

# Methoden zur Anpassung einer Binauralsynthese an verschiedene raumakustische Parameter des Abhörtraumes

Nicolas Pachatz, Stephan Werner, Florian Klein<sup>1</sup>

<sup>1</sup> Technische Universität Ilmenau, 98693 Ilmenau

<sup>1</sup> {nicolas.pachatz; stephan.werner; florian.klein}@tu-ilmenau.de

## Motivation

Die Binauralsynthese ermöglicht es, jeden möglichen Punkt im Raum über ein stereophones Wiedergabesystem zu simulieren. Da Kopfhörer nur zwei Schallwandler besitzen, die sich direkt vor den Ohren befinden, fehlen hier, bis auf Pegelunterschiede zwischen linken und rechten Ohr, jegliche Richtungsinformationen.

Ein Problem, welches bei der Binauralsynthese in Verbindung mit Kopfhörern auftritt, ist der Raumdivergenzeffekt. Durch diesen kommt es zu einer verringerten Externalisierung und vermehrten Im-Kopf-Ortung der Audiosignale. Dafür gilt es, Algorithmen zu finden, die durch Anpassung der akustischen Parameter eines Raumes diesen Effekt minimieren. Hierbei sollte jedoch beachtet werden, dass die Anpassungen in einem Maß ausgeführt werden, die trotzdem die akustische Charakteristik des Ausgangsraumes erhalten [1].

## Der Raumdivergenzeffekt

Der Raumdivergenzeffekt entsteht durch eine abweichende Raumakustik des wahrgenommenen Raumes zum Abhörtraum. Durch diese Divergenz kommt es zu einer verringerten Externalität bei der Wiedergabe binauraler Signale.

Die Ursache für diesen Effekt liegt begründet in der Erwartungshaltung an die Raumakustik eines Raumes. Dadurch kommt es zu einem kognitiven Missverständnis.

Zudem wird der Effekt durch die optischen Merkmale des Raumes bestätigt und unterstützt. Stimmt beides überein, nimmt das Gehirn eine Erwartungshaltung gegenüber der Raumakustik an.

Kommt es zu einer ausreichend großen Abweichung der Muster zur Erkennung eines Raumes im auditorischen System, kann eine korrekte Verbindung des Gehörten mit den erwarteten Mustern nicht mehr gewährleistet werden [1].

Je stärker sich der erwartete und der gehörte Raum unterscheiden, desto stärker wirkt sich der Raumdivergenzeffekt aus. Über die Anpassung von verschiedenen raumakustischen Parametern des gehörten auf den befindlichen Raum des Hörers soll dieser Effekt minimiert werden [2].

## Anpassungen der raumakustischen Parameter

Die Anpassung erfolgt mittels Auftrennung von binauralen Raumimpulsantworten (BRIRs) in das direkte Schallsignal, die frühen Reflexionen und den diffusen Nachhall. Diese drei

Teilbereiche werden je nach Methode getrennt betrachtet und angepasst.

Die vorgestellten Methoden werden jedoch nicht gegenübergestellt, sondern sind aufeinander aufbauend (Abbildung 1).

