

Dmitry Grigoriev, David Ackermann, Sönke Pelzer, Stefan Weinzierl

Ein psychologisches Messinstrument für die akustische Wahrnehmung von Räumen für Musik und Sprache: Stimulus-Erzeugung

Conference paper | Published version

This version is available at <https://doi.org/10.14279/depositonce-8808>



Grigoriev, Dmitry; Ackermann, David; Pelzer, Sönke; Weinzierl, Stefan (2016): Ein psychologisches Messinstrument für die akustische Wahrnehmung von Räumen für Musik und Sprache: Stimulus-Erzeugung. In: Proceedings of the Inter-Noise 2016 : 45th International Congress and Exposition on Noise Control Engineering : towards a quieter future : August 21-24, 2016, Hamburg. Berlin: Deutsche Gesellschaft für Akustik e.V. pp. 309–312.

Terms of Use

Copyright applies. A non-exclusive, non-transferable and limited right to use is granted. This document is intended solely for personal, non-commercial use.

Ein psychologisches Messinstrument für die akustische Wahrnehmung von Räumen für Musik und Sprache: Stimulus-Erzeugung

Dmitry Grigoriev¹, David Ackermann¹, Sönke Pelzer², Stefan Weinzierl¹,

¹ TU Berlin, FG Audiokommunikation, 10587 Berlin, Deutschland,

Email: dmitry.grigoriev@campus.tu-berlin.de, {david.ackermann, stefan.weinzierl}@tu-berlin.de

² RWTH Aachen, Institut für Technische Akustik, 52074 Aachen, Email: spe@akustik.rwth-aachen.de

Einleitung

In einem zweistufigen Verfahren wird ein Messinstrument für die akustische Wahrnehmung von Räumen entwickelt. Zunächst wurde durch eine Fokusgruppe von Experten aus dem Bereich der raumakustischen Forschung und Planung ein Konsens erzielt im Hinblick auf ein möglichst vollständiges Vokabular von Attributen zur Beschreibung raumakustischer Qualitäten. Um diese Attribute zu bewerten wird ein großes Sample von 70 unterschiedlichen, durch dynamische Binauralsynthese generierte raumakustische Umgebungen erstellt und deren raumakustische Parameter bestimmt. Als Anregung dieser virtuellen Räume dienen drei Kombinationen von Audio Content und akustischer Quelle (Sprecher, Trompete als Soloinstrument und Sinfonieorchester). Für die Auralisation eines vollzähligen Orchesters steht bis heute allerdings kein polyphones, nachhall- und übersprechfreies Audiomaterial zur Verfügung. Vorhandene Aufnahmen beinhalten meist nur ein bis zwei Instrumente pro Stimme ([1], [2]). Um dennoch eine vollständige Streichersektion eines Symphonieorchesters zu simulieren, wurde ein Verfahren [3] weiterentwickelt, welches einzelne nachhallfreie Aufnahmen eines Streichinstruments mit Hilfe des Phasenvocoders [4] vervielfältigt. Unter Berücksichtigung der korrekten Richtcharakteristik der Instrumente, der Bewegung der Musiker und einer charakteristischen Tonhöhenverteilung klassischer Musik wurden auf diese Weise binaurale Raumimpulsantworten (BRIR's) für 190 verschiedene modellbasierte Stimuli für die weitere Untersuchung erzeugt.

Methode

Durch eine dynamische, d.h. der Kopfposition der Hörer nachgeführte Binauralsynthese lässt sich eine perzeptiv hoch plausible Simulation von akustischen Szenen erzeugen [5], [6]. In der vorliegenden Studie wurde diese Möglichkeit genutzt, um die große Bandbreite raumakustischer Umgebungen für Sprache und Musik durch ein Sample von Computermodellen für Räume mit unterschiedlicher Größe, mittlerer Absorption und für unterschiedliche architektonische Typen abzubilden (vgl. Abb. 1), aus welchen binaurale Raumimpulsantworten für die Binauralsynthese erzeugt wurden.

