

# Länge von Außenohr-Impulsantworten und Auswahl von Reflexionen für interaktive auditive virtuelle Umgebungen

Andreas Silzle

Institut für Kommunikationsakustik, Ruhr-Universität Bochum, andreas.silzle@ruhr-uni-bochum.de

## Einleitung

Die Generierung einer plausiblen interaktiven auditiven virtuellen Umgebung erfordert einen hohen Rechenaufwand, siehe [1]. Die Anzahl der notwendigen Rechenoperationen ist stark abhängig von der Ordnung der simulierten Reflexionen und der damit stark ansteigenden Anzahl an zu berechnenden Spiegelschallquellen. Sie übersteigt leicht die auf einem heutigen PC vorhandene Rechenkapazität, siehe oberste durchgezogene Kurve in Abbildung 2.

Für die Optimierung einer interaktiven auditiven virtuellen Umgebung ist es hilfreich die relevanten Qualitätselemente auf der physikalischen Ebene und die für die Applikation wichtigen Qualitätsmerkmale auf der Wahrnehmungsebene in einer Taxonomie zu identifizieren. Mit einer effizienten Suchmethode sind dann diejenigen Einstellungen der Qualitätselemente zu finden, die a) den Qualitätsansprüchen des Benutzers dieser Applikation (Quality of Experience) genügen und b) dies mit einem möglichst geringen Rechenaufwand erreichen, siehe [3].

Für die folgende Untersuchung wurden diejenigen Qualitätselemente ausgewählt, die die meisten Rechenoperationen beanspruchen und die stark die wahrgenommene Qualität beeinflussen: Länge der Außenohr-Impulsantworten und Anzahl der Spiegelschallquellen.

## Länge der Außenohr-Impulsantworten

Die Außenohr-Impulsantworten sind die berechnungsintensivsten Filter, die bei jeder Spiegelschallquellenberechnung durchlaufen werden müssen. In Abbildung 1 ist die Signalverarbeitungskette für eine Spiegelschallquellenberechnung der zweiten Ordnung dargestellt.

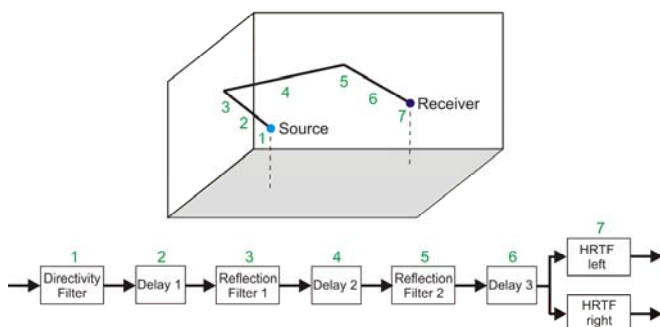


Abbildung 1: Signalverarbeitungskette für eine Spiegelschallquellenberechnung zweiter Ordnung, [2]

Die Wandreflexionsfilter können als IIR-Filter erster oder maximal zweiter Ordnung implementiert werden und benötigen damit 10..20 Multiply-Accumulate-Operationen. Die Außenohr-Impulsantworten oder Head-Related Impulse Responses (HRIR) haben nach ihrer Faltung mit den inversen Lautsprecher- und den inversen Kopfhörerübertragungsfunktionen eine effektive Länge von 256 Koeffizienten.

Durch eine Reduktion auf die Minimum-Phasen-Impulsantwort, die keine perzeptive Beeinträchtigung darstellt [4], konzentriert sich der Hauptenergieanteil am Beginn der Impulsantwort. Damit ist eine Verkürzung auf 128 Koeffizienten und weniger möglich, ohne dass ein signifikanter Fehler im Amplitudenfrequenzgang auftritt. Insbesondere für die berechneten Reflexionen kann darüber hinaus davon ausgegangen werden, dass die hochfrequente Feinstruktur in den HRIRs nicht relevant ist und deswegen eine weitere deutliche Verkürzung der FIR-Filter vorgenommen werden kann.

IKA-SIM [2], ein System zur Erzeugung einer auditiven virtuellen Umgebung, wurde in der Art erweitert, dass alle in dieser Untersuchung verwendeten Parameter in Echtzeit einstellbar zur Verfügung stehen. Somit ist jede Veränderung eines Parameters unmittelbar hörbar. Dies verkürzt die Untersuchungszeit bei einem großen Parameterraum drastisch im Vergleich zu Untersuchungsreihen mit vorproduzierten Tondateien.

