

Implementierung und Evaluation einer akustischen Raumsimulation für audiologicalhe Testungen

Zusammenfassung

In alltäglichen Hörsituationen ist die akustische Kommunikation von verschiedenen, anspruchsvollen akustischen Umgebungen geprägt. In der Regel treten Nebengeräusche auf und zusätzlich wird der Schall in Räumen von Wänden und Gegenständen reflektiert. Reflexionen haben einen Einfluss auf die Diskrimination von Sprachinformationen, jedoch werden audiologicalhe Routinemessungen weitgehend in Messkabinen mit geringer Nachhallzeit durchgeführt. Das Ziel dieser Arbeit war die Implementierung und Evaluation eines Raumsimulationssystems zur Untersuchung der Hörleistung in alltagsnahen Hörsituationen mit unterschiedlicher Raumakustik und Nachhall.

Mit Hilfe einer Raumsimulations-Software wurde die Schallausbreitung in Abhängigkeit der Raumgeometrie und der Raumakustik berechnet. Über ein Multikanal-Wiedergabesystem mit 128 Lautsprechern werden den Ergebnissen der Simulation folgend der Anteil des Direktschalls und der Raumreflexionen wiedergegeben. Somit können beliebige Räume unter realitätsnahen und kontrollierten akustischen Bedingungen im Labor rekonstruiert werden.

Das Verfahren wurde anhand eines Raummodells eines leeren Hörsaals für verschiedene Absorptionsgrade evaluiert. Technische Messungen der Nachhallzeit (RT) und des Deutlichkeitsgrades (D50) zeigten gute Übereinstimmung mit den Modellvorgaben. Zusätzlich wurden interaurale Pegeldifferenzen (ILDs) aufgezeichnet und in Bezug zu Sprachverständlichkeitsschwellen gebracht. Es zeigte sich, dass die ILDs mit zunehmendem Nachhall um bis zu 10 dB abnehmen. Gleichzeitig ist im Vergleich zwischen Freifeld und Nachhallkondition eine Verschlechterung des spatial release from masking (SRM) um bis zu 5,5 dB SNR zu verzeichnen.

Die Untersuchung der Diskrimination von Sprache mittels des vorgestellten Raumsimulations-Systems erweist sich als sinnvolle Ergänzung zu etablierten audiologicalhen Messverfahren. Ein unter bestimmten raumakustischen Bedingungen erzielttes Sprachverstehen kann reproduzierbar und zuverlässig untersucht werden.

Schlüsselwörter: Raumsimulation, Nachhall, Sprachverstehen, räumliche Entmaskierung, SRM, interaural level difference, ILD

Einleitung

Im Alltag treten häufig komplexe Hörsituationen mit Nebengeräuschen auf. Zusätzlich wird in geschlossenen Räumen der Direktschall mit Nachhall überlagert. Typische Hörsituationen im Alltag sind beispielsweise ein Stimmgewirr vieler Sprecher in einem Restaurant, am Bahnhof oder im Supermarkt. Das Sprachverstehen hängt sehr stark von Anzahl und räumlicher Verteilung der Störgeräuschquellen ab [1]. Da in der Diagnostik und Therapie von Hörstörungen sowie bei der Anpassung und Kontrolle von Hörsystemen üblicherweise Testverfahren in akustisch optimierten Hörkabinen mit sehr niedriger Nachhallzeit durchgeführt werden, kann die tatsächliche Hörleistung im Alltag nicht abgebildet werden. Hier wer-

den weiterhin nur ein bis zwei Schallquellen (Lautsprecher) eingesetzt, wodurch die möglichen Einfallsrichtungen von Nutz- (d.h. Sprache) und Störschall limitiert sind. Simulationsverfahren bieten die Möglichkeit, reale Hörsituationen nachzubilden um die Hörleistung von Menschen mit und ohne Hörminderung genauer untersuchen zu können [2], [3], [4].

Die Summe aller Reflektionen innerhalb eines abgegrenzten Raumes wird als Nachhall definiert [5]. Dabei wird die abgestrahlte akustische Information einer Schallquelle in geschlossenen Räumen von diesen Reflexionen überlagert. Das am Empfänger aufgenommene Signal lässt sich in Direktschall, frühe und späte Reflektionen einteilen. Frühe Reflexionen treten bis zu 50–80 ms nach der ersten Schallwelle auf und erweisen sich teilweise als

Anja Eichenauer¹

Uwe Baumann¹

Tobias Weißgerber¹

¹ Audiologische Akustik, Klinik für HNO-Heilkunde, Universitätsklinikum Frankfurt am Main, Deutschland

nützlich für das Sprachverstehen, da sie als Verstärkung der Information des Direktschalls angesehen werden können [6]. Späte Reflexionen treten mit größerer Verzögerung auf und werden als diffuse „Hallfahne“ oder als separates Echo wahrgenommen und können zu einer Verschlechterung des Sprachverstehens führen. Die akustischen Eigenschaften eines Raumes werden durch verschiedene Maße wie z.B. Nachhallzeit (engl. „reverberation time“, RT), Deutlichkeitsmaß (C50), Klarheitsmaß (C80), Deutlichkeitsgrad (D50) oder Stärkemaß (G) ausgedrückt. Die Nachhallzeit beschreibt dabei die Zeit, die benötigt wird, bis der Schalldruckpegel eines Schalls auf ein Tausendstel seines Ausgangswertes (d.h. um 60 dB) gesunken ist. Das Deutlichkeitsmaß C50 beschreibt das Verhältnis zwischen der Schallenergie im Zeitraum bis 50 ms nach dem Direktschall und der Schallenergie ab dem Zeitpunkt 50 ms. Das Klarheitsmaß C80 ist definiert durch das Verhältnis zwischen der Schallenergie in der Zeit bis 80 ms nach dem Direktschall und der Schallenergie ab dem Zeitpunkt 80 ms. Der Deutlichkeitsgrad D50 beschreibt das Energieverhältnis zwischen den frühen Reflexionen der ersten 50 ms zur gesamten Schallenergie. Das Stärkemaß G setzt den Schallpegel einer kugelförmig abstrahlenden Schallquelle ins Verhältnis zum Pegel derselben Quelle in 10 m Abstand. Die genannten raumakustischen Kenngrößen werden unter anderem durch die Größe (d.h. Volumen) des Raumes und die akustischen Eigenschaften der Begrenzungsflächen wie Decken, Boden, Wänden, etc., beeinflusst.

Je nach akustischen Eigenschaften eines Raumes wird die zeitliche und spektrale Struktur des Quellsignals verändert. Die Überlagerung des Direktschalls mit Nachhall kann mit der Wirkung eines Störgeräusches verglichen werden und zusätzlich die Modulationstiefe des Nutzsignals verringern. Aufgrund der Abhängigkeit zwischen Raumgröße/-beschaffenheit und Sprachverstehen werden nach DIN 18041:2016-03 (Hörsamkeit in Räumen) geeignete Nachhallzeiten abhängig vom Volumen des Raumes empfohlen [7]. Je nach Art der Nutzung gibt es somit unterschiedliche Empfehlungen für Musik, Sprache, Unterricht, etc. Bei einem Raumvolumen von 1.000 bis 5.000 m³ liegen die optimalen Nachhallzeiten für Unterricht bei ca. 0,75 bis 1 s, bei inklusivem Unterricht hingegen sollte die Nachhallzeit bei angegebenem Raumvolumen 0,6 bis 0,8 s nicht überschreiten.

In dieser Arbeit wird die Weiterentwicklung eines bereits bestehenden Freifeldwiedergabesystems mit 128 Lautsprechern [8] für reproduzierbare, flexible und möglichst plausible Wiedergabe von Schallen in Räumen unterschiedlicher akustischer Beschaffenheit beschrieben. Die Evaluation beinhaltet die messtechnische Analyse raumakustischer Größen sowie von interauralen Pegelunterschieden in drei Nachhallkonditionen. Weiterhin soll der Einfluss von Nachhall auf das Sprachverstehen im Störgeräusch unter Betrachtung der räumlichen Entmaskierung (engl. „spatial release from masking“, SRM) bei normalhörenden Probanden untersucht werden.