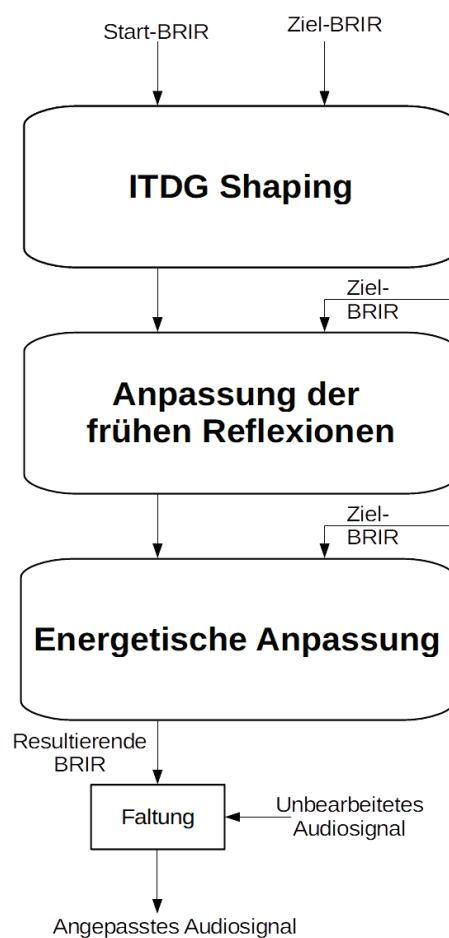


Abbildung 1 - Ablaufplan für die Anwendung von allen 3 Methoden zur Anpassung der BRIRs mit anschließender Faltung mit dem Audiosignal

## Energetische Anpassung

Bei dieser Methode werden zwei energetische Parameter einer BRIR betrachtet: Die *Energy Decay Curve* (EDC) und die *Direct to Reverberant Ratio* (DRR).

Bei der Anpassung der EDC werden beide Kurven verglichen und ein Quotient jedes Samples gebildet, aus denen ein Array von Faktoren errechnet wird, welches, multipliziert mit jedem Wert der EDC des Ausgangssignals, dieses energetisch so verändert, dass es die EDC der zweiten Impulsantwort ergibt.

Wie in Abbildung 2 zu sehen, können die EDCs bzw. BRIRs unterschiedliche Länge besitzen, was bei einer Quotientenbildung die Werte am Ende der kürzeren Impulsantwort unverhältnismäßig stark erhöhen würde.

Da in diesem Abschnitt ohnehin nahezu Stille herrscht, wird das Array mit einem Fenster multipliziert, dessen Schräge eine Länge von 20% der Länge des Arrays besitzt. Somit wird die Verstärkung unterdrückt und ein natürlicher Roll-off der BRIR gewährleistet.

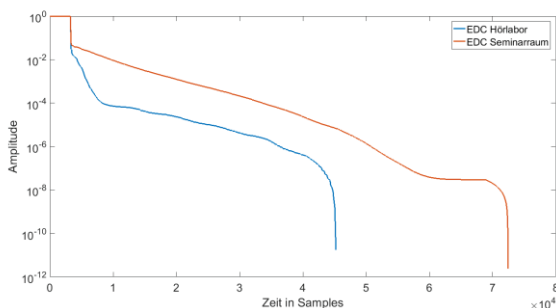


Abbildung 2 - Energy Decay Curves der BRIRs von 2 Räumen

Der zweite Teil dieser Methode besteht darin, die DRR, also das Verhältnis zwischen Direktschall und Nachhallenergie, anzupassen.

Wie bei der EDC werden die Werte der Ziel-BRIR und der resultierenden des vorherigen Schrittes bestimmt und miteinander verglichen. Allerdings wird nun nur ein Faktor gebildet, der mit dem gesamten Diffusschallanteil multipliziert wird. Somit ändert man nicht die temporale Energieverteilung, sondern nur die Gesamtenergie im Nachhall. Durch die Anpassung der EDC ist diese zwar schon recht ähnlich, bei einigen Kombinationen von Entfernungen hat sich jedoch gezeigt, dass sich die DRR auch noch weiter vom Zielwert entfernen kann. Aus diesem Grund ist die nachträgliche Anpassung der DRR notwendig, um eine ähnliche Lautheit des Diffusschalls zu erzielen.

Entgegen der EDC-Anpassung werden die Kanäle nicht getrennt betrachtet, sondern es wird ein Gesamtwert der DRR gebildet. In einigen Arbeiten findet die Anpassung nach den gemessenen Werten des der Schallquelle näheren Kanals statt. Das würde aber bei Schallquellen in der Nähe der Sagittalebene zu starken Abweichungen führen, da hier beide Kanäle eine ähnliche Gewichtung haben und ungleichmäßige reflektive Eigenschaften des Raumes keine Beachtung finden. Eine individuelle Betrachtung der Kanäle war jedoch nicht möglich, da es hier durch unterschiedliche Abhörpositionen in beiden Räumen zu einer Verschiebung der Schallquelle

durch Änderung der ILD kommen würde, durch welche ein Kanal stärker verändert wird als der andere.