Im Hörversuch wurde die Anzahl der Filterkoeffizienten pro Reflexionsordnung eingestellt. Die Versuchspersonen wurden angewiesen, die Wahrnehmbarkeitsschwelle für die kritischste Kombination aus Schallquellen- und Empfängerposition<sup>1</sup> zu finden. Die Einstellungen wurde sowohl für männliche Sprache (SQAM-CD) als auch für das von jedem Experten individuell als am kritischsten empfundene Signal durchgeführt. Zwei der drei mit dieser Art der Einstellung von auditiven virtuellen Umgebungen erfahrenen Testhörer entschieden sich für das Rauschpulssignal (300 ms weißes Rauschen, 300 ms Pause), der dritte wählte das Kastagnetsignal (SQAM-CD). Jegliche Änderung der räumlichen oder klanglichen Wahrnehmung im Vergleich zur Maximal-einstellung sollte vermieden werden. In Tabelle 1 sind die für jede Reflexionsordnung gefundenen Werte für die beiden unterschiedlichen Raumgrößen angegeben.

Ordnung	Kleiner Raum (70 m <sup>3</sup> )		Großer Raum (700 m <sup>3</sup> )	
	Sprache	Krit. Signal	Sprache	Krit. Signal
0	64	80	32	64
1	48	48	32	32
2	32	48	24	24
3	24	32	24	24
4	24	32	24	24

Tabelle 1: Maximal eingestellte Länge der gekürzten Außenohr-Impulsantworten pro Reflexionsordnung;  $f_s = 44100$  Hz

Da die hier eingestellten Filterlängen aus Implementierungsgründen die Eingangsverzögerung der Außenohr-Impulsantworten enthalten (im Mittel 6 Koeffizienten), bewegen

<sup>1</sup> Extreme Positionen nahe der Raumbegrenzungsflächen und Annäherungen zwischen Quelle und Empfänger kleiner als ein Meter wurden ausgeschlossen.

sich die Werte in einem sehr ähnlichen Bereich, wie diejenigen, die von Lehnert und Richter [5] gefunden wurden (20 Koeffizienten). Der kleinere Raum scheint in Bezug auf die Filterlängen etwas anspruchsvoller zu sein, insbesondere für die kritischen Signale. Die sich ergebende Gesamtanzahl der notwendigen Rechenoperationen pro Ordnung ist in der gestrichelten zweiten Kurve in Abbildung 2 aufgetragen.

## Auswahl von Reflexionen

Ein weiterer sehr aussichtsreicher Ansatz zur Rechenzeitoptimierung ist eine sinnvolle Auswahl und Beschränkung der zu berechnenden Spiegelschallquellen. Der erste Schritt besteht darin, nur diejenigen Spiegelschallquellen auszuwählen, die auch wahrgenommen werden (Irrelevanzreduktion). In einem zweiten Schritt wird eine überschwellige Reduktion der Spiegelschallquellen vorgenommen, die den unterschiedlichen Qualitätsansprüchen der unten aufgeführten Applikationen gerecht werden muss.

Eine Reduzierung der Anzahl der Spiegelschallquellen kann mit unterschiedlichen Strategien vorgenommen werden. Die folgenden Strategien wurden im Experiment untersucht:

- S1: Begrenzung der Ordnung der Spiegelschallquellen
- S2: Zeitliche Begrenzung nach dem Direktschall
- S3: Pegelgrenze unterhalb des Direktschallpegels.
- S4: Ausnutzung der Nachverdeckung

Die Nachverdeckung, bei der leise Tonsignale, die direkt hinter lauten Tonsignalen folgen, nicht mehr wahrgenommen werden, wird z.B. in [6] beschrieben. Binaurale Modelle, die diesen Effekt ausnutzen wollen, befinden sich noch in der Definitionsphase, z.B. [7]. Für die Reduzierung von Reflexionen wurde das folgende sehr vereinfachte monaurale Nachverdeckungsmodell untersucht. Die Nachverdeckungskurve wird durch eine e-Funktion mit einer Zeitkonstante  $\tau$  und einer gewissen Dauer  $T$  beschrieben. Sie wird nach jeder Reflexion mit einer Pegelabsenkung  $L_s$  und einem zeitlichen Abstand von 1 ms angesetzt. Die Reflexionen, die unterhalb dieser Kurve liegen, werden nicht auralisiert.