Material und Methoden

Aufbau des Raumsimulations-Systems

Für den Aufbau des Simulationssystems stand am Schwerpunkt Audiologische Akustik der HNO Universitätsklinik Frankfurt am Main ein reflexionsarmer Raum zur Verfügung. Dieser Raum ist mit einem Wiedergabesystem mit 128 Lautsprechern (Rechteckarray in der Horizontalebene) ausgestattet, welches bisher entweder unter Freifeldbedingungen (z.B. [8], [9], [10]) oder zur richtungsunabhängigen Simulation von Diffusschall [11] genutzt wurde. Die bestehende Anlage wurde um eine Komponente zur Raumsimulation erweitert.

Raummodell

Für die Raumsimulation wurde in der Software ODEON (ODEON A/S, Lyngby, Dänemark) ein dreidimensionales Modell eines leeren Hörsaals (siehe Abbildung 1) mit einem Raumvolumen von 3.520 m³ verwendet. Die Höhe des Raums beträgt 7,5 bis 10 m. Dieses Raummodell ist in der Software frei verfügbar (Dateiname: Example.par). In ODEON können Objekten und Wänden Absorptionseigenschaften bzw. spezifische Materialeigenschaften zugeordnet werden. Basierend auf der genannten Raumgeometrie wurden zwei Raummodelle mit unterschiedlichen Absorptionseigenschaften aller Raumflächen (d.h. Wände, Decke, Boden) erstellt. Die frequenzunabhängigen Absorptionsgrade betragen 80% und 60%. Dabei entspricht 80% Absorption einer mittleren Nachhallzeit von 350 ms und 60% einer Nachhallzeit von 510 ms.

Es wurden drei Schallquellen im vorderen Halbraum der Horizontalebene bei -60° , 0° und 60° relativ zum Empfänger positioniert. Dabei hatte die frontale Schallquelle aus 0° einen Abstand von 5 m zur Hörposition, die seitlichen Schallquellen waren 10 m von der Hörposition entfernt. Abbildung 1 zeigt in der linken Abbildung eine Skizze der Raumgeometrie des Raumes mit drei Schallquellen und einem Empfänger.

Im Raummodell wurde die Schallausbreitung eines vom Sender richtungs- und frequenzunabhängig abgestrahlten Signals zu den Empfängern berechnet (Spiegelquellenmodell und „ray-radiosity-Verfahren“, [12]). Es wurden Reflexionen bis zur 10. Ordnung berücksichtigt und in einem Reflektogramm dargestellt. Für jede Reflektion beinhaltet das Reflektogramm die zeitliche Verzögerung der Reflektionen, die Oktavpegel zwischen 63 und 8.000 Hz, die horizontalen Schalleinfallswinkel sowie die vertikalen Schalleinfallswinkel. Abbildung 1 rechts zeigt grafisch die Schallausbreitung von Quelle P1 ausgehend. Der abgestrahlte Schall trifft bei seiner Ausbreitung mehrmals auf Raumbegrenzungen und wird wiederholt reflektiert. Der Direktschall wird in dunkelrot und Reflexionen erster Ordnung in grün dargestellt.

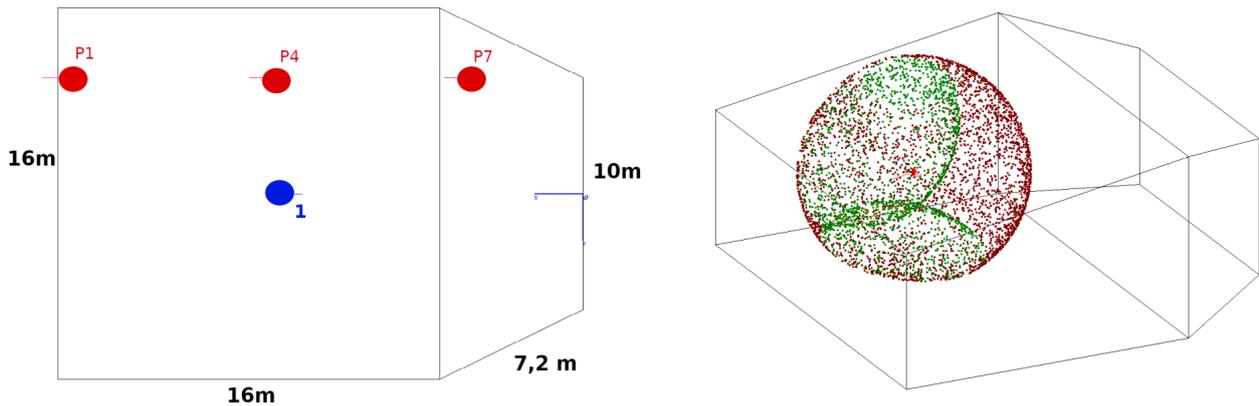


Abbildung 1: Linke Abbildung: Blick von oben auf das Hörsaalmodell mit Empfänger, blauer Punkt (1) und Schallquellen: P4 (0°) und P1/P7 ($\pm 60^\circ$); rechte Abbildung: Schallausbreitung im dreidimensionalen Hörsaalmodell von Quelle P1

Raumsimulation

Basierend auf den mittels ODEON berechneten Reflektogrammen wurde die Raumsimulation im reflexionsarmen Raum realisiert. Abbildung 2 zeigt eine Skizze des Labors mit den Positionen der Schallquellen.

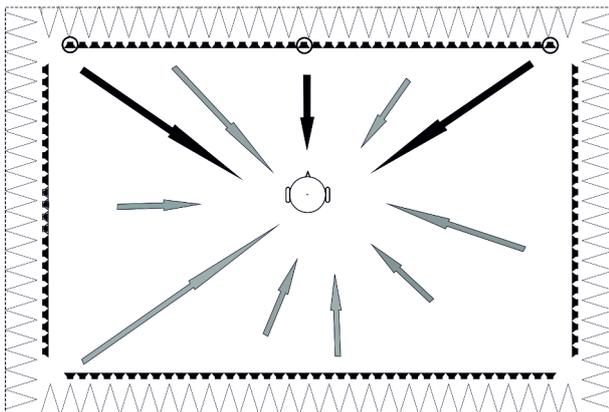


Abbildung 2: Anordnung der 128 Lautsprecher des Wiedergabesystems im reflexionsarmen Raum

Mittelpunkt (Hörposition) mit skizzierterm Probanden; Kreise: Lautsprecher, deren Position den simulierten Schallquellen P1, P4 und P7 entsprechen. Der Direktschall wird durch schwarze Pfeile skizziert, frühe Reflexionen und diffuser Hall durch graue Pfeile.

Es wurde das Verfahren des am nächsten gelegenen Lautsprechers angewendet [13], [14]. Dies bedeutet, dass bei der Wiedergabe für jede berechnete Reflexion der Lautsprecher mit dem geringsten Abstand im Horizontalwinkel gewählt wird. Reflexionen mit einem vertikalen Einfallswinkel größer als $\pm 30^\circ$ wurden nicht berücksichtigt, da es sich im Wiedergaberaum um eine horizontale Lautsprecheranordnung handelt. Während eine vollständige Integration (d.h. Mapping) aller Elevationswinkel auf die Horizontalebene einen vollständigen Erhalt der Reflexionsdichte der Erstreflexionen ermöglicht, führt diese Variante zu sehr deutlich hörbaren Klangverfärbungen. Als Kompromiss zwischen Reflexionsdichte und Klangverfärbung haben wir deshalb nach subjektiven Tests in

Vorversuchen beschlossen, vertikale Einfallswinkel bis $\pm 30^\circ$ über die horizontalen Lautsprecher wiederzugeben. Basierend auf dem Reflektogramm wurde für jede Reflexion entsprechend ihrer zeitlichen Verzögerung und ihres Spektrums eine Impulsantwort generiert.

Im nächsten Schritt wurde jeder Reflexion der im Horizontalwinkel nächstliegende Lautsprecher zugeordnet und die Impulsantworten je Lautsprecher summiert. 80 ms nach dem Direktschall wurden späte Reflexionen mittels eines statistischen Modells (Feedback-Delay-Network, FDN, bezogen von den Quality & Usability Labs, Institut für Softwaretechnik und Theoretische Informatik, Technische Universität Berlin) als diffuser Nachhall eingeblendet. Hierzu wurde im Zeitfenster zwischen 80 ms und 120 ms nach Direktschall der RMS-Pegel aller generierten Reflexionen bestimmt. Der diffuse Hall wurde im Pegel an die zuvor generierten Reflexionen angepasst. Der diffuse Nachhall wurde unter Einbeziehung der frequenzabhängigen RT generiert und gleichermaßen auf alle 128 Kanäle eingeblendet. Zum Ausgleich der aufgrund der Rechteckanordnung individuellen Übertragungseigenschaften der Lautsprecher zur Raummitte (Amplitudenfrequenzgang und Laufzeit) wurde für jeden Lautsprecher individuell eine Entzerrung durchgeführt („finite impulse response“ (FIR) Filter).