Die ausgewählten akustischen Umgebungen sollen eine große Bandbreite an musikalischen Aufführungsräume repräsentativ abdecken. Um die Eigenschaften der Räume systematisch zu variieren, werden die Raummodelle in

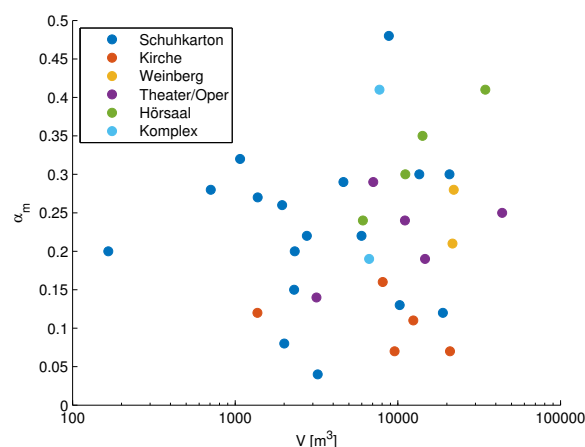


Abbildung 1: Darstellung des mittleren Absorptionsgrads α_m über das Raumvolumen V für die 6 Raumgeometrien der 35 Räume. (Kreise: 1. Hörposition, Kreuze: 2. Hörposition)

sechs Geometrien, sechs Raumvolumen und zwei mittleren Absorptionsgrade unterteilt. Die Modelle orientierten sich überwiegend an real existierenden Vorbildern, ohne den Anspruch diese exakt nachzubilden.

Position Quellen und Empfänger

Zur Bestimmung der raumakustischen Parameter und zur Auralisation des Soloinstruments und des Sprechers wurden in jedes Raummodell eine Quelle und zwei Empfänger platziert.

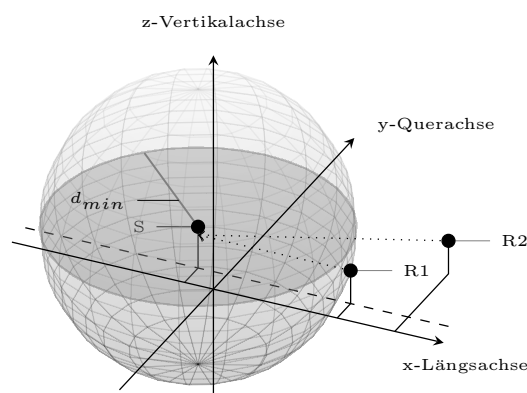


Abbildung 2: Positionierung der Quelle (S) und der beiden Empfänger (R1, R2) unter Berücksichtigung des Mindestabstands d_{min} und Vermeidung symmetrischer Positionen nach [7] und [8].

In Anlehnung an die einschlägigen Standards [7], [8]

wurde die Quelle an einer für die Nutzung des Raumes üblichen Position angeordnet, ihr akustisches Zentrum liegt 1,5 m über dem Boden. Die Empfänger befinden sich im Zuhörerbereich, sind mindestens 2 m voneinander entfernt, ihr akustisches Zentrum liegt 1,2 m über dem Boden, mit einem Mindestabstand d_{min} zwischen Quell- und Empfängerposition in Höhe des doppelten Hallradius. Der Blickwinkel beider Empfänger wird so auf die Quelle ausgerichtet, dass bei 0° Azimut die Quellposition auf der Medianebene der Empfänger liegt. Der Elevationswinkel der Empfänger beträgt 0° relativ zu der Ebene, die von der Längs- und Querachse aufgespannt wird (vgl. Abb. 2). Die 66 Orchesterinstrumente sind an den entsprechend der Nutzung des Raumes üblichen Positionen angeordnet, ihr akustisches Zentrum liegt 1,2 m über dem Boden. Die Orchesteraufstellung (vgl. Abb. 3) der Instrumente ist für alle Räume identisch und orientiert sich an den Bestuhlungsplänen des Deutschen Symphonie-Orchesters Berlin. Die Empfängerpositionen bleiben hierbei unverändert.

