

HYBRIDES SPIEGELQUELLENMODELL - EINFLUSS DER WINKELGENAUIGKEIT

Boris Jung

Technische Universität Berlin, TUB
Boris.Jung@gmx.de

Christoph Schulze

Technische Universität Berlin, TUB
seackone@hotmail.de

ABSTRACT

Im Rahmen des Kurses "Virtuelle Akustik" an der TU Berlin wurde der Einfluss der azimuthalen Winkelauflösung verwendeter HRIRs auf die Qualität der damit erstellten binauralen Impulsantworten untersucht. Die BRIRs wurden mittels eines hybriden Spiegelquellen-Modells (ISM, "Image Source Modell") für kubische Räume erstellt.

Hierfür wurde zunächst das Modell in Matlab implementiert und mit Hilfe der zur Verfügung gestellten FABIAN HRIRs binaurale Impulsantworten erstellt. Mit diesen wiederum wurde hallarmes Audiomaterial für die vergleichenden Hörversuche aufbereitet.

Es zeichnete sich ab, daß die ursprüngliche Winkelauflösung der HRIR mit insgesamt 11345 Quellpositionen, reduziert werden kann, ohne merkliche Qualitätsverluste in Kauf nehmen zu müssen. Dies ermöglicht eine effizientere Nutzung des Spiegelquellenmodells (ISM).

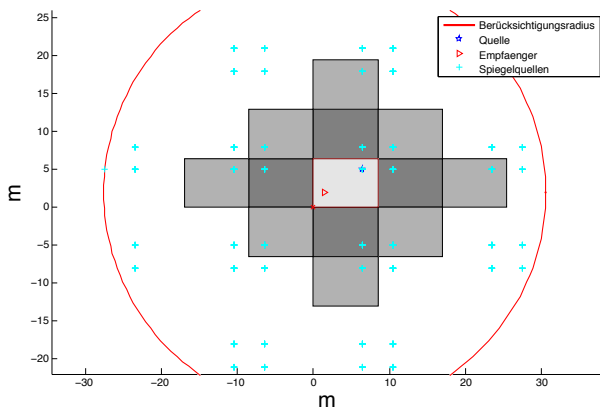


Figure 1: Spiegelquellenmodell - Positionen

1. MATLAB IMPLEMENTIERUNG

Zur Realisierung der Hörversuche, musste sowohl ein Spiegelquellenmodell für kubische Räume, als auch ein stochastic Reverb Modell in Matlab implementiert werden.

1.1. ISM - Positionen der Quellen

Abbildung 1 zeigt schematisch das Prinzip des implementierten Spiegelquellenmodells für einen bestimmten kubischen Raum. Zur Verdeutlichung wurden die "Spiegelräume" erster und zweiter Ordnung ebenfalls eingezeichnet.

Die Positionen des Hörers als rotes Dreieck, der Quelle in blau sowie der Spiegelquellen in Cyan verdeutlichen die sehr schnell zunehmende Anzahl an Quellen. Daher dient ein Eingabeparameter des Scripts, *dist_lim*, zur Begrenzung des Berücksichtigungsradius, hier in rot dargestellt, beziehungsweise, des maximalen Delays einer Spiegelquelle. (Resultierend aus der Entfernung und der Schallgeschwindigkeit)

Zur Berechnung der Positionen der einzelnen Spiegelquellen, wurde Formel 1 verwendet.

$$p_{ISM}(\vec{u}, \vec{l}) = (1 - 2\vec{u})\vec{x}_s + 2\vec{l}\vec{L} \quad (1)$$

mit $\vec{u} = (u, v, w) \in \{0, 1\}$, $\vec{l} = (l, m, n) \in \mathbb{Z}^3$, $\vec{L} = (L_x, L_y, L_z)$ (Raumlängen in m) und $\vec{x}_s = (x_s, y_s, z_s)$ (Quellposition in m)

Für die Berechnung der Verzögerung der einzelnen Spiegelquellen ist jedoch nur die relative Position zum Empfänger nach Gleichung 2 relevant.

$$p_{ISM,rel}(\vec{u}, \vec{l}) = p_{ISM}(\vec{u}, \vec{l}) - p_r = [x_{rel} \ y_{rel} \ z_{rel}]^T \quad (2)$$

Durch korrekte Wahl des Berücksichtigungsradius lässt sich der Rechenaufwand bereits deutlich reduzieren - sehr viele Quellen können vernachlässigt werden. Wie viele genau ist abhängig davon ob man ein reines ISM oder ein hybrides Modell (ISM kombiniert mit einem stochastischem Reverb) verwendet.