Die generierte 128-kanalige Raumimpulsantwort („room impulse response“, RIR) wurde für die Schallwiedergabe mit dem gewünschten einkanaligen Audiosignal gefaltet. Die Darbietung des Raumklangs erfolgte simultan über alle 128 Lautsprecher. Es wurden RIRs beider Konditionen mit Nachhall, für je drei Direktschallpositionen ($\pm 60^\circ$, 0°) erzeugt. Abbildung 3 zeigt ein Flussdiagramm der Raumsimulation mit Direktschall, Erstreflexionen und diffuserm Hall.

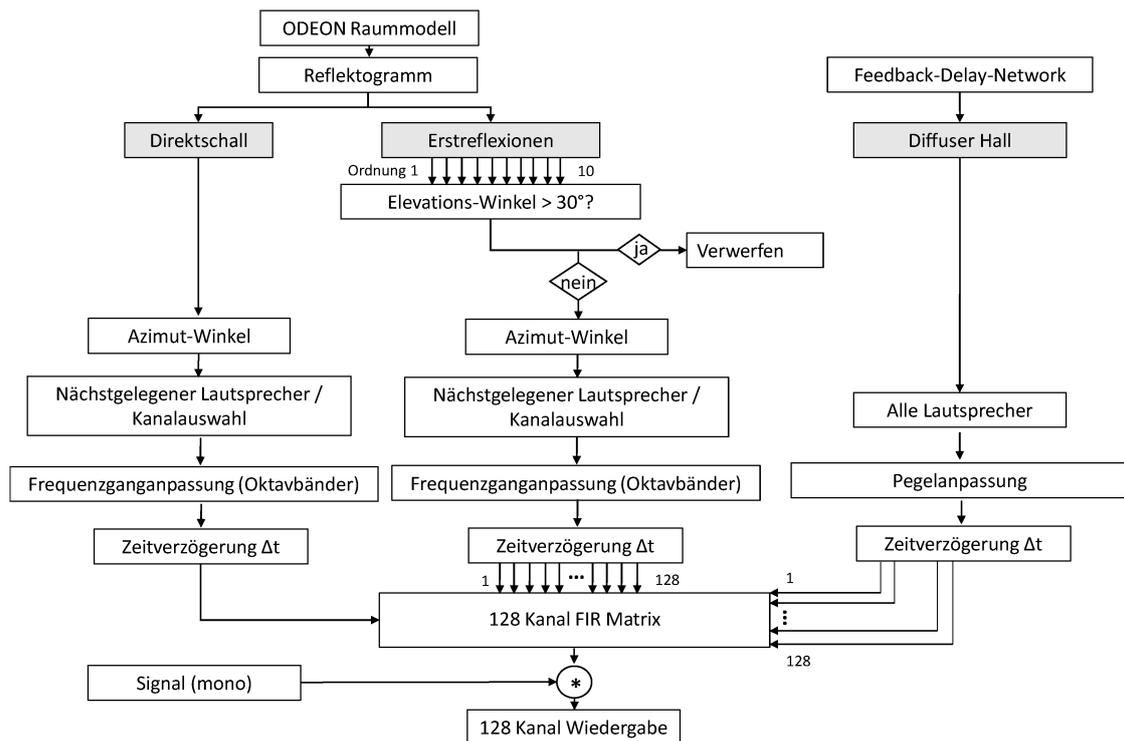


Abbildung 3: Flussdiagramm der Raumsimulation mit Direktschall, Erstreflexionen und diffusem Hall.

Evaluation des Raumsimulations-Systems

Um die Qualität der Simulation zu überprüfen, wurden im ersten Schritt RT und D50 messtechnisch ermittelt und mit den Modelldaten verglichen. Im zweiten Schritt wurde das System unter audiologischen Gesichtspunkten evaluiert, indem Sprachtests durchgeführt und die Ergebnisse einer Bewertung unterzogen wurden. Hierbei wurde der Einfluss der Nachhallzeit auf das Sprachverstehen und SRM untersucht. Zusätzlich wurden mit einem Kunstkopf aufgenommene kopfbezogene (engl. „head-related“) IRs (HRIRs) auf interaurale Pegeldifferenzen (engl. „interaural level differences“, ILDs) untersucht und in Relation zum SRM betrachtet.

Technische Evaluation

Material und Methode

Die RT und D50 wurden sowohl pro Oktavfrequenzband von 63 Hz bis 8.000 Hz als auch breitbandig gemittelt analysiert und mit den Berechnungen des Raummodells verglichen. Die Analyse erfolgte für eine Schallquelle bei 0°. Als Stimulus wurde ein Sinussweep (Frequenzbereich von 50 Hz bis 22.050 Hz) der Dauer 0,5 s genutzt. Die Messungen an der Hörposition erfolgte mit einem Messmikrofon vom Typ 4155 (Brüel & Kjær, Nærum, Dänemark) am Impedanzwandler 2.669 und Messverstärker Nexus des gleichen Herstellers. Die Impulsantworten wurden mit der Akustiksoftware ARTA (Artalabs, Kastel Luksic, Kroatien) untersucht.

Zusätzlich wurden die IRs am Kopf- und Rumpfsimulator (Brüel & Kjær Type 4100 mit Mikrofonen Type 4190-L-002) gemessen. Die Schalldarbietung erfolgte aus 60°. Aus den Ergebnissen der Messungen wurden ILDs abgeleitet. Alle Messungen wurden in beiden simulierten Konditionen mit Absorptionsgraden 60% und 80% und unter Freifeldbedingungen (d.h. ~100% Absorption) durchgeführt.

Ergebnisse

Für die Freifeldmessung ergab sich eine mittlere RT von 0,05 s. Die RT-Werte pro Oktavband sind für Messkondition 2 mit 80% Absorption und Messkondition 3 mit 60% Absorption in Tabelle 1 angegeben. Die durch das Modell mittels ODEON ermittelte RT lag in der Kondition mit 80% Absorption im Mittel über alle Frequenzen bei 0,35 s; für 60% Absorption bei 0,51 s. Die Messungen mit dem Simulationssystem im reflexionsarmen Raum ergaben bei 80% Absorption 0,39 s und bei 60% Absorption 0,58 s Nachhallzeit. Es zeigen sich die größten Abweichungen in den tieffrequenten Bändern, insbesondere im Bereich und unterhalb der Grenzfrequenz des reflexionsarmen Raumes (~200 Hz).

Tabelle 2 zeigt beispielhaft für Messkondition 2 (80% Absorption) und Messkondition 3 (60% Absorption) das D50-Maß pro Oktavfrequenz. Die D50 Werte lagen im Modell im Mittel über alle Frequenzen bei 95% und 91% (Kondition 2 und 3 entsprechend). In den Messungen ergaben sich 96% in Kondition 2 und 87% in Kondition 3. Ähnlich wie bei den RT-Werten liegen die größten Abweichungen im tieffrequenten Bereich. Mittels der am Kopf- und Rumpfsimulator aufgenommenen HRIRs wurden

Tabelle 1: Nachhallzeit RT pro Oktavband; angegeben ist die Oktavmittenfrequenz.

f (Hz)	63	125	250	500	1.000	2.000	4.000	8.000	Ø
RT Modell (s)	0,37	0,40	0,34	0,33	0,33	0,33	0,32	0,30	0,35
RT Messung (s)	0,36	0,55	0,48	0,40	0,35	0,37	0,34	0,34	0,39

f (Hz)	63	125	250	500	1.000	2.000	4.000	8.000	Ø
RT Modell (s)	0,55	0,57	0,53	0,51	0,50	0,50	0,47	0,40	0,51
RT Messung (s)	0,52	0,77	0,60	0,59	0,43	0,54	0,48	0,44	0,58