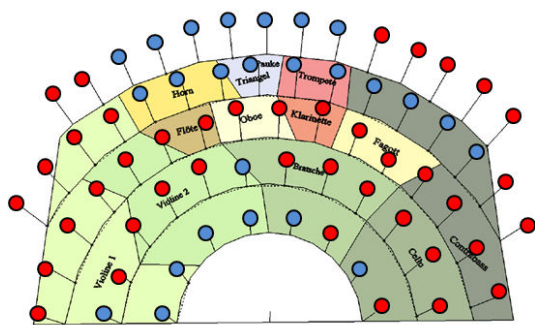


Abbildung 3: Orchesteraufstellung: Die Orchesterstimmen sind in Flächen verschiedener Farben unterteilt. Jeder Kreis repräsentiert eine Quelle (rot: reale Aufnahme, blau: vervielfältigte Spur).

Simulationsparameter/Richtcharakteristik

Als Richtcharakteristik der Empfänger wurden die Außenohrübertragungsfunktionen (HRTFs) des Kunstkopf-Torso-Simulator (HATS) FABIAN gewählt [9]. Zur Bestimmung der raumakustischen Parameter wurde mit omnidirektionaler Quelle simuliert, die binauralen Raumimpulsantworten für eine Anregung mit einem Sprecher, einem Soloinstrument und Symphonieorchester wurden durchgehend mit jeweils spezifischer Richtcharakteristik erzeugt.

Diese zunächst messtechnisch erfassten Richtcharakteristiken [10] wurden anhand der Bewegungsdaten, die in einem Live-Auftritt unter konzertähnlichen Bedingungen mit einem Motion-Tracking System aufgenommen wurden [11] gemittelt, was zu einer räumlichen Glättung der Spektren führte. Darüber hinaus wurden die z.T. stark Tonhöhen-abhängigen Richtcharakteristiken einer gewichteten Mittelung unterzogen, die sich an der Tonhöhenverteilung in den Sinfonien von L.v. Beethoven orientiert [12]. Abschließend wurden die Richtcharakteristiken auf die jeweils mittlere Energie im Frequenzband normiert, ähnlich der Diffusfeldentzer-

rung eines Mikrofons. Die resultierenden Daten werden im OpenDAFF Format [13] an die Simulationssoftware übergeben.

Akquise modellbasierter BRIRs mit RAVEN

Bei einer für eine plausible Simulation erforderlichen, horizontalen Winkelauflösung von 2° [14] und einer konstanten Elevation von 0° wurden für jeden Empfänger 180 BRIRs pro Quelle erfasst. Zur Berechnung der BRIRs wurde die raumakustische Simulationssoftware RAVEN eingesetzt [15], unterschiedliche Kopforientierungen wurden über das interne Animationsmodul realisiert. Dabei wird ein Quellsignal mit im Abstand der doppelten Nachhallzeit aufeinander folgenden Diracstößen eingespielt und die Kopfposition des Empfängers vor jedem nachfolgenden Diracstoß um 2° horizontal gedreht. Somit lässt sich ein Audiostream mit sequentiell aneinandergereihten BRIRs für alle Kopfdrotationen generieren. Die einzelnen BRIRs wurden nach [16] in Direktschall und frühe Reflexionen sowie in diffusen Nachhall getrennt und in das SOFA-Format [17] konvertiert. Die Auralisation erfolgt mit dem SoundScene Renderer [18].

Vervielfachung der Tonspuren

Bei der Untersuchung einzelner, zur gleichen Orchesterstimme gehörender Instrumente einer Orchesteraufnahme lassen sich sowohl im Zeit- als auch Frequenzbereich Unterschiede erkennen. Im Zeitbereich variieren die Anschlagszeitpunkte (Onset), die Tondauer und die Lautstärke eines gespielten Tons. Im Frequenzbereich schwankt die gespielte Tonhöhe um den notierten Ton. Durch Manipulation dieser Variablen (Onset, Tondauer, Dynamik und Tonhöhe) wurde somit die künstliche Vervielfachung einer Solostimme realisiert, wobei sich der Algorithmus an einer Verteilung für die einzelnen Parameter orientierte, die für ein Spitzenorchester empirisch bestimmt wurde [19].

