

Exposé zur Masterarbeit

Untersuchung des Zusammenhangs physikalischer Maße zur Bestimmung der *mixing time* in Raumimpulsantworten und ihrem perzeptiven Äquivalent

Linda Kosanke, Matrikelnummer 301943, lindakosanke@yahoo.de

Technische Universität Berlin

Fakultät I

Institut für Sprache und Kommunikation

Fachgebiet Audiokommunikation

Zusammenfassung

Die dynamische Binauralsynthese erfordert große Datensätze binauraler Raumimpulsantworten. Der Nachhall der Impulsantworten verhält sich nach einer bestimmten Zeit stochastisch, die zeitliche Feinstruktur kann zunehmend schlechter erkannt werden. Nach Ablauf einer gewissen Zeit kann der Nachhall unabhängig vom Ort gewählt werden. Damit ist eine Reduktion sowohl des Datensatzes als auch des Echtzeitrechenaufwandes möglich, ohne einen Qualitätsverlust hinnehmen zu müssen. Ziel der Arbeit ist es, den Zeitpunkt zu bestimmen, ab dem eine Manipulation des Nachhalls nicht mehr wahrnehmbar ist. Dazu werden in neun Räumen mit systematisch variierten mittleren Absorptionsgraden und Volumina binaurale Impulsantworten gemessen. In einem Hörversuch wird untersucht, ab welchem Zeitpunkt die Überblendung eines Direktschalls in einen diffusen Nachhall einer anderen Impulsantwort nicht mehr wahrnehmbar ist. Die Ergebnisse des Schwellwertversuches werden deskriptiv und inferenzstatistisch ausgewertet. Mittels multipler Regression und Korrelation sollen Zusammenhänge zu physikalischen Prädiktoren untersucht werden.

Einleitung und Fragestellung

Virtuelle akustische Umgebungen können auf Basis modellgenerierter oder empirischer binauraler Raumimpulsantworten (BRIRs) erzeugt werden. Die Auralisation kann mittels dynamischer Binauralsynthese realisiert werden, bei der ein nachhallfreies Quellsignal mit den BRIRs entsprechend der aktuellen Kopfposition eines Hörers in Echtzeit gefaltet wird. Der Rechenaufwand für dieses Verfahren ist sehr hoch, so dass perzeptive Reduktionsmöglichkeiten gesucht werden, ohne dass die Auralisation an Qualität verliert.

Impulsantworten von Räumen setzen sich aus Direktschall, frühen Reflexionen und spätem Nachhall zusammen. Der Direktschall ist der Schall, der ohne Umwege von der Quelle zum Empfänger gelangt. Die nachfolgenden schwächeren Einzelreflexionen entstehen durch Schallrückwürfe der Seitenwände, der Decke und des Bodens. Sie sind in der Impulsantwort deutlich erkennbar und auch perzeptiv unterscheidbar. Nach einer bestimmten Zeit werden diese Reflexionen so dicht, dass Feinstrukturen nicht mehr wahrnehmbar sind. Die Impulsantwort verhält sich stochastisch, wenn kein Zusammenhang zum Ausgangszustand des Schallfeldes besteht. Der stochastische Nachhall ist diffus, d.h. er enthält keine Information über die Einfallsrichtung des Schalls und besitzt an jedem Punkt im Raum die gleiche Energiedichte (Schroeder, 1959). Der Übergang in den stochastischen, späten Nachhall wird auch als physikalische mixing time bezeichnet (Polack, 1993). Im Gegensatz dazu handelt es sich bei der perzeptiven mixing time um den Zeitpunkt, ab dem der Nachhall einer Raumimpulsantwort ohne hörbare Konsequenzen durch einen Nachhall einer anderen beliebigen Sender-/Empfängerkonstellation ausgetauscht werden kann. Die Reduktion des Rechenaufwandes bei der Auralisation kann realisiert werden, da der Nachhall der Impulsantwort nicht mehr dynamisch entsprechend der Kopfposition eines Hörers aktualisiert werden muss, sondern durch einen für den gesamten Raum repräsentativen Nachhall ersetzt werden kann. Ziel dieser Arbeit ist es, den Zusammenhang der perzeptiven mixing time mit den physikalischen Eigenschaften von Räumen und geeignete Prädiktoren zur Vorhersage der perzeptiven mixing time zu finden.