Ø bezeichnet die im Mittel über die Frequenzbänder berechnete Nachhallzeit

Oben: Kondition 2, Absorptionsgrad 80%; unten: Kondition 3, Absorptionsgrad 60%

Tabelle 2: Deutlichkeitsgrad D50 pro Oktavband; angegeben ist die Oktavmittenfrequenz und im Mittel über die Frequenzbänder.

f (Hz)	63	125	250	500	1.000	2.000	4.000	8.000	Ø
D ₅₀ Modell (%)	97	96	98	98	98	98	98	99	98
D ₅₀ Messung (%)	98	92	95	92	97	98	98	99	96

f (Hz)	63	125	250	500	1.000	2.000	4.000	8.000	Ø
D ₅₀ Modell (%)	89	88	90	91	91	91	92	94	91
D ₅₀ Messung (%)	82	86	72	87	90	92	94	93	87

Oben: Kondition 2, Absorptionsgrad 80%; unten: Kondition 3, Absorptionsgrad 60%

für alle drei Konditionen ILDs berechnet. Abbildung 4 zeigt die ILDs als Funktion der Frequenz. In allen Konditionen liegt die höchste ILD bei ca. 4 kHz. Bis 4 kHz steigen die ILDs mit steigender Frequenz steil an, ab etwa 5 kHz sinken sie leicht und bleiben bis 10 kHz nahezu konstant. Je stärker der Nachhall, umso geringer ist die ILD. Zwischen Kondition 1 (Freifeldkondition) und Kondition 3 (starker Nachhall) liegt zwischen 3.000 Hz und 5.000 Hz ein ILD-Unterschied von bis zu 10 dB.

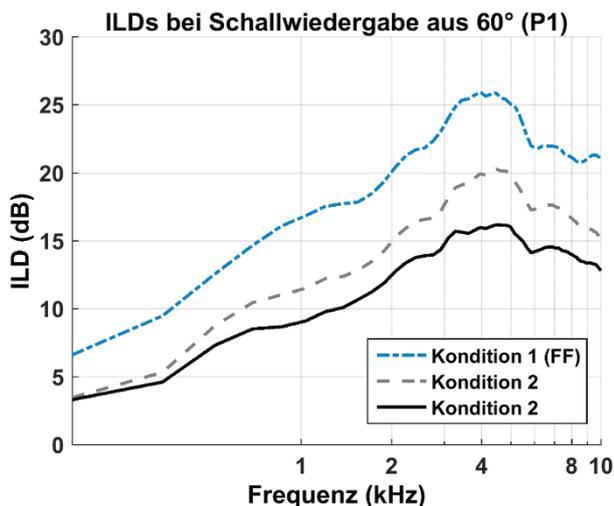


Abbildung 4: ILDs (dB) als Funktion der Frequenz für alle drei Messkonditionen: Kondition 1: Freifeld, Kondition 2: 80% Absorption, Kondition 3: 60% Absorption

Sprachtests

Material und Methode

Das Sprachverstehen im Störgeräusch wurde an normalhörenden (NH) Probanden mit dem Oldenburger Satztest (OLSA, [15], [16], [17] in den Anordnungen SON0 (Signal bei 0°, Störgeräusch bei 0°) und SON60 (Signal bei 0°, Störgeräusch bei ±60°) bestimmt. Der Abstand zwischen Proband und dem Lautsprecher bei 0° war 1 m, zu den Lautsprechern bei 60° betrug der Abstand 1,80 m. Der Direktschall des Sprachsignals wurde für alle Konditionen auf 60 dB SPL festgelegt. Das Sprachsignal blieb im Pegel konstant, während der Störgeräuschpegel adaptiv geändert wurde. Das Signal-Rauschverhältnis (SNR) bei Testbeginn lag bei +5 dB. Alle Probanden absolvierten eine Trainingsliste. Anschließend wurden in randomisierter Reihenfolge Messungen mit 100%, 80% und 60% Absorption jeweils mit den Anordnungen SON0 und SON60 durchgeführt. Als Rauschsignal wurde das Oldenburger Rauschen (OINoise) verwendet. Pro Messung wurden 20 OLSA-Sätze abgespielt. Der Proband saß in der Mitte des reflexionsarmen Raumes und wählte die Antworten über einen Touchscreen-Monitor aus.

An den Messungen nahmen 17 NH Probanden (Alter: 26,7±8,0 Jahre, 3 weibl., 14 männl.) teil. Die Ergebnisse wurden mittels Kolmogorow-Smirnow-Test auf Normalverteilung untersucht. Da der Test die Annahme einer Normalverteilung bestätigte, wurden statistische Auswertungen mittels T-Test durchgeführt. Bei Mehrfachvergleichen wurde das Signifikanzniveau mittels der Bonferroni-Methode korrigiert.

Ergebnisse

Abbildung 5 zeigt Boxplots der Sprachverständlichkeitsschwellen (SVS) für die drei Konditionen in den beiden Anordnungen SON0 (Signal 0°, Noise 0°) und SON60 (Signal 0°, Noise ±60°). Für das Sprachverstehen bei räumlich überlagerter Sprach- und Störschallwiedergabe von vorne (SON0) ergab sich in der Freifeldbedingung im Median eine SVS von -7,4 dB SNR. In Kondition 2 reduziert sich diese auf -5,8 dB SNR. Die Ergebnisse sind signifikant unterschiedlich ($p=0,024$). In Kondition 3 beträgt die SVS noch -4,7 dB SNR. Diese ist signifikant schlechter als die FF-Bedingung ($p<0,001$) und als die Bedingung mit 80% Absorption ($p=0,01$). Bei räumlicher Trennung der Sprach- und Störschallquellen (SON60) verbesserte sich die SVS in der Freifeldbedingung auf -13,9 dB SNR. Bei einem Absorptionsgrad von 80% reduziert sich die SVS auf -8,8 dB SNR, beim Absorptionsgrad von 60% beträgt sie noch -6 dB SNR. Alle Ergebnisse unterscheiden sich signifikant ($p<0,001$).

Der Gewinn in der SVS durch die räumliche Trennung (SRM) wurde durch Differenzbildung der SVS der Konditionen SON60 und SON0 bestimmt. Die Ergebnisse des SRM sind in Abbildung 6 dargestellt.

Im Freifeld beträgt der Vorteil durch die räumliche Entmaskierung 6,8 dB. Bei einem Absorptionsgrad von 80% reduziert sich der SRM auf 2,7 dB, beim Absorptionsgrad von 60% beträgt er nur noch 1,3 dB. Die Konditionen mit Raumsimulation unterscheiden sich signifikant von der Freifeldbedingung ($p<0,001$) und der SRM bei 60% Absorption ist signifikant geringer als bei 80% ($p=0,016$).

Diskussion und Ausblick

Ziel dieser Arbeit war es, das vorhandene Freifeldwiedergabesystem um eine Raumsimulation zu erweitern und diese zu evaluieren. Dazu wurden Raumsimulationen für einen Modell-Hörsaal durchgeführt und die akustischen Kenngrößen RT und D50 zwischen den Vorhersagen des Modells und der Laborwiedergabe verglichen. Zusätzlich wurde durch Einführung von zwei verschiedenen Absorptionsgraden der Effekt von zwei verschiedenen langen Nachhallzeiten auf die Sprachdiskrimination im Störgeräusch und den SRM geprüft.