Bei Letzterem müssen nur die Quellen berücksichtigt werden, deren Pulse vor der sogenannten *perceptual mixing time* t_{mp50} plus der entsprechenden Fade Dauer, den Empfänger erreichen.

1.2. ISM - Berechnung der Impulsantworten

Mit Hilfe der relativen Positionen $p_{ISM,rel}(\vec{u}, \vec{l})$, lassen sich nun die Entfernungen der einzelnen Quellen zum Empfänger und damit die Verzögerung bestimmen. Gleichung 3 zeigt die Berechnung der Raumimpulsantwort.

$$RIR(t) = \sum_{\vec{u}=0}^1 \sum_{\vec{l}=-N}^N a(\vec{u}, \vec{l}) \cdot \delta(t - \tau(\vec{u}, \vec{l})) \quad (3)$$

Der Amplitudenfaktor $a(\vec{u}, \vec{l})$ ist abhängig von der Position der Spiegelquelle und den Reflektionskoeffizienten der zugehörigen Wände. Weiterhin wird die Abschwächung in Abhängigkeit der Entfernung einer idealen Punktquelle berücksichtigt.

$$a(\vec{u}, \vec{l}) = \frac{\beta_{x,1}^{|u-l|} \beta_{x,2}^{|l|} \beta_{y,1}^{|v-m|} \beta_{y,2}^{|m|} \beta_{z,1}^{|w-n|} \beta_{z,2}^{|n|}}{4\pi \cdot d(\vec{u}, \vec{l})} \quad (4)$$

mit $\vec{u} = (u, v, w) \in \{0, 1\}$, $\vec{l} = (l, m, n) \in \mathbb{Z}^3$

Formel 4 zeigt die Berechnungsgrundlage der Amplituden der einzelnen Quellen. In Abbildung 2 ist eine nach Formel 3 berechnete Raumimpulsantwort dargestellt, die gestrichelte Senkrechte markiert den Zeitpunkt t_{mp50} .

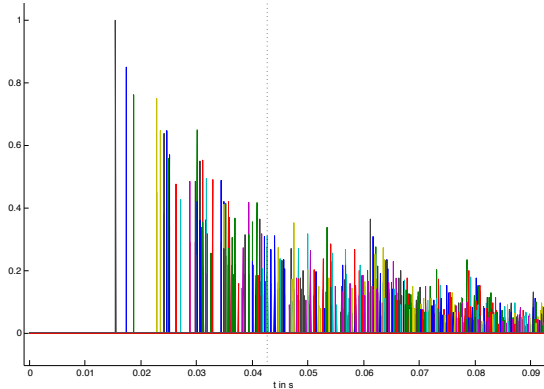


Figure 2: gewichtete Pulse der einzelnen Spiegelquellen

Um nun die binaurale Impulsantwort des Raums zu bestimmen, müssen die Anteile jeder einzelnen Spiegelquelle an der RIR mit der HRIR mit korrespondierendem Azimutal und Elevationswinkel gefaltet und anschließend aufsummiert werden, wie in Formel 5 dargestellt.

$$BRIR(t) = \sum_{\vec{u}=0}^1 \sum_{\vec{l}=-N}^N A(\vec{u}, \vec{l}, t) * HRIR(\varphi(\vec{u}, \vec{l}), \vartheta(\vec{u}, \vec{l}), t) \quad (5)$$

mit $A(\vec{u}, \vec{l}, t) = (a(\vec{u}, \vec{l}) \cdot \delta(t - \tau(\vec{u}, \vec{l})))$

und $\varphi(\vec{u}, \vec{l})$, $\vartheta(\vec{u}, \vec{l})$ den Azimutal-/Elevationswinkel der \vec{u}, \vec{l} zugehörigen Spiegelquelle.

Das Ergebnis einer solchen Berechnung einer BRIR für Raum 1 und Position 1 ist in Abbildung 3 ohne Normalisierung dargestellt.