Stand der Forschung

Zur Bestimmung des Übergangzeitpunkts von Anfangsreflexionen der Raumimpulsantwort zu diffusem Nachhall existieren im Wesentlichen drei physikalisch motivierte Ansätze:

Die einfachsten Maße sind feste Zeiten. Sie liegen zwischen 50 und 200 ms (vgl. Bradley & Souloire, 1995; Begault, 1992; Hidaka *et al.*, 1995).

Weiterhin wurden viele Maße veröffentlicht, die auf den physikalischen Eigenschaften des Raumes und daraus abgeleiteten Modellannahmen beruhen. Reichardt & Lehmann (1978), Cremer (1961), Rubak & Johansen (1999) und Schmidt & Ahnert (1973) leiten die mixing time aus einer bestimmten Reflexionsdichte ab, die notwendig ist, damit Einzelreflexionen nicht mehr wahrnehmbar sind. Die Berechnungsvorschriften basieren im Wesentlichen nur auf dem Volumen bzw. dessen Wurzel. Rubak & Johansen (1999) legen der Berechnung die mittlere freie Weglänge zugrunde. Hidaka *et al.* (2007) versuchte, die empirischen Ergebnisse eines Prädiktors der physikalischen mixing time per Regression über die Nachhallzeit vorherzusagen.

Außerdem wurden verschiedene Maße zur Prädiktion der physikalischen mixing time entwickelt, die auf der Auswertung von gemessenen Raumimpulsantworten basieren. Abel & Huang (2006) und Stewart & Sandler (2007) stellten Methoden vor, die prüfen, ob eine im diffusen Schallfeld annehmbare Normalverteilung der Amplituden im Zeitbereich vorliegt. Ein Maß von Hidaka *et al.* (2007) beruht auf der zeitlichen Korrelation zwischen den Schallenergien ab dem Zeitpunkt des Direktschalls und späteren Zeitpunkten. Eine weitere Methode nutzt Matching Pursuit, um die Impuls-

antwort zu zerlegen und aus der so entstandenen zeitlichen Abfolge von Reflexionen den Zeitpunkt des Übergangs in ein diffuses Verhalten zu finden (Defrance *et al.*, 2009).

Bisher gibt es wenige empirische Untersuchungen der perzeptiven mixing time. In einem kleinen Hörsaal wurde sie von Meesawat & Hammershøi (2003) untersucht. Dabei wurden verschiedene Kombinationen von frühen und späten Impulsantwortanteilen von unterschiedlichen Quell-/Empfängerrelationen getestet. In diesem speziellen Raum wurde positionsunabhängig eine perzeptive mixing time von ca. 40 ms gefunden. Die perzeptive mixing time eines großen Hörsaals wurde von Lindau *et al.* (2007) untersucht. Die mixing time war erwartungsgemäß höher (140 ms) und außerdem von der Position der Quelle und des Empfängers abhängig. Zudem stellte sich ein Schlagzeugsample als besonders kritisch heraus. Bisher existiert noch keine Studie, in der die perzeptive mixing time systematisch in verschiedenen Räumen und unter Nutzung der dynamischen Binauralsynthese untersucht wurde.

Methode

Der Einfluss von Volumen und relativer Nachhalldauer auf die perzeptive mixing time soll systematisch untersucht werden. Da Volumen und Nachhalldauer sich nicht unabhängig voneinander verändern lassen, werden die Räume bezüglich der vom Volumen unabhängigen mittleren Absorptionsfläche variiert. Die Faktoren Volumen und mittlere Absorptionsfläche bzw. relative Nachhallstärke werden jeweils dreifach gestuft, so dass Datensätze binauraler Impulsantworten aus insgesamt 9 Räumen benötigt werden. Tabelle 1 zeigt das Untersuchungsschema mit den Faktoren Volumen und mittlerem Absorptionsgrad. Die Räume werden so ausgewählt, dass der Abstand zwischen den Nachhallzeiten mindestens 10 % beträgt, damit Unterschiede gerade wahrnehmbar sind. Gemäß Hidaka *et al.* (2007) ist die mixing time in Rechteckräumen maximal. Die Auswahl beschränkt sich dementsprechend auf im Wesentlichen rechteckförmige Räume. Bei den Räumen handelt es sich um ein Tonstudio, Seminarräume und Hörsäle der TU Berlin, einen Konzertsaal, einen Aufnahmesaal und eine Kirche.