Ähnlich wie in anderen Arbeiten wurde eine Raumsimulation basierend auf Reflektogrammen aus einem Raummodell realisiert [2], [13], [14] und um den diffusen Nachhall durch Nutzung eines Feedback-Delay-Netzwerks ergänzt. Für die Wiedergabe der Erstreflexionen wurde der Ansatz des nächstgelegenen Lautsprechers gewählt. Dieses Verfahren wurde eingesetzt, da die Integration in das bereits vorhandene Testsetup mit 128 Lautsprechern ohne technische Umbauten im Wiedergaberaum erfolgen konnte und aufgrund des geringen Durchmessers der Einzellautsprecher (und dem damit verbundenen geringen Lautsprecherabstand) eine äußerst gute Winkelauflösung in der Horizontalebene erreicht werden kann. Andere bereits in der Literatur beschriebene Systeme für akusti-

sche Raumsimulation mit oft weniger Lautsprechern verwenden beispielsweise die Verfahren „Higher Order Ambisonics“ (HOA, z.B. [2]), Vector Base Amplitude Panning (VBAP, [18]) oder Wellenfeldsynthese (WFS, [19]). Verfahren wie HOA und WFS haben jedoch den Nachteil, dass aufgrund der begrenzten Zahl an Lautsprechern bzw. des nicht unendlich kleinen Abstand zwischen zwei Lautsprechern eine mehr oder weniger stark ausgeprägte räumliche Unterabtastung (sogenanntes „spatial aliasing“) stattfindet, d.h. das synthetisierte Schallfeld ist oberhalb der Aliasfrequenz nicht mehr physikalisch korrekt. Bei WFS liegt, je nach Aufbau und Lautsprecherabstand, die Aliasfrequenz üblicherweise bei etwa 1–4 kHz. Das hier verwendete Verfahren des nächstgelegenen Lautsprechers wurde favorisiert, da im gesamten relevanten Frequenzbereich immer eine physikalisch korrekte Wiedergabe der Einzelreflexionen ermöglicht wird. Der Nachteil des hier beschriebenen Systems ist, dass nur eine Wiedergabe in der Horizontalebene möglich ist. Um eine dreidimensionale Raumsimulation zu ermöglichen, bietet sich die Nutzung von HOA oder VBAP an. Hierzu müssen dann aber auch Lautsprecher mit Elevation installiert werden. Bei einer konstanten Hörerposition im sogenannten „Sweet-Spot“ (Mitte des Wiedergabearrays) kann auch bei der Nutzung von HOA bei entsprechendem Wiedergabesetup eine für die Hörforschung ausreichend korrekte räumliche Auflösung erreicht werden. In der dreidimensionalen Umsetzung kann der diffuse Nachhall beispielsweise auch mit Ambisonics der 1. Ordnung auf Basis der Einhüllenden erzeugt werden [2]. Anstelle der Nutzung eines Raummodells können bei dem Wiedergabeverfahren HOA auch dreidimensionale Mikrofonaufnahmen (d.h. gemessene Impulsantworten) realer Räume zur Raumsimulation genutzt werden.

Während mit dem hier genutzten Verfahren keine dynamisch bewegten Schallquellen erzeugt werden können, ist dies mit anderen Verfahren möglich. Je nach Fragestellung und gewünschter Hörsituation ist daher die Wahl des genutzten Wiedergabeverfahrens mit den jeweiligen Vor- und Nachteilen individuell zu überdenken. Die Arbeiten von Grimm et al. [20] und Ahrens und Koautoren [21] geben einen guten Überblick, welche Verfahren sich für welche Fragestellung am besten eignen.

Die Evaluation der hier vorgestellten Raumsimulation zeigte eine gute Übereinstimmung zwischen Messungen und Modell-Daten. Sowohl RT als auch D50 zeigten nur unterhalb der Grenzfrequenz des zur Verfügung stehenden reflexionsarmen Raumes größere Abweichungen. Das in dieser Studie simulierte Auditorium hat ein Volumen von 3.520 m³ und entsprechend der Messungen eine maximale RT von 580 ms. Die mit den verwendeten Absorptionsgraden beobachteten Nachhallzeiten unterschritten die nach DIN-Norm [7] geforderte maximale mittlere Nachhallzeit von 900 ms für Unterricht/Kommunikation bzw. 1,2 s für Sprache/Vortrag deutlich. Dennoch zeigte sich bereits eine stark reduzierte Sprachverständlichkeit.

In den Messungen zum Sprachverstehen in der räumlichen Anordnung SON0 zeigte sich in Kondition 3 (60%

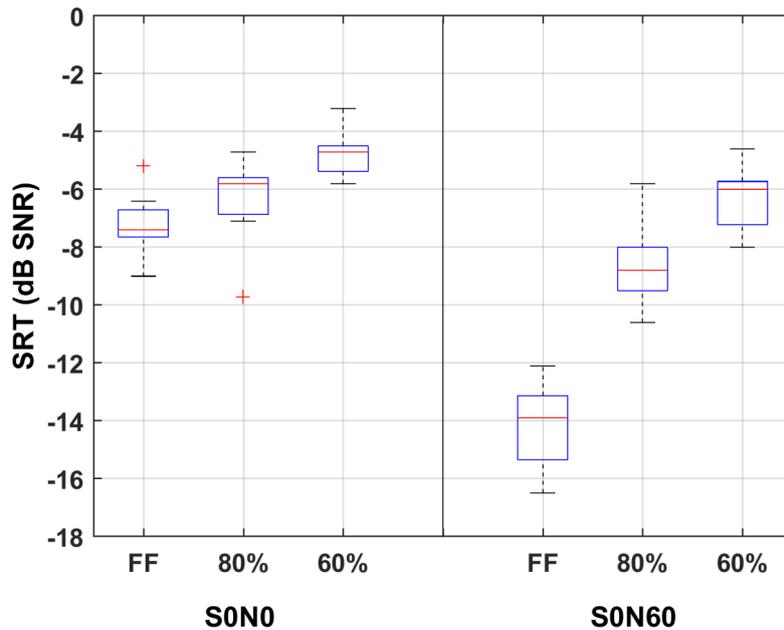


Abbildung 5: Sprachverständlichkeits-Schwelle (SVS) für drei Messkonditionen: 1. Freifeld, 2. RT 390 ms (80 % Absorption), 3. RT 580 ms (60 % Absorption); Messungen jeweils für die Sprachwiedergabe aus 0° und Störschallwiedergabe aus 0° und 60° (SON0 und SON60)

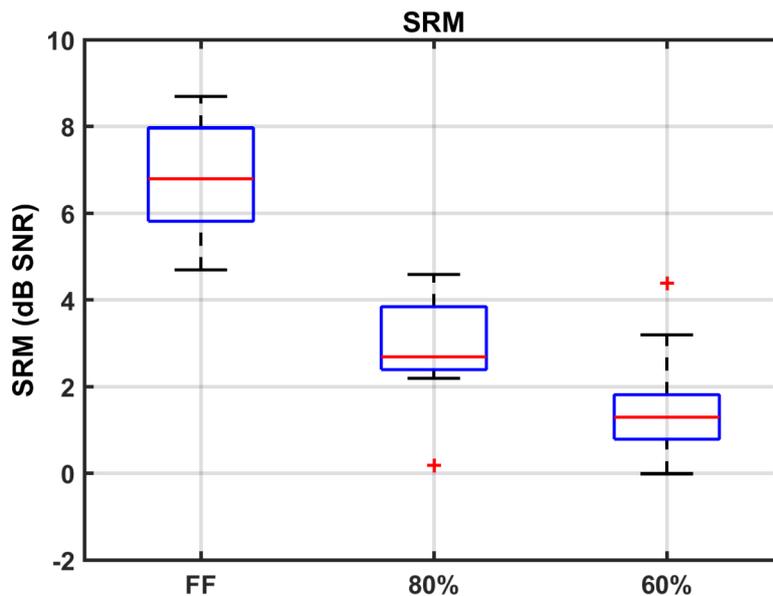


Abbildung 6: Spatial release from masking (SRM) aus SON0 und SON60 der drei Messkonditionen: 1. Freifeld, 2. RT 390 ms (80 % Absorption), 3. RT 580 ms (60 % Absorption)

Absorption, RT 580 ms) im Vergleich zur Freifeldsituation eine Verschlechterung von 2,7 dB SNR. In Kondition 2 (80% Absorption, RT 390 ms) betrug diese Verschlechterung 1,6 dB SNR. Bedingt durch frühe Reflexionen erreicht der abgegebene Schall das Ohr mit Zeitverzögerung mehrere Male und späte Reflexionen bewirken zusätzlich, dass der Schallpegel nur relativ langsam abfällt. Dies führt bei Sprachdarbietungen dazu, dass Sprachanteile von reflektiertem Schall maskiert werden. Einzelne Phänomene können somit überdeckt werden, was zu einem reduzierten Sprachverstehen führt. Die RT und D50 drücken dies ebenfalls aus: Während in Kondition 1 (Freifeldbedingung) eine minimale RT vorliegt und 100%

der Schallenergie innerhalb von 50 ms integriert wird, liegen in Kondition 3 im Mittel 13% der Schallenergie außerhalb der ersten 50 ms nach dem Direktschall. Hierdurch kann auf einen erhöhten Effekt der Maskierung von Sprache durch Nachhall geschlossen werden. In der SON60-Kondition bei 60% Absorption verschlechterte sich die SVS im Vergleich zur FF-Bedingung um 7,9 dB SNR. Dies spiegelte sich insbesondere auch im SRM wieder. Lediglich 1,3 dB SNR Verbesserung konnten durch die räumliche Trennung von Stör- und Nutzschaquelle noch erzielt werden. Abbildung 4 zeigt die ILDs der drei gemessenen akustischen Konditionen als Funktion der Frequenz. Zwischen der Kondition 1 (Freifeld) und