Aus diesem Grund ist es notwendig, einen Mittelwert aus den Kanälen zu bilden, der auf beide angewendet wird. Dieser wird aus einem Produkt von den errechneten Energiewerten der Kanäle und einer zum Abstand der Schallquelle abhängigen Gewichtung errechnet. Der Abstand ergibt sich aus der Summe der kürzesten Distanz von Schallquelle zur Oberfläche des Kopfes und des Weges auf der Oberfläche zum jeweiligen Ohr. Die Kopfform wurde dabei als annähernd kugelförmig angenommen. Die Resultierende dieser Methode besitzt nun die gleiche DRR und EDC-Form der Zielfunktion.

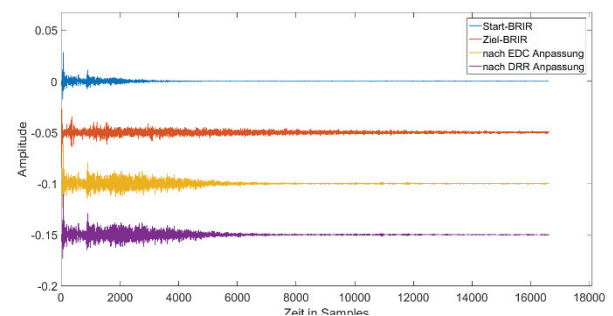


Abbildung 3 - Diffusschallanteil des Raumes "Hörlabor" auf "Seminarraum" (vgl. Abbildung 2) vor und nach der Energieanpassung

## Anpassung der Initialzeitlücke

In dem Verfahren wird die Initialzeitlücke (ITDG) der Start- auf die Ziel-BRIR angepasst. Dies geschieht mittels Kompression bzw. Expansion des Bereiches zwischen Direktschall und der ersten Reflektion. In diesem Bereich befinden sich in der Regel keine charakterisierenden Anteile für den Raum. Der Übergangspunkt zwischen Direktschall und Nachhallanteil wird nach Füg [3] mit  $0,6 \cdot \text{ITDG}$  nach dem Direktschallindex beschrieben. Beide Fenster überlappen sich bei Amplitudenwerten von 0,5.

Die Anpassung geschieht innerhalb von 2 Abschnitten. Der erste deckt den Bereich von Start- bis Endpunkt ab und ist ausschlaggebend für die Verschiebung. Der zweite Abschnitt betrifft den restlichen Diffusschallanteil und staucht oder streckt diesen bis zu einem konstanten Endpunkt, der so gewählt war, dass er sich bei der originalen und modifizierten Impulsantwort im diffusen Nachhall befindet. Dieser kann als stochastisches Rauschen angesehen werden und ist damit unabhängig von der Entfernung.

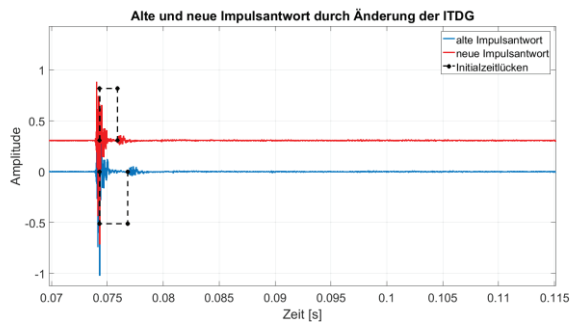


Abbildung 3 - Änderung der ITDG durch Kompression/Expansion

### Anpassung der frühen Reflexionen mittels Dynamic Time Warping

Diese Methode versucht, die Position der einzelnen Reflexionen zwischen Start- und Ziel-BRIR anzunähern. Eine konventionelle Interpolation im Zeitbereich würde zu Schmiereffekten führen, durch welche die Amplituden der einzelnen Reflexionen weniger stark ausgeprägt erscheinen lässt und somit zeitlich „verschmiert“. Somit würden neue Reflexionen an Positionen erzeugt werden, die in beiden BRIRs nicht vorkamen [4].