	kleines Volumen	mittleres Volumen	großes Volumen
großes α	Studio EN324 $\alpha=0.36$ (RT=0.39 s), V=216 m ³	Hörsaal H0104 $\alpha=0.28$ (RT=1.15 s), V=3300 m ³	Konzertsaal UdK $\alpha=0.33$ (RT=1.52 s), V=8298 m ³
mittleres α	Seminarraum EN111 $\alpha=0.26$ (RT=0.62 s), V=224 m ³	Hörsaal HE101 $\alpha=0.23$ (RT=1.67 s), V=5179 m ³	Hörsaal H0105 $\alpha=0.23$ (RT=2.08 s), V=8500 m ³
kleines α	Seminarraum EN191 $\alpha=0.17$ (RT=0.79 s), V=182 m ³	Aufnahmesaal Teldex $\alpha=0.2$ (RT=1.83 s), V=3647 m ³	Jesus-Christus-Kirche $\alpha=0.23$ (RT=2.36 s), V=7417 m ³

Tabelle 1: untersuchte Räume, mittlerer Absorptionsgrad bzw. Nachhallzeit und Volumen systematisch variiert (3x3)

In allen Räumen werden Messungen durchgeführt, die den interaktiven Hörversuch ermöglichen. Die Messungen sind in zwei Abschnitte unterteilt:

- monaurale Nachhallzeitmessung und Hallradiusbestimmung
- Messen der BRIRs mit dem Kunstkopftorsosimulator FABIAN

Als Quelle wird ein frequenzgangkompensierter 3-Wege-Dodekaeder verwendet, der auf der Bühne bzw. im vorderen Drittel entlang der Längsrichtung des Raumes positioniert wird. Die Nachhallzeit wird aus monauralen Impulsantworten, die mit einem Messmikrofon an drei Mikrofonpositionen im Diffusfeld gemessen werden, bestimmt. Die binauralen Impulsantworten werden mit dem institutseigenen Messsystem FABIAN gemessen, einem Messroboter bestehend aus Kopf und Torso mit integrierten Mikrofonen an der Position der Ohren und beweglichem Halsgelenk (Lindau *et al.*, 2007). Der Messroboter wird im doppelten Hallradius zur Quelle auf einem repräsentativen Hörerplatz mittig im Raum aufgebaut und frontal zum Lautsprecher ausgerichtet. Die binauralen Impulsantworten werden in 1°-Schritten für horizontale Kopfpositionen zwischen -80° und $+80^\circ$ gemessen.

Die perzeptive mixing time wird in einem Hörversuch untersucht. In einem adaptiven 3-AFC-Test sollen die Versuchspersonen die originale dynamische binaurale Simulation von einer manipulierten Variante unterscheiden. In der originalen Simulation entsprechen die Impulsantworten immer der aktuellen Kopfposition eines Hörers. In der manipulierten Simulation werden nur die frühen Anteile der Impulsantworten entsprechend der aktuellen Kopfposition nachgeführt, der späte Nachhall wird immer der BRIR, die mit der Kopfposition nach vorn gemessen wurde, entnommen und ist konstant. Im adaptiven Schwellwertverfahren wird der Übergangzeitpunkt zwischen frühen und späten Anteilen der Impulsantworten gemäß dem Antwortverhalten der Versuchsperson während des Versuchs variiert.

Als Stimulus wird ein Schlagzeugsample verwendet, das sich in früheren Untersuchungen als besonders kritisch zeigte (Lindau *et al.*, 2007). Die Versuchspersonen sind Studenten und Mitarbeiter des Fachgebiets.