Kondition 3 (60% Absorption) zeigen sich verminderte ILDs über das gesamte Frequenzspektrum. Zwischen 3.000 Hz und 5.000 Hz ist der ILD-Unterschied maximal, er beträgt etwa 10 dB. Diese ILD-Reduktion erklärt den deutlich reduzierten SRM unter Nachhall, die Abschattung des Schalls durch den Kopf ist durch das diffuse Schallfeld verringert. Die durch den Kopfschatteneffekt entstehende SNR-Verbesserung am kontralateralen Ohr ist somit deutlich geringer und resultiert in einem verminderten Effekt. Kidd und Koautoren [22] demonstrierten den Effekt von Nachhall auf den SRM in einem Raum, der in einer Kondition mit Plexiglas und in der Vergleichskondition mit Schaumstoff ausgekleidet war. Der Gewinn durch räumliche Trennung von Nutz- und Störschall um 90° verringerte sich von 8 dB SNR auf 2 dB SNR in der Plexiglas-Kondition und ist vergleichbar mit den in dieser Arbeit gezeigten Resultaten.

Die Ergebnisse bestätigen die lange bekannte Beobachtung des Einflusses der Raumakustik auf die Sprachdiskrimination. Eine Raumakustik mit sehr geringer Nachhallzeit fördert gutes Sprachverstehen. Eine lange Nachhallzeit hingegen führt zu erhöhter Maskierung des direkten Sprachsignals durch zusätzliche Nachhallanteile des Störgeräuschs und des Nutzsignals selbst. Lange Nachhallzeiten haben somit eine negative Wirkung auf die Sprachdiskrimination. Der Einfluss der Raumakustik wird gegenwärtig in den klinischen audiologischen Prüfverfahren zur Sprachverständlichkeit nicht berücksichtigt. Die Bestimmung der Hörleistung im Raumsimulationssystem stellt eine sinnvolle Ergänzung zu etablierten audiologischen Messverfahren dar. Zukünftig sollen auch mit Hörgeräte- sowie Nutzern von Cochlea-Implantaten Hörtests in simulierten akustischen Hörumgebungen durchgeführt werden. Hierdurch können zusätzliche Informationen über den Einfluss der Versorgungsart und den Effekt von Signalvorverarbeitung auf die Sprachdiskrimination gewonnen werden. Eine Ergänzung der Messkonditionen ist geplant und soll zu weiteren Einblicken in die Auswirkung raumakustischer Größen auf die Verständlichkeit von Sprache führen. Insbesondere die Untersuchung von Klassenräumen mit unterschiedlich guter raumakustischer Optimierung ist hier bei verschiedenen Abständen zwischen Schallquelle (Lehrer) und Hörposition (Schüler) von Interesse. Ebenso soll der Einfluss der Raumakustik auf das Richtungshören untersucht werden. Dank der Raumsimulation können nun Hörsituationen unter realitätsnahen aber kontrollierten akustischen Bedingungen im Labor rekonstruiert werden.

Anmerkungen

Interessenkonflikte

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Danksagung

Die Autoren danken der Firma Cochlear Deutschland GmbH & Co. KG. sowie der Moessner-Stiftung Frankfurt am Main für die Unterstützung der Studie.

Literatur

1. Cox RM, Alexander GC. Hearing aid benefit in everyday environments. *Ear Hear.* 1991 Apr;12(2):127-39. DOI: 10.1097/00003446-199104000-00009
2. Minnaar P, Favrot S, Buchholz JM. Improving hearing aids through listening tests in a virtual sound environment. *Hear J.* 2010;63(10):40-42. DOI: 10.1097/01.HJ.0000389926.64797.3e
3. Revit LJ, Killion MC, Compton-Conley CL. Developing and testing a laboratory sound system that yields accurate real-world results. *Hear Rev.* 2007;14(11):54.
4. Weissgerber T. Ein Wiedergabesystem mit Wellenfeldsynthese zur Simulation alltäglicher Hörumgebungen [A sound reproduction system using wave field synthesis to simulate everyday listening conditions]. *HNO.* 2019 Apr;67(4):265-271. DOI: 10.1007/s00106-019-0635-5
5. Kuttruff H. *Room Acoustics.* 6th Ed. Boca Raton: CRC Press; 2016. DOI: 10.1201/9781315372150
6. Bradley JS, Sato H, Picard M. On the importance of early reflections for speech in rooms. *J Acoust Soc Am.* 2003 Jun;113(6):3233-44. DOI: 10.1121/1.1570439
7. DIN 18041:2016-03. *Hörsamkeit in Räumen - Anforderungen, Empfehlungen und Hinweise für die Planung.* Berlin: Beuth Verlag; 2016.
8. Weissgerber T, Bandeira M, Brendel M, Stöver T, Baumann U. Impact of Microphone Configuration on Speech Perception of Cochlear Implant Users in Traffic Noise. *Otol Neurotol.* 2019 03;40(3):e198-e205. DOI: 10.1097/MAO.0000000000002135
9. Weissgerber T, Rader T, Baumann U. Impact of a moving noise masker on speech perception in cochlear implant users. *PLoS ONE.* 2015;10(5):e0126133. DOI: 10.1371/journal.pone.0126133
10. Weissgerber T, Rader T, Baumann U. Effectiveness of Directional Microphones in Bilateral/Bimodal Cochlear Implant Users-Impact of Spatial and Temporal Noise Characteristics. *Otol Neurotol.* 2017 12;38(10):e551-e557. DOI: 10.1097/MAO.0000000000001524
11. Weissgerber T, Neumayer HL, Baumann U. Sprachverständlichkeitsschwellen mit Cochlea Implantat und mit CI-Simulation in Abhängigkeit vom Pegelverhältnis zwischen Direktschall und Diffusschall. *Z Audiol.* 2016;55(1):14-9.
12. Christensen CL, Rindel JH. A new scattering method that combines roughness and diffraction effects. In: *Proceedings Forum Acusticum; 2005; Budapest, Hungary.* p. 344-52.
13. Seeber BU, Kerber S, Hafter ER. A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hear Res.* 2010 Feb;260(1-2):1-10. DOI: 10.1016/j.heares.2009.11.004
14. Favrot S, Buchholz, J. M. LoRA: A Loudspeaker-Based Room Auralization System. *Acta Acustica.* 2010;96(2):364-75. DOI: 3813/AAA.918285
15. Wagener K, Kühnel V, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil I: Design des Oldenburger Satztests. *Z Audiol.* 1999;38(1):4-15.

16. Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil II: Optimierung des Oldenburger Satztests. *Z Audiol.* 1999;38(2):44-56.
17. Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil III: Evaluation des Oldenburger Satztests. *Z Audiol.* 1999;38(3):86-95.
18. Pulkki V. Virtual sound source positioning using vector base amplitude panning. *J Audio Eng Soc.* 1997;45(6):456-66.
19. Berkhout AJ. A holographic approach to acoustic control. *J Audio Eng Soc.* 1988;36(12):977-95. DOI: 10.1111/j.1532-5415.1988.tb05800.x
20. Grimm G, Ewert S, Hohmann V. Multi-channel loudspeaker reproduction and virtual acoustic environments in the context of hearing aid evaluation. *J Acoust Soc Am.* 2016;140(4):2999. DOI: 10.1121/1.4969292
21. Ahrens A, Marschall M, Dau T. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. *Hear Res.* 2019 06;377:307-17. DOI: 10.1016/j.heares.2019.02.003
22. Kidd G, Mason CR, Brughera A, Hartmann WM. The Role of Reverberation in Release from Masking Due to Spatial Separation of Sources for Speech Identification. *Acta Acustica.* 2005;91(3):526-36.