Da HRIRs nicht beliebig fein aufgelöst sein können, müssen hier in den allermeisten Fällen die Winkel zwischen Empfänger- und Spiegelquellposition gerundet werden. Durch größeres Runden lässt sich die Datenmenge der HRIR reduzieren, ob und ab wann dies einen negativen Effekt auf die wahrgenommene Qualität der BRIR zeigt, soll in den Hörversuchen erörtert werden.

1.3. Späte Reflektionen - Stochastischer Reverb

Da die späten Anteile einer Raumimpulsantwort mit fortschreiten der Zeit immer undefinierter werden, kann auch hier der Rechenaufwand reduziert werden. Eine Möglichkeit hierzu, ist den späten

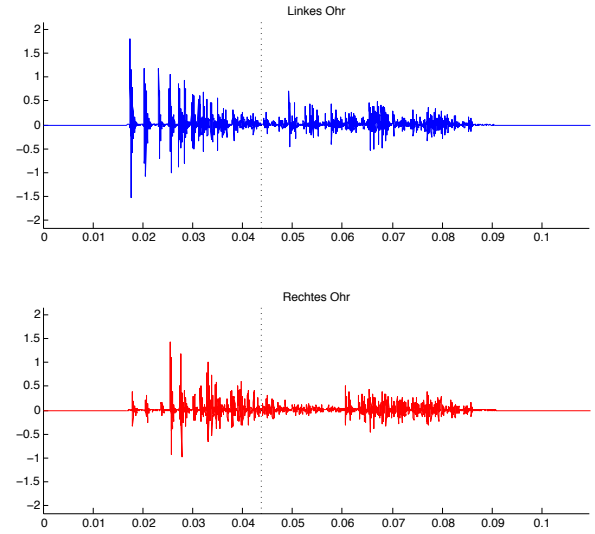


Figure 3: BRIR erzeugt mit reinem Spiegelquellenmodell

Teil einer BRIR nicht über das Spiegelquellenmodell sondern über einen sogenannten *stochastic reverb* zu erzeugen.

Ein weißes Rauschen, wurde mit Hilfe der *Auditory Modelling Toolbox* mit einer Rekonstruktionsfilterbank bearbeitet, die einzelnen Bänder gewichtet und summiert. Im Anschluss daran wird das Abklingen der RIR nach Formel 6 und 7 berücksichtigt.

$$h(t) = h_0 e^{-\delta t} \quad (6)$$

Wobei

$$\delta = \frac{cA}{8V} \quad (7)$$

und A der äquivalenten Absorptionsfläche und V dem Raumvolumen entsprechen.

Nach Borß und Martin¹ wurde die gewünschte Kohärenz der beiden Kanäle erzeugt, welche in Formel 8 dargestellt ist.

$$\hat{\gamma}_{x,y}(\omega) = \text{sinc}\left(\frac{\omega}{\omega_0}\right) \max\left(0, 1 - \frac{\omega}{\omega_1}\right) \quad (8)$$

Hierbei entspricht $\omega_0 = 2\pi 550\text{Hz}$ einem ungefähren Ohrabstand von 10cm. $\omega_1 = 2\pi 2700\text{Hz}$ ist ein manuell optimierter Wert, um die Kohärenz bei höheren Frequenzen zu verringern.

Die beiden unabhängigen Rauschsignale $x(t)$ und $y(t)$ wurden somit nach dem Rekonstruktionsfilter und der Berücksichtigung des Abklingens einer RIR, nach Formel 9, 10 und 11 bearbeitet, um die gewünschte Kohärenz nach Formel 8 zu erhalten. Abbildung 4 zeigt die so erstellten Rauschsignale.

$$\begin{aligned} X_{\hat{\gamma}}(\omega) &= H_{\alpha}(\omega)X(\omega) + H_{\beta}(\omega)Y(\omega) \\ Y_{\hat{\gamma}}(\omega) &= H_{\alpha}(\omega)Y(\omega) + H_{\beta}(\omega)X(\omega) \end{aligned} \quad (9)$$

¹Borß, Christian and Rainer Martin. *An improved parametric Model for perception-based design of virtual acoustics*. In: AES 35th International Conference. London, UK, 2009