Die oben genannten Methoden zur Bestimmung der physikalischen mixing time werden in Matlab implementiert und eingesetzt. Mit linearer Regression wird der Zusammenhang zwischen physikalischen Maßen und perzeptiven Schwellwerten untersucht.

Mittels Varianzanalyse wird getestet, welche Effekte sich auf die perzeptive mixing time bezüglich Volumen und mittlerem Absorptionsgrad ergeben und ob Interaktionen auftreten. Mit Hilfe einfacher und multipler Regression sollen optimale Prädiktoren für die perzeptive mixing time gefunden werden.

Zeitplan

August 2009 bis Dezember 2009

- Literaturrecherche und -durchsicht
- Implementierung der physikalischen Maße in Matlab
- Messung der BRIR-Datensätze
- Vorbereitung des Hörversuchs

Januar 2010 bis Februar 2010

- Durchführung des Hörversuchs
- Auswertung des Hörversuchs

März 2010 bis Juli 2010

- Aufbereiten der Ergebnisse
- Verfassen der Arbeit

Literatur

- Abel, Jonathan S., & Huang, Patty. 2006. A Simple, Robust Measure of Reverberation Echo Density. *In: Proc. of the 121st AES Convention, San Francisco.*
- Begault, Durand R. 1992. Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems. *J. Audio Eng. Soc.*, **40**(11), 895–904.
- Bortz, Jürgen. 2005. *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bortz, Jürgen, & Döring, Nicola. 2006. *Forschungsmethoden und Evaluation*. Heidelberg: Springer.
- Bradley, John S., & Souloire, Gilbert A. 1995. The influence of late arriving energy on spatial impression. *J. Acoust. Soc. Am.*, **97**(4), 2263–2271.
- Cremer, Lothar. 1961. *Statistische Raumakustik*. Stuttgart: S. Hirzel Verlag.
- Defrance, G., Daudet, L., & Polack, J.-D. 2009. Using Matching Pursuit for Estimating Mixing Time Within Room Impulse Responses. *Acta Acustica united with Acustica*, **95**, 1071–1081.
- Hidaka, Takayuki, Beranek, Leo L., & Okano, Toshiyuki. 1995. Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls. *J. Acoust. Soc. Am.*, **98**(2), 988–1007.
- Hidaka, Takayuki, Yamada, Yoshinari, & Nakagawa, Takehiko. 2007. A new definition of boundary point between early reflections and late reverberation in room impulse responses. *J. Acoust. Soc. Am.*, **122**(1), 326–332.
- Lindau, Alexander, Hohn, Torben, & Weinzierl, Stefan. 2007. Binaural resynthesis for comparative studies of acoustical environments. *In: Proc. of the 112th AES Convention, Vienna.*
- Meesawat, Kittiphong, & Hammershøi, Dorte. 2003. The time when the reverberation tail in a binaural room impulse response begins. *In: Proc. of the 115th AES Convention, New York.* preprint 5859.
- Polack, Jean-Dominique. 1993. Playing Billiards in the Concert Hall: The Mathematical Foundations of Geometrical Room Acoustics. *Applied Acoustics*, **38**, 235–244.
- Reichardt, W., & Lehmann, U. 1978. Raumeindruck als Oberbegriff von Räumlichkeit und Halligkeit, Erläuterungen des Raumeindrucksmaßes R. *Acustica*, **40**(5), 277–290.
- Rubak, Per, & Johansen, Lars G. 1999. Artificial Reverberation based on a Pseudorandom Impulse Response II. *In: Proc. of the 106th AES Convention, Munich.* preprint 4900.

- Schmidt, Wolfgang, & Ahnert, Wolfgang. 1973. Einfluss der Richtungs- und Zeitdiffusität auf den Raumeindruck. *Wiss. Z. Techn. Univers. Dresden*, **22**(2), 313–317.
- Schroeder, Manfred R. 1959. Measurement of Sound Diffusion in Reverberation Chambers. *J. Acoust. Soc. Am.*, **31**(11), 1407–1417.
- Stewart, Rebecca, & Sandler, Mark. 2007. Statistical Measures Of Early Reflections Of Room Impulse Responses. *In: 10th Int. Conference on Digital Audio Effects, Bordeaux*.