Korrespondenzadresse:

Anja Eichenauer
Audiologische Akustik, Klinik für HNO-Heilkunde,
Universitätsklinikum Frankfurt am Main,
Theodor-Stern-Kai 7, 60590 Frankfurt am Main,
Deutschland, Tel.: + 49 69 6301 7804
anja.eichenauer@kgu.de

Bitte zitieren als

Eichenauer A, Baumann U, Weißgerber T. Implementierung und Evaluation einer akustischen Raumsimulation für audiologische Testungen. *GMS Z Audiol (Audiol Acoust).* 2020;2:Doc06.
DOI: 10.3205/zaud000010, URN: urn:nbn:de:0183-zaud0000108

Artikel online frei zugänglich unter

<https://www.egms.de/en/journals/zaud/2020-2/zaud000010.shtml>

Veröffentlicht: 11.09.2020

Copyright

©2020 Eichenauer et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.

Implementation and evaluation of acoustic room simulation for audiological testing

Abstract

In everyday life listening situations, communication is characterized by a variety of challenging acoustic scenes. Usually, background noise sources are present and, additionally, direct sound in rooms is reflected from walls and surfaces. These reflections affect speech perception. However, audiological assessments are usually performed in rooms with low reverberation time. This work aims at implementing and evaluating a room simulation system to examine hearing ability in everyday life listening situations with different acoustic properties and reverberation.

With a room simulation software sound propagation was calculated in relation to the room geometry and acoustic room parameters. The direct sound and its reflections are presented according to their direction by using a multichannel playback system with 128 loudspeakers. Any desired room can be reconstructed in the laboratory under realistic and controllable acoustic conditions.

The room simulation procedure was evaluated using an empty auditorium with different degrees of absorption properties. Technical measurements of reverberation time (RT) and distinctness (D50) showed good agreement with the model data. Additionally, interaural level differences (ILDs) were recorded and correlated with speech reception thresholds. ILDs are reduced up to 10 dB with increasing reverberation. At the same time, spatial release from masking (SRM) compared to free field conditions decreased up to 5.5 dB SNR.

The determination of speech discrimination using the introduced room simulation system is a useful complement to established audiological measurements. The hearing performance under defined acoustic properties can be determined reproducibly and reliably.

Keywords: room simulation, reverberation, speech perception, spatial release from masking, SRM, interaural level difference, ILD

Introduction

In everyday life, complex hearing situations with background noise often occur. In addition, in closed rooms direct sound is superimposed with reverberation. Typical listening situations in everyday life are, for example, babble noise of many speakers in a restaurant, at the train station or in a supermarket. Speech perception is highly dependent on the number and spatial arrangement of noise sources [1]. Test procedures in the diagnosis and therapy of hearing disorders as well as in the fitting and testing of hearing systems are usually performed in acoustically optimized listening booths with very low reverberation times. Therefore, the actual hearing performance in everyday life cannot be reproduced. Here, only one or two sound sources (loudspeakers) are used, which limits the possible directions of incidence of target sound (i.e. speech) and noise. Simulation methods offer the

possibility to simulate everyday listening situations in order to examine the hearing performance of people with and without hearing loss [2], [3], [4].

The sum of all reflections within a delimited room is defined as reverberation [4]. The acoustic information produced by a sound source in closed rooms is superimposed by these reflections. The signal that arrives at the receiver can be divided into direct sound, early and late reflections. Early reflections occur up to 50–80 ms after the first sound wave and are partly beneficial for speech perception because they provide amplification of the information of the direct sound [6]. Late reflections occur with more delay and are perceived diffusely or as a separate echo with the consequence of deteriorated speech perception. The acoustic properties of a room are expressed by various measures such as reverberation time (RT), clarity measures (C50/C80), “Deutlichkeit”/definition (D50) or strength (G). The reverberation time describes the time it takes for the sound pressure level to decrease to one thousandth of its initial value (i.e. by

Anja Eichenauer¹
Uwe Baumann¹
Tobias Weißgerber¹

¹ Audiological Acoustics, Clinic for Otolaryngology, University Hospital Frankfurt am Main, Germany

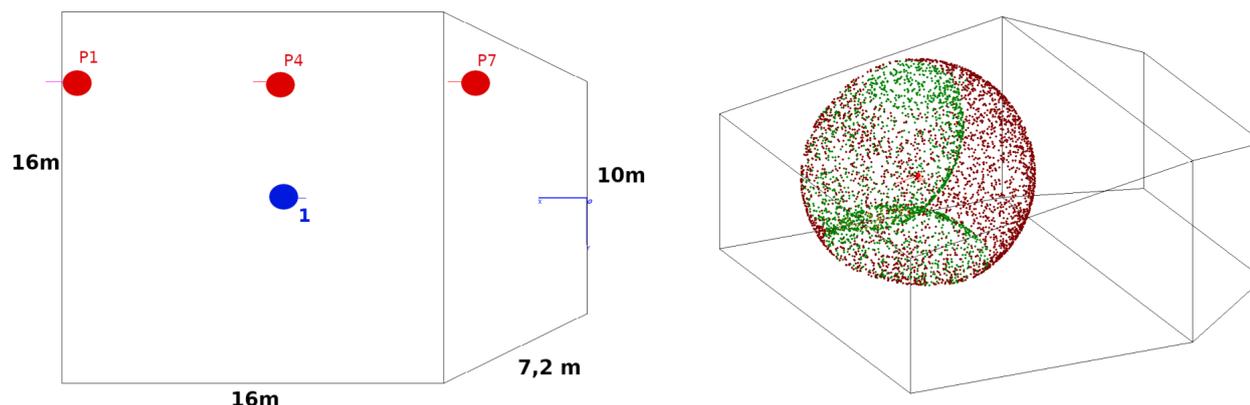


Figure 1: Left figure: top view of lecture hall model with receiver: blue dot (1) and sound sources: P4 (0°) and P1/P7 (±60°); right figure: sound propagation in the three-dimensional auditorium model from source P1

60 dB). The clarity $C50$ describes the ratio between the sound energy in the period up to 50 ms after direct sound and the sound energy from 50 ms onwards. Clarity $C80$ is defined by the ratio between the sound energy up to 80 ms after the direct sound and the sound energy from 80 ms onwards. The definition $D50$ describes the energy ratio between the early reflections of the first 50 ms to the total sound energy. The strength G puts the sound level of a spherically radiating sound source in relation to the level of the same source at a distance of 10 m. The mentioned room acoustic parameters are influenced by the size (i.e. volume) of the room and the acoustic properties of the boundary surfaces such as ceilings, floor, walls, etc. Depending on the acoustic properties of a room, the temporal and spectral structure of the reproduced signal is changed. The superposition of direct sound with reverberation can be compared to the effect of a masking noise and can additionally reduce the modulation depth of the target signal.

Due to the dependency between room size/condition and speech perception, DIN 18041:201603 (audibility in rooms) recommends suitable reverberation times depending on the volume of the room [7]. Depending on the room usage, there are different recommendations for music, speech, teaching, etc. For room volumes between 1,000 to 5,000 m³, the optimum reverberation time for teaching is approx. 0.75 to 1 s, rooms for teaching lessons for a handicapped audience should not exceed 0.6 to 0.8 s reverberation time for the specified room volume.

This study describes the further development of an already existing free field reproduction system with 128 loudspeakers [8] for reproducible, flexible and plausible reproduction of sounds in rooms with different acoustic properties. The evaluation includes the analysis of room acoustic parameters and interaural level differences in three reverberated conditions. Furthermore, the influence of reverberation on speech perception in noise will be investigated under consideration of spatial release from masking (SRM) in normal hearing subjects.

Materials and methods

Implementation of the room simulation system

For the implementation of the simulation system, an anechoic chamber was available at the Audiological Acoustics Department of the ENT University Hospital of the J.W.-Goethe University Frankfurt am Main. This room is equipped with a reproduction system with 128 loudspeakers (rectangular array in the horizontal plane), which was previously used either under free field conditions (e.g. [8], [9], [10]) or for the direction-independent simulation of diffuse sound [11]. The existing system was extended by a component for room simulation.