Darum wurde eine Interpolation mittels *Dynamic Time Warping* (DTW) gewählt. Dieser Algorithmus findet vor allem Anwendung in der Spracherkennung und dient dort zum Vergleich von Sprachmustern. Mithilfe der adaptiven Zeitnormierung ist es möglich, auch unterschiedlich betonte Wörter zu vergleichen. Die DTW-Funktion bildet 2 Wertefolgen aufeinander ab und sucht zwischen ihnen den „kostengünstigsten“ Pfad.

Um ein besseres Ergebnis zu erzielen, wird vor der Methode die ITDG angepasst. Somit befinden sich die beiden ersten Reflexionen beider BRIRs an der gleichen Position. Würde man darauf verzichten, könnten zu starke Veränderungen der Wellenform auftreten, da nicht nur die Form, sondern auch die globale Position dieser angepasst werden müsste [1].

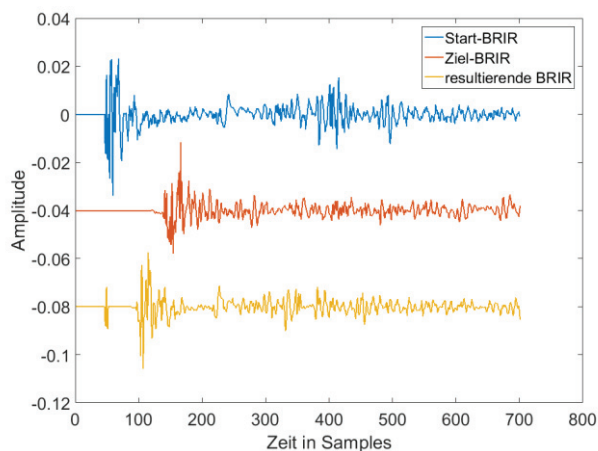


Abbildung 4 - Frühe Reflexionen vor und nach der Anpassung ohne vorherige Anpassung der ITDG. Durch die ungleiche Anfangsposition entstehen in der resultierenden BRIR Artefakte.

## Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurden Algorithmen entwickelt und vorgestellt, welche es ermöglichen, die akustischen Eigenschaften zweier Räume anzunähern. Das Ziel davon war eine Verbesserung der Externalisierung im Fall von Raumdivergenz beim Hören binauraler Signale über Kopfhörer. Die Ergebnisse können in Zukunft genutzt werden, um bei binauraler Kopfhörerwiedergabe in natürlichen Umgebungen oder bei späteren Hörtests eine verbesserte Wiedergabe zu erzielen.

In welchem Maße diese Beeinflussung stattfindet und ob sie zu einem perzeptiv besseren Ergebnis bei der Binauralsynthese führt, muss jedoch trotzdem in zukünftigen Tests überprüft werden. Außerdem existieren weitere Parameter, welche für die Entwicklung anderer Methoden ausschlaggebend sein können.

## Literatur

- [1] Georg Plenge. "Über das Problem der Im-Kopf-Lokalisation". In: *Acustica* 26.5 (1972), S. 241–252
- [2] Nicolas Pachatz, "Untersuchung zur Relevanz raumakustischer Parameter bei Anpassung eines binauralsynthesystems an die Raumakustik des Abhörspaces", Bachelorarbeit, Fachgebiet elektronische Medientechnik, Technische Universität Ilmenau, 2017.
- [3] S. Füg, "Untersuchungen zur Distanzwahrnehmung von Hörereignissen bei Kopfhörerwiedergabe," Masterarbeit, Technische Universität Ilmenau, 2012.
- [4] A. Lindau und S. Roos, "Perceptual evaluation of discretization and interpolation for motion-tracked binaural (MTB) recordings," in 26. Tonmeistertagung - VDT International Convention, November 2010.