klangverfärbung bezogen auf eine Referenz, bedeuten damit nicht automatisch, dass die Systeme auch ähnlich klingen. Die Aussage ist lediglich, dass sie beide gleich weit entfernt von der Referenz eingeordnet werden. Dies ist von Interesse, da es dazu führen kann, dass die wahrgenommene Audioqualität für zwei in der Klangverfärbung gleich bewerteten Systeme sich durchaus unterscheiden kann. Daher kann aus den vorliegenden Ergebnissen nicht geschlossen werden, ob ein Wellenfeldsynthese-System mit einem geringem Lautsprecherabstand eine bessere Audioqualität aufweist als ein Stereophonie-System.

## Modellierung

Die Ergebnisse aus dem Hörversuch wurden versucht mit Hilfe eines auditorischen Modells zu erklären. Dabei gibt es in der Literatur schon eine Reihe von Ansätzen, die sich mit der Modellierung besonders der Detektion von Klangverfärbung beschäftigt haben. So ist hier das A0-Kriterium [3] zu nennen und im Rahmen von Stereophonie-Wiedergabe das Modell von Pulkki [4]. Weiterhin haben Moore und Tan [5] ein auditorisches Modell zur Vorhersage der Natürlichkeit für Stimuli, die einen Kammfilter aufwiesen entwickelt. Die von den Versuchspersonen und dem Modell bewertete Natürlichkeit dürfte in diesem Fall stark mit der Klangverfärbung korreliert sein.

Im Folgenden wird nun zunächst ein einfaches Modell beschrieben, welches auf dem von Pulkki beschriebenen basiert und dieses auf die Stimuli aus dem Hörversuch angewendet. Anschließend werden die Ergebnisse mit Vorhersagen des Modells nach Moore und Tan verglichen.

Abbildung 3 zeigt das Blockschaltbild des verwendeten auditorischen Modelles. Zur Vereinfachung ist jeweils nur ein Ohr dargestellt. Das binaurale Hören ist im Moment als Aufsummation der internen Spektren implementiert, wie dies auch Pulkki durchgeführt hat. Am Anfang wird das verwendete Audiomaterial jeweils mit der HRTF für die Referenz und das zu untersuchende System gefaltet. Anschließend durchlaufen die Signale eine Bandpassfilterung im Mittelohr und werden mit einer Gammatonfilterbank in mehrere Frequenzkanäle aufgespalten. Pro Frequenzkanal wird die Lautheitskompression per Potenzierung nach Zwicker angenähert und anschließend wird in jedem Kanal der RMS Wert in dB ermittelt und die Differenz zwischen Referenz- und Testsignalkanälen gebildet. Wird in einem Kanal ein Schwellwert von 1 dB überschritten, wird angenommen, dass die Versuchsperson eine Klangverfärbung detektieren kann [4]. Um nun eine Vorhersage der Stärke der Klangverfärbung zu erhalten wird für jeden dieser Frequenzkanäle die Werte oberhalb der Detektionsschwelle aufsummiert und daraus ein Einzahlwert für die Vorhersage der Klangverfärbung gebildet. Anschließend wird die Vorhersage noch mit einem freien Parameter multipliziert um die Vorhersage an die wirklichen Ergebnisse zu fitten.

Die Vorhersagen sind als durchgezogene und gestrichelte Linie in Abbildung 2 zusammen mit den Ergebnissen aus dem Hörversuch eingetragen. Es ist zu erkennen, dass die Vorhersagen insbesondere im Fall der Sprachstimuli gut mit den Ergebnissen aus dem Hörversuch übereinstimmen, lediglich für Stereophonie überschätzt das Modell die Klangverfärbung. Im Fall der Rauschstimuli sind ab einem Lautsprecherabstand von 4 cm größere Abweichungen zu erkennen und das Modell unterschätzt die Klangverfärbung.

Das vorgestellte Modell wurde in einem weiteren Schritt ausgebaut und mit mehreren freien Parameter versehen um ein besseres Fitten an Daten zu erlauben. Eine Beschreibung aller Parameter findet sich in Ende [6]. Da die Anzahl an Daten aber zur Zeit zu gering ist um ein sinnvolles Fitten mit mehreren Parametern zu ermöglichen, wird dieses Modell hier nicht weiter betrachtet. Stattdessen ist das von Moore und Tan beschriebene Modell implementiert worden und mit den Standard Parametern aus ihrer Studie auf die Stimuli angewendet worden. Dabei hat sich gezeigt, dass die Vorhersagen sich nicht signifikant von den mit dem oben beschriebenen und in Abbildung 2 dargestellten Ergebnissen unterscheiden.