2. HÖRVERSUCHE

$$H_{\beta}(\omega) = \begin{cases} \sqrt{\frac{1}{2} (1 - \sqrt{1 - \hat{\gamma}_{x,y}^2(\omega)})} & \text{für } \hat{\gamma}_{x,y}(\omega) \geq 0 \\ -\sqrt{\frac{1}{2} (1 - \sqrt{1 - \hat{\gamma}_{x,y}^2(\omega)})} & \text{für } \hat{\gamma}_{x,y}(\omega) < 0 \end{cases} \quad (10)$$

$$H_{\alpha}(\omega) = \sqrt{1 - H_{\beta}^2(\omega)} \quad (11)$$

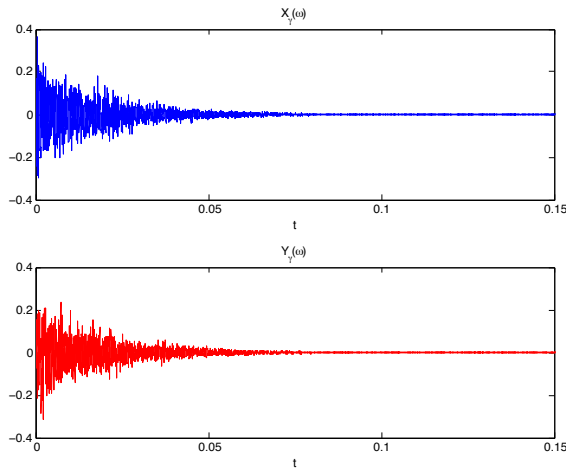


Figure 4: Erstellte Rauschsignale mit definierter Kohärenz

1.4. Kombination der beiden Anteile

Um nun mit Hilfe der beiden Modelle eine hybride BRIR zu erstellen, muss der Zeitpunkt des Übergangs von einem zum andern Modell bestimmt werden. Nach Lindau et al.² sei der Quotient aus Raumvolumen und Absorptionsoberfläche ein guter Anhaltspunkt zur Bestimmung der *perceptual mixing time* t_{mp50} , also dem Mix-Zeitpunkt, bei dem zwischen einem hybriden Modell und einem reinem ISM nicht mehr differenziert werden kann. Formel 12 stellt die von Lindau et al. optimierte Berechnung dar.

$$t_{mp50} = 20 \frac{V}{S} + 12 [ms] \quad (12)$$

Implementiert wurden drei verschiedene Mix Typen: hardcut, linearer Fade und logarithmischer Fade. Alle starten bei der in Formel 12 berechneten mixing time.

Nach Borß und Martin sollten bereits vor t_{mp50} in sehr geringem Maße Anteile des stochastic Reverbs beigemischt werden, um einen natürlicheren Klang zu erzeugen. In der erstellten Implementierung ist dies über die Variable *early factor* regelbar, welche standardmäßig auf 0.05 eingestellt ist, und somit auch schon vor der perceptual mixing time 5% des Rauschsignals beimischt.

²Lindau, Alexander; Linda Kosanke; and Stefan Weinzierl. *Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses*. In: J. Audio Eng. Soc., 60(11), pp. 887-898, 2012.

Der Einfluss der azimuthalen Winkelauflösung der HRIR auf die Qualität der BRIR sollte mittels Hörversuchen geprüft werden. Um diese durchführen zu können, musste im Voraus Audiomaterial auf bestimmte Weise aufbereitet werden um mit Hilfe der Ergebnisse der Hörversuche qualitative Aussagen zu erhalten. Ziel war eine grobe Eingrenzung des Bereichs der Winkelauflösung, ab dem deutliche Unterschiede hörbar sind.

Table 1: Raum Abmessungen

	Raum 1	Raum 2
Länge (in m)	8.5	4.5
Breite (in m)	6.5	4
Höhe (in m)	3.5	2.5

2.1. Vorbereitung

Um festzustellen welchen Einfluss die azimuthale Winkelgenauigkeit bei der Auswahl der HRIR auf die BRIR hat, wurden stichprobentypisch vergleichende Hörversuche durchgeführt. Es wurden die BRIRs zwei verschiedener Räume mit jeweils drei unterschiedlichen Quell-/Empfängerpositionen mit verschiedener Azimutalaufösung erzeugt. Die Abmessungen der beiden Räume sowie die Reflektionskoeffizienten, welche für beide Räume benutzt wurden, sind in den Tabellen 1 und 2 aufgeführt.