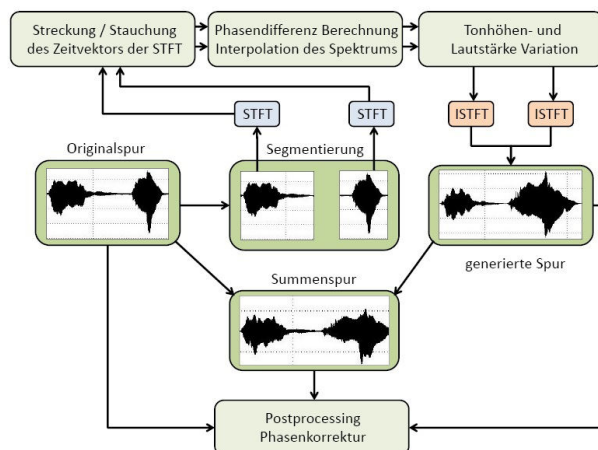


Abbildung 4: Vervielfachung einer Tonspur: Zunächst wird die Originalspur segmentiert, danach u.a. mit Hilfe des Phasenvocoders nach [4] die Variablen Onset, Tondauer, Lautstärke und Tonhöhe variiert und schließlich, als wieder zusammengesetzte generierte Spur von Kammfiltereffekt-Artefakten beseitigt. [(I)STFT: (Inverse) Kurzzeit-Fourier-Transformation]

Segmentierung, Phasenvocoder und Variation der Variablen

Für die Vervielfachung der Tonspuren wurde der Ansatz aus [3] aufgegriffen und weiterentwickelt (vgl. Abb. 4). Damit innerhalb eines gespielten Tons keine abrupten Tonhöhen-, Onset- oder Tondaueränderungen auftreten, muss die Tonspur in einzelne kurze Abschnitte (in der Größe eines Einzeltons oder weniger aufeinanderfolgender Töne) aufgeteilt werden.

Auf die extrahierten Einzelabschnitte wird eine Kurzzeit-Fourier-Transformation (STFT) mit $N = 2048$ Samples und einem Overlap von 75% angewandt und die Länge des Zeitvektors der einzelnen Abschnitte, basierend auf der Onset- und Tondauerverteilung [19], zufällig variiert. Darauf folgend wird die Phasendifferenz zwischen zwei Zeitfenstern mit der *scaled phase locking* Methode [4] berechnet und das Betragsspektrum interpoliert. Die einzelnen Abschnitte werden abwechselnd vergrößert bzw. verkleinert, damit keine Gesamtverzögerung bzw. -beschleunigung des Audioinhalts entsteht. Ausgehend von der Dynamik- und Tonhöhenverteilung werden die Abschnitte zusätzlich in der Lautstärke und Tonhöhe zufällig variiert.

Bei der Addition der so gewonnenen Audiospuren kann in schnell gespielten Passagen oder gleichmäßig übergehenden Tönen ein hörbarer Kammfiltereffekt auftreten. Dieser entsteht durch fluktuierende Einbrüche in bestimmten Frequenzbereichen der Summenspur, die durch ungünstige Phasenlage der einzelnen Tonspuren hervorgerufen werden. Um diesem Effekt vorzubeugen wird in einem Postprocessingschritt die Phase der generierten Spur frequenzspezifisch angepasst, sodass bei Addition mit der Originalspur, weniger Fluktuationen im Betragsspektrum der Summentonspur entstehen und somit der Kammfiltereffekt reduziert wird.