Das bestätigt die Vermutung, dass die von Moore und Tan untersuchte Natürlichkeit von Kammfilter behafteten Stimuli mit der dazugehörigen Klangverfärbung korreliert ist und dieses Modell daher ebenfalls geeignet erscheint um dieses vorhersagen zu können.

## Zusammenfassung

Zur Verwendung der Wellenfeldsynthese werden Lautsprechergruppen benötigt, die in der Regel einen Lautsprecherabstand zwischen 10 cm und 30 cm aufweisen. Die vorliegende Studie hat die Auswirkungen dieser Lautsprecherabstände auf den Frequenzgang dieser Systeme und der damit verbundenen Wahrnehmung der Klangverfärbung untersucht. Dabei zeigte sich, dass die genannten Systeme eine starke Klangverfärbung aufweisen und diese zudem stärker ist als es bei einem Stereophonie-System der Fall wäre. Könnten jedoch sehr kleine Lautsprecherabstände von 4 cm oder kleiner realisiert werden, würde die Klangverfärbung hingegen geringer sein als für Stereophonie.

Weiter konnte gezeigt werden, dass ein auf Lautheitsunterschieden in einzelnen Frequenzbändern basierendes auditorisches Modell in der Lage ist die Ergebnisse für kleine Lautsprecherabstände vorherzusagen. Bei größeren Lautsprecherabständen kann es zu teilweise starken Unterschätzungen der Klangverfärbung kommen. Weiter zeigte sich, dass das Modell zur Vorhersage der Natürlichkeit bei Kammfiltersignalen von Moore und Tan [5] ähnliche Ergebnisse liefert und ebenfalls zur Untersuchung von Klangverfärbung geeignet erscheint.

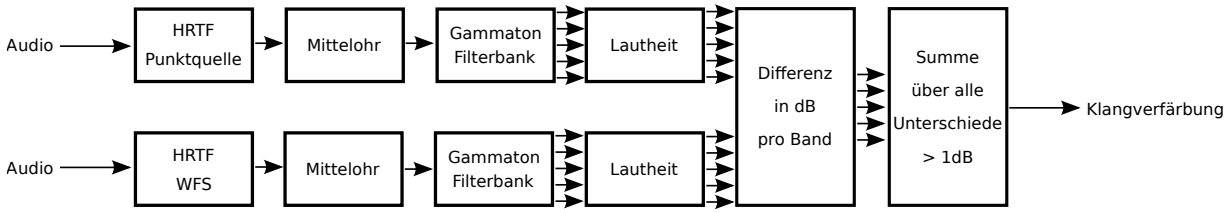


**Abbildung 3:** Blockschaltbild des verwendeten auditorischen Modells. Zur Vereinfachung ist jeweils nur die Verarbeitung in einem Ohr dargestellt.

## Danksagung

Diese Studie wurde gefördert durch EU FET grant Two!Ears, ICT-618075.

## Literatur

[1] Wittek H. Perceptual differences between wavefield synthesis and stereophony, University of Surrey, 2007

[2] H. Wierstorf - Perceptual Assessment of Sound Field Synthesis. (2014), Dissertation, Technische Universität Berlin

[3] M. Salomons - Coloration and binaural decoloration of sound due to reflections. (1995), Dissertation, Technische Universiteit Delft

[4] V. Pulkki - Coloration of Amplitude-Panned Virtual Sources. 110th Convention of the Audio Engineering Society (2001), Paper 5402

[5] B. C. J. Moore, C.-T. Tan - Development and Validation of a Method for Predicting the Perceived Naturalness of Sounds Subjected to Spectral Distortion. Journal of the Audio Engineering Society (2004), 52(9), 900-914

[6] C. Ende - Auditorische Modellierung der Klangverfärbung in der Wellenfeldsynthese. (2014), Bachelorarbeit (2014), Technische Universität Berlin