Table 2: Reflektionskoeffizienten

β_{x1}	β_{x2}	β_{y1}	β_{y2}	β_{z1}	β_{z2}
0.8	0.6	0.6	0.6	0.7	0.5

Für jeden Raum wurden 3 BRIRs mit unterschiedlichen Quell- und Empfängerpositionen erzeugt. Der Berücksichtigungsradius, welcher mit Hilfe des Parameters *dist_lim* definiert wird, wurde jeweils so gewählt, dass die Anzahl der berücksichtigten Quellen im gleichen Bereich lag. Positionen und Anzahl berücksichtigter Quellen sind für die beiden Räume in den Tabellen 3 und 4 aufgeführt.

Table 3: Quell- und Empfängerpositionen Raum 1

	Pos 1	Pos 2	Pos 3
x_r	1.5	1.5	1.5
y_r	2	2	2
z_r	1.7	1.7	1.7
x_s	6.5	2	2.5
y_s	5	5	1.5
z_s	1.7	1.7	1.7
dist_lim	5	10	25
Quellen	525	606	480

Die zur Erstellung der BRIRs genutzten HRIRs, eine Freifeld Näherung mit dem Kopf- und Torsosimulator FABIAN, lagen uns

Table 4: Quell- und Empfängerpositionen Raum 2

	Pos 1	Pos 2	Pos 3
x_r	1.2	1.2	2.25
y_r	0.8	0.8	2
z_r	1.7	1.7	1.7
x_s	4	1.1	2.5
y_s	3.5	0.8	1.5
z_s	1.7	1.7	1.7
dist_lim	5	35	10
Quellen	657	511	693

in einem 2° Raster für die Elevation und unterschiedlichen azimuthalen Auflösungen vor. Um nun zu prüfen ob wirklich der volle Datensatz notwendig ist, wurde die Auflösung in mehreren Schritten reduziert. Die originale und resultierenden azimuthalen Schrittweiten in Grad sind in Tabelle 5 aufgelistet. Als Vergleich wurden ebenfalls für alle Raum Setups Versionen mit 0° für Elevation und Azimut, sowie mit original Elevation und 0° Azimut erstellt.

Table 5: Azimutale Schrittweiten der verwendeten HRIR

Elevation	original	high	medium	low	very low
90	360	360	360	360	360
88	45	45	45	45	45
86, 84	18	18	18	18	36
82, 80	10	10	10	20	30
78	9	9	9	18	36
76...72	6	6	12	24	30
70, 68	5	5	10	20	30
66...50	3	3	9	18	30
48...-48	2	4	10	20	30
-50...-64	3	3	9	18	30

Somit wurden für beide Räume mit je drei Positionen fünf BRIRs mit unterschiedlicher azimuthaler Auflösung der Spiegelquellen erstellt, sowie die beiden Kontroll BRIRs mit 0° und $0^\circ/0^\circ$. Um die Hörversuche realisierbar zu machen, wurden mit diesen BRIRs je zwei verschiedene hallarme Audiosignale gefaltet. Hierbei handelte es sich um eine Sprachaufnahme und um einen kurzen Schlagzeugpart. Somit resultierten insgesamt 84 Audiodateien, welche anschließend im Hörversuch, in ausgewählten Paaren, verglichen wurden.

2.2. Durchführung

Da ein großangelegter Hörversuch den Rahmen des Kurses sprengen würde, wurden nur stichprobenartig die verschiedenen Signale verglichen. Dies geschah sowohl durch die beiden Autoren, welche bewusst ausgewählte Auflösungen miteinander verglichen, als auch in Blindtests mit neutralen Versuchspersonen. Diese sollten nach dem Hören zweier Signale, angeben, ob ein Unterschied hörbar war.

2.3. Auswertung

Grundsätzlich vermittelten die bearbeiteten Audiosignale eine gute Räumlichkeit - verschiedene Raumgrößen ließen sich leicht unterscheiden. Desweiteren war auch bei gleicher Raumgröße deutlich wenn sich die Quell-Position mit gleichem aber auch unterschiedlichem Abstand zum Empfänger veränderte. Eine Ortung vorne-hinten konnte relativ deutlich am dumpferen Klang vollzogen werden. Rechts-links konnte ebenfalls geortet werden, allerdings entstanden hier teilweise Fehleindrücke beispielsweise aufgrund von geringer Wandentfernung des Empfängers.