Room model

The software package ODEON (ODEON A/S, Lyngby, Denmark) was used with to implement a three-dimensional model of an empty lecture hall (see Figure 1) with a room volume of 3,520 m³. The height of the room is between 7.5 to 10 m. This room model is freely available and delivered as example with ODEON (file name: Example.par). In ODEON, absorption properties or specific material properties can be assigned to objects and walls. Based on the room geometry mentioned above, two room models with different absorption properties of all room surfaces (i.e. walls, ceiling, floors) were generated. The frequency independent absorption coefficients were 80% and 60%. Thereby, 80% absorption corresponds to an average reverberation time of 350 ms and 60% to a reverberation time of 510 ms.

Three sound sources were positioned in the front half space of the horizontal plane at -60°, 0° and 60° relative to the receiver. The frontal sound source (0°) was 5 m apart from the listening position, the lateral sound sources were 10 m apart from the listening position. Figure 1 (left) shows a sketch of the room geometry of the room with three sound sources and one receiver. In the room model, the sound propagation of a signal emit-

ted by the transmitter independently of direction and frequency was calculated (mirror source model and “ray-radiosity method”, [12]). Reflections up to the 10th order were calculated and stored in a reflectogram. For each reflection, the reflectogram contains the time delay of the reflections, the octave levels between 63 and 8,000 Hz, the horizontal sound incidence angles and the vertical sound incidence angles. Figure 1 (right) displays sound propagation. During its propagation, the radiated sound hits room boundaries several times and is reflected repeatedly. The direct sound is shown in dark red and first-order reflections in green.

Room simulation

Based on the reflectograms calculated by ODEON, the room simulation was realized in a low-reflective room. Figure 2 shows a sketch of the laboratory room with the positions of the sound sources.

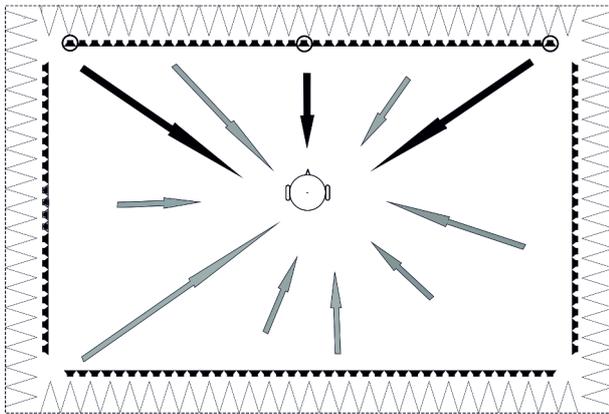


Figure 2: Arrangement of the 128 loudspeakers of the sound reproduction system in the anechoic chamber

Centre (listening position) with sketched test person; circles: speakers with positions corresponding to the simulated sound sources P1, P4 and P7; direct sound is outlined by black arrows, early reflections and diffuse reverberation by grey arrows.

The nearest loudspeaker method was used [13], [14]. This means that for each calculated reflection, the speaker with the smallest distance of the horizontal angle is selected for playback. Reflections with an elevation angle of incidence greater than $\pm 30^\circ$ were not taken into account because the speakers of the present sound reproduction system are arranged horizontally. While a complete integration (i.e. mapping) of all elevation angles in the horizontal plane allows a complete preservation of the reflection density of the initial reflections, this variant leads to clearly audible sound coloration. After preliminary listening experiments, vertical angles of incidence up to $\pm 30^\circ$ were discarded as a compromise between reflection density and sound coloration. Based on the reflectogram, an impulse response was generated for each reflection according to its time delay and spectrum.

In the next step, each reflection was assigned to the loudspeaker closest to the calculated horizontal angle and the impulse responses per loudspeaker were summed. Late reflections were inserted as diffuse rever-

beration using a statistical model 80 ms after direct sound incidence (Feedback Delay Network, FDN retrieved from Quality & Usability Labs, Institution for Software Technology and Theoretic Informatics, Technical University Berlin). For this purpose, the RMS levels of all generated reflections were determined in the time window between 80 ms and 120 ms after direct sound. The level of the diffuse reverberation was adjusted in level to the previously generated reflections. The diffuse reverberation was generated using the frequency-dependent RT and faded in equally on all 128 channels. To compensate for the individual transmission characteristics of the loudspeakers to the center of the room (amplitude frequency response and running time) due to the rectangular arrangement, equalization was carried out individually for each loudspeaker (finite impulse response (FIR) filter). The generated 128-channel room impulse response (RIR) was convolved with the desired single-channel audio signal for sound reproduction. The spatial sound was presented simultaneously via all 128 loudspeakers. RIRs of both conditions with reverberation were generated, each for three direct sound positions ($\pm 60^\circ$, 0°). Figure 3 shows a flowchart of the room simulation with direct sound, early reflections and diffuse reverberation.

Evaluation of the room simulation system

To check the quality of the simulation, RT and D50 were measured and compared with the model data in the first step. In a second step, the system was evaluated from an audiological point of view by conducting speech tests and evaluating the results. The influence of the reverberation time on speech perception and SRM was investigated. In addition, head-related IRs (HRIRs) recorded with an artificial head were examined for interaural level differences (ILDs) and considered in relation to the SRM.

Technical evaluation

Material and method

The RT and D50 were analyzed per octave frequency band from 63 Hz to 8,000 Hz as well as broadband averaged and compared with the calculations of the room model. The analysis was performed for a sound source at 0° . A sinus sweep (frequency range from 50 Hz to 22,050 Hz) of 0.5 s duration was used as stimulus. The measurements at the listening position were made with a measuring microphone of type 4155 (Brüel & Kjær, Nærum, Denmark) and an impedance converter 2,669 connected to a measuring amplifier (Nexus) of the same manufacturer. The impulse responses were analyzed with the acoustic software ARTA (Artalabs, Kastel Luksic, Croatia). In addition, the IRs were measured at the head and torso simulator (Brüel & Kjær Type 4100 with microphones Type 4190-L-002). The sound presentation was made from 60° . ILDs were derived from the results of

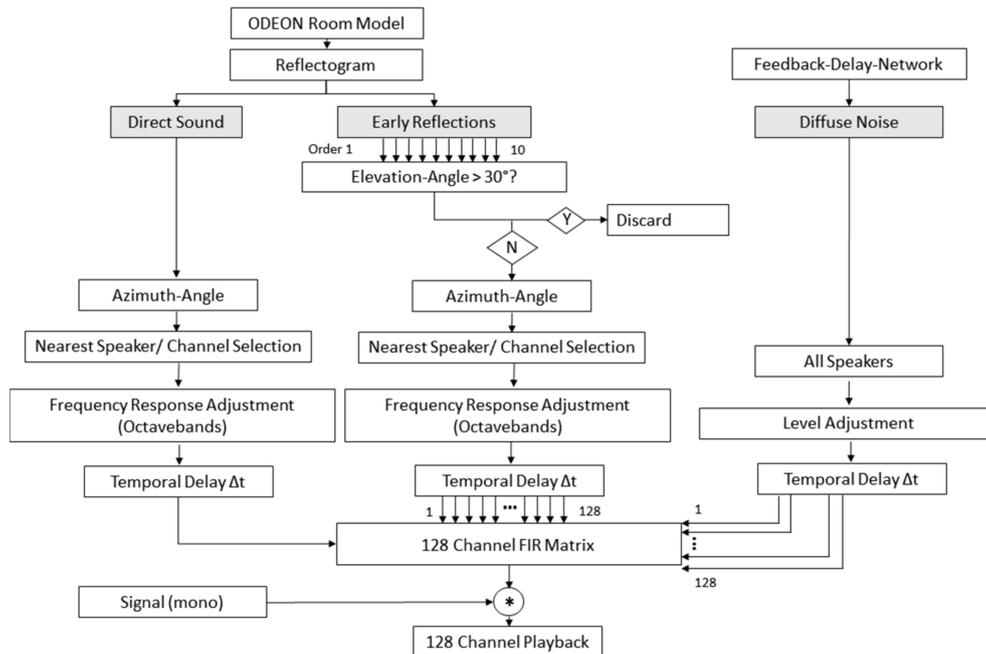


Figure 3: Flow chart of the room simulation with direct sound, early reflections and diffuse reverberation.

the measurements. All measurements were performed in both simulated conditions with absorption coefficients of 60% and 80% and