Die Parameter Reflexionsordnung, Pegelreduktion, Zeitreduktion und die drei Nachverdeckungsparameter wurden jeweils unabhängig voneinander für die folgenden drei Qualitätsszenarien eingestellt:

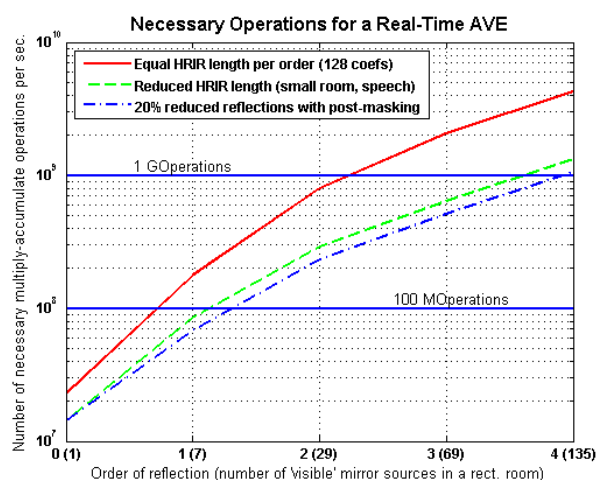
- Q0: Kein wahrnehmbarer Unterschied
- Q1: Hochqualitative Anwendung (Virtuelles Museum, Edutainment)
- Q2: Geringqualitative Anwendung (Virtueller Chat-Room)

Die jeweils kritischste Einstellung der drei Testhörer ist in Tabelle 2 aufgeführt. Bei dem hier vorgestellten einfachen Nachverdeckungsmodell ist es sehr schwierig eine positionsunabhängige Qualitätseinstellung zu finden. Darüber hinaus schwankt die Anzahl der unter die Verdeckung fallenden Reflexionen stark mit der Position und damit auch die eingesparte Rechenleistung. Als Richtwert kann eine mittlere Reduktion von 20% für die Wahrnehmbarkeitsschwelle Q0 angegeben werden.

Qualitätsszenario	Kleiner Raum ( $T_{60} \approx 0.3$ s)			Großer Raum ( $T_{60} \approx 0.6$ s)		
	Q0	Q1	Q2	Q0	Q1	Q2
S1: Reflexionsordnung	4	2	1	4	4	2
S2: Pegelgrenze [dB]	25	15	10	30	25	20
S3: Zeitgrenze [ms]	35	20	15	100	50	40
S4: Zeitkonstante $\tau$ [ms]	20	20	30	50	50	50
Dauer $T$ [ms]	7	15	15	15	20	20
Pegelabsenkung $L_s$ [dB]	4	3	3	4	1.5	1.5

**Tabelle 2: Unterschiedliche Methoden der Reflexionsreduktion für männliche Sprache, kritischste Werte**

Die sich ergebende Reduktion des Rechenaufwandes bei Filterkürzungen und dem Einsatz von Nachverdeckung wird in Abbildung 2 gezeigt. Die Pegel- oder Zeitreduktion sind weitaus ineffizienter als die Ordnungsreduktion, die ebenfalls dort abgelesen werden kann.



**Abbildung 2: Anzahl der notwendigen Rechenoperationen pro Sekunde für den kleinen Raum;  $f_s = 44100$  Hz**

Die aufgeführten Experteneinstellungen werden für weitere Parameter fortgeführt, um ein möglichst umfassendes Bild der Beziehungen zwischen den relevanten Qualitätselementen und Qualitätsmerkmalen einer interaktiven virtuellen auditiven Umgebung zu erhalten. Für die so gefundenen 'Quality of Service'-Einstellungen wird dann eine Verifizierung mit einer größeren Versuchspersonenanzahl durchgeführt.

## Literatur

- [1] Silzle, A. (2003). Welche Rechenleistung wird für die Echtzeitimplementierung von Virtueller Akustik benötigt?. DAGA 2003. Aachen
- [2] Silzle, A., P. Novo, and H. Strauss (2004). IKA-SIM: A System to Generate Auditory Virtual Environments. 116th AES Convention. Berlin, Germany
- [3] Silzle, A. (2005). Taxonomie und Optimierungsmethode für Systeme zur Generierung von auditiven virtuellen Umgebungen. DAGA '05. München, Germany
- [4] Minnaar, P., et al. (1999). Audibility of all-pass components in binaural synthesis. 106th Convention of the AES. Munich, Germany
- [5] Lehnert, H. and M. Richter (1995). Auditory Virtual Environment: Simplified Treatment of Reflections. 15th International Congress on Acoustics (ICA). Trondheim, Norway
- [6] Püschel (1988). Prinzipien der zeitlichen Analyse beim Hören (Diss.), 3. Physikalisches Institut. Universität Göttingen.
- [7] Buchholz, J., J. Mourjopoulos, and J. Blauert (2001). Room Masking: Understanding and Modelling the Masking of Room Reflections. 110th AES Convention. Amsterdam, The Netherlands