Die ersten Hörversuche fanden nur mit den ersten drei Reduktionen der Winkelauflösung statt. Hier konnte kein eindeutiges Ergebnis erzielt werden. Bei den Kontrollversuchen mit den 0° und $0^\circ/0^\circ$ BRIRs, wurde jedoch klar, dass der räumliche Eindruck stark von den Winkeln der genutzten HRIR abhängig sein muss. So fiel der Raumeindruck in diesen beiden Versionen fast vollständig in sich zusammen und klang - erwarteterweise - sehr frontal.

Nachdem das ursprüngliche Konzept um die weitere Winkelauflösungsstufe *very low* erweitert wurde, konnten bei nahezu allen Hörversuchen, dieser neuen Auflösung ein deutlich anderer Raumklang zugeschrieben werden.

Weiterhin wurden abwechselnd die Versionen *original* und *low*, sowie *low* und *very low* verglichen. Die Ergebnisse deuteten klar daraufhin, dass der Unterschied erst bei einem Bereich der Auflösung der zwischen den untersten beiden Stufen liegt, hörbar wird. Dies war bei beiden Räumen und allen Positionskonstellationen der Fall. Trotzdem ist nicht auszuschließen, daß der Schwellwert der Winkelauflösung auch von diesen Faktoren beeinflusst wird. Da das gesamte Hörversuchsmaterial mit den gleichen Reflexionsfaktoren erstellt wurde, ist auch hier eine Beeinflussung nicht ausgeschlossen.

3. FAZIT

Es wurde ein Matlab Script, bestehend aus mehreren Unterfunktionen, erstellt, mit dem ein hybrides Spiegelquellenmodell für kubische Räume beliebiger Abmessungen mit frequenzunabhängigen Reflexionskoeffizienten berechnet werden kann. Die Anzahl der zu berücksichtigenden Quellen kann über den Berücksichtigungsradius und dem Summenindex N definiert werden.

Für die Art der Überblendung zwischen Spiegelquellen- und stochastic Reverb-Anteil kann zwischen drei verschiedenen Modi gewählt werden.

Weiterhin wurde eine Funktion erstellt, welche monaurales Audiomaterial mit den zuvor erzeugten binauralen Impulsantworten faltet und je nach Eingabeparameter abspielt und/oder im *wav* Format abspeichert.

Hallarmes Audiomaterial wurde für die Hörversuche auf diese Weise mit unterschiedlichen BRIRs aufbereitet. Die Winkelauflösung der verwendeten HRIRs wurden zunächst von 2° bis auf 20° reduziert, jedoch ohne deutliche Ergebnisse zu liefern. Die Signale, welche mit schlechter aufgelösten BRIRs erzeugt wurden, konnten nicht zuverlässig identifiziert werden.

Zur weiteren Kontrolle wurden zwei BRIRs, basierend auf nur einer HRIR beziehungsweise auf nur einem azimuthal Winkel, untersucht. Das damit aufbereitete Material konnte eindeutig unterschieden werden.

In einem weiteren Versuch wurde die Winkelauflösung auf 30° reduziert. Hier wurden bei fast allen Hörversuchen die Unter-

schiede richtig identifiziert. Eine mögliche Schlussfolgerung hieraus wäre, daß der Schwellwert der Winkelauflösung der HRIR für einen hörbaren Unterschied, in den hier untersuchten Konstellationen von Raumparametern (Maße, Reflektionskoeffizienten) und Positionsparametern (Quell-/Empfängerpositionen, Wandabstände), im Bereich zwischen 20° und 30° liegen sollte.

Weiterhin zu klären wäre, in welchem Maße andere Faktoren diesen Schwellwert beeinflussen. Neben den bereits genannten Raum- und Positionsparametern, könnten hier auch Abhängigkeiten von HRIR Parametern oder dem verwendeten Audiomaterial vorhanden sein. Generell wird erwartet, dass mit steigender Komplexität des Modells, beispielsweise durch Einführung weiterer Objekte im Raum oder durch kompliziertere Raumformen, der Schwellwert eher steigen wird.

Es zeichnete sich jedoch die klare Tendenz ab, dass für die meisten einfachen Fälle, der Datensatz von über 11.000 HRIRs deutlich reduziert werden kann ohne merkliche Qualitätseinbußen zu verzeichnen. In genau welchem Maß und Kontext, muss Gegenstand zukünftiger Untersuchungen sein.