Ergebnisse

Es wurden in 35 virtuellen akustischen Umgebungen binaurale Raumimpulsantworten mit horizontaler Winkelauflösung von 2° bei konstanter Elevation von 0° für eine Anregung mit einem Sprecher und einer Trompete als Soloinstrument für zwei Empfängerpositionen erzeugt. Für den Audiodienst 'Orchester' wurden, basierend auf den Aufnahmen des 3. Satzes der 4. Brahms Symphonie [2], mit dem hier vorgestellten Verfahren aus jeweils drei Aufnahmen für die 1. und 2. Violine insgesamt 12 bzw. 11 Spuren generiert und weiterhin aus jeweils einer Bratschen-, Cello-, und Kontrabassaufnahme, 10 Bratschen-, 9 Cello- und 8 Kontrabassaufnahmen erstellt. Mit weiteren drei Horn-, jeweils zwei Fagott-, Oboen-, Flöten-, Klarinetten-, Trompetenaufnahmen und jeweils einer Triangel-, Pauken- und Kontrafagottaufnahme liegt ein Orchester mit einer Besetzung von 66 Musikern vor. BRIR's mit einer typischen Orchesteranordnung dieser 66 Quellen konnten in 25 Räumen für zwei Hörerpositionen akquiriert werden.

Die 35 virtuellen akustischen Umgebungen repräsentieren unterschiedliche architektonische Typen. Die Raumvolumina erstrecken sich über ein Bereich von

166 m³ bis 43790 m³ (vgl. Abb. 1) 18 Raummodelle besitzen einen mittleren Absorptionsgrad $\alpha_m > 0,24$ und 17 Modelle einen mittleren Absorptionsgrad $\alpha_m \geq 0,24$ (vgl. Abbildung 1).

Die raumakustischen Parameter nach [7] wurden je Raum und für beide Empfängerpositionen bestimmt und liegen in einer Auflösung über 31 Terzbänder von 20 Hz bis 20 kHz vor (vgl. Tabelle 1). Im Fall des Symphonieorchesters wurden die raumakustischen Parameter über alle 66 Quellpositionen gemittelt.

Tabelle 1: Minima, Maxima und Median der simulierten akustischen Parameter.

Parameter	Min.	Med.	Max.
T_{30} [s]	0,56	1,54	7,08
EDT [s]	0,46	1,53	4,84
G [dB]	-1,47	7,46	18,91
C_{80} [dB]	-6,62	2,04	11,19
D_{50}	0,10	0,46	0,86
T_s [s]	0,02	0,09	0,37
J_{LF}	0,002	0,286	0,525
J_{LFC}	0,004	0,302	0,466
L_J [dB]	-12,83	-7,47	-4,56
$IACC_{All}$	0,18	0,38	0,75
BR	0,66	1,20	2,53

Diskussion

Für einen Hörversuch zur Entwicklung eines psychologischen Messinstruments für die raumakustische Wahrnehmung wurden Stimuli für die dynamische Binauralsynthese erzeugt. Die Verwendung von simulierten Umgebungen ermöglicht es hierbei, ein für die Aufführung von Sprache und Musik repräsentatives Sample von Räumen mit großer raumakustischer Varianz zu verwenden. Eine besondere Herausforderung bildet die Simulation einer Orchesteraufführung, sowohl im Hinblick auf die Synthese vollständiger Orchesterstimmen auf der Grundlage von nur wenigen nachhallfreien Aufnahmen, als auch die Verwendung jeweils korrekter Richtcharakteristiken.

Auch die absoluten Schallleistungen der einzelnen Instrumente ließen sich aus der Dokumentation der nachhallfreien Aufnahmen nicht rekonstruieren, daher musste der Wiedergabepegel der Soloinstrumente und des Orchesters von drei Experten (Musiker, Tonmeister) in einem Vorversuch eingestellt werden.

Aus den Problemen, diese Parameter (Richtcharakteristik, Wiedergabepegel) für Auralisationen auf eine physikalisch zuverlässige Weise ohne manuelle Feinarbeit zu konfigurieren, resultieren eine Reihe von Anforderungen an die Produktion und die Dokumentation von zukünftigem, nachhallfreiem Audio Content.

In einem informellen Hörversuch wurde die vorliegende Auralisation dennoch bereits jetzt als plausibles Abbild realer Solostimmen bzw. eines realen Orchesters wahrgenommen.

Literatur

- [1] Pätynen, J. et al. (2008): „Anechoic recording system for symphony orchestra.”, *Acta Acustica united with Acustica*, 94, 856-865.
- [2] Vigeant, M. C. et al. (2010): „Multi-Channel Orchestral Anechoic Recordings for Auralizations.” In: *Proc. of the International Symposium on Room Acoustics*, Melbourne.
- [3] Pätynen, J. et al. (2011): „Simulation of the violin section sound based on the analysis of orchestra performance.” In: *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, S. 173-176.
- [4] Laroche, J.; Dolson, M. (1999): „Improved Phase Vocoder Time-Scale Modification of Audio.” In: *IEEE Transactions on Speech and Audio Processing*, 7(3), S. 323-332.
- [5] Lindau, A. et al. (2007): „Binaural resynthesis for comparative studies of acoustical environments.” In: *Proc. of the 122th AES Convention*, Wien.
- [6] Lindau, A.; Weinzierl, S. (2012): „Assessing the plausibility of virtual acoustic environments.” *Acta Acustica united with Acustica* 98, 804-810.
- [7] DIN (2009): *DIN EN ISO 3382-1:2009: Akustik - Messung von Parametern der Raumakustik - Teil 1: Aufführungsräume*. Berlin: Beuth
- [8] DIN (2008): *DIN EN ISO 3382-2:2008: Akustik - Messung von Parametern der Raumakustik - Teil 2: Nachhallzeit in gewöhnlichen Räumen*. Berlin: Beuth
- [9] Lindau, A.; Weinzierl, S. (2007): „FABIAN-schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden.” In: *Fortschritte der Akustik: Tagungsband d. 33. DAGA*, Stuttgart, 33(2) S. 633-634.
- [10] Pollow, M. et al. (2010): „Musical Instrument Recording for Building a Directivity Database.” In: *Fortschritte der Akustik: Tagungsband d. 36. DAGA*, Berlin.
- [11] Steger, D. et al. (2015): „Spielbewegungen von Musikinstrumenten und deren Bedeutung für das Klangergebnis - Ein experimenteller Zugang durch Motion Tracking klassischer Orchesterinstrumente und Auralisation der Bewegungsdaten.” In: *Fortschritte der Akustik: Tagungsband d. 41. DAGA*, Nürnberg.
- [12] Quiring, R., Weinzierl, S. (2016): „Tonhöhenverteilungen im klassischen Orchesterrepertoire.” In: *Fortschritte der Akustik: Tagungsband d. 42. DAGA*, Aachen.
- [13] Wefers, F. (2010): „OpenDAFF - Ein freies quell-offenes Software-Paket für richtungsabhängige Audiodaten.” In: *Fortschritte der Akustik : 36. Deutsche Jahrestagung für Akustik*, Berlin.
- [14] Lindau, A.; Weinzierl, S. (2009): „On the Spatial Resolution of Virtual Acoustic Environments for Head Movements in Horizontal, Vertical and Lateral Direction.” In: *Proc. of the EAA Symposium on Auralization*, Espoo.
- [15] Schröder, D. (2011): *Physically Based Real-Time Auralization of Interactive Virtual Environments*. Dissertation, Rheinisch-Westfälische Technische Hochschule, Fakultät für Elektrotechnik und Informationstechnik, Aachen.
- [16] Lindau, A. et al. (2010): „Perceptual Evaluation of Physical Predictors of the Mixing Time in Binaural Room Impulse Responses.” In: *Proc. of the 128th AES Convention*, London.
- [17] AES (2015): *AES69-2015: AES standard for file exchange - Spatial acoustic data file format*.
- [18] Geier, M. et al. (2008): „The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods.” In: *Proc. of the 124th AES Convention*, Amsterdam.
- [19] Recke, T. (2011): *Zum orchestralen Klang, eine Analyse der Verteilung von Onset-Zeiten und Tonhöhen in der Streichergruppe eines Spitzenorchesters*, Masterarbeit, Fachgebiet Audiokommunikation, TU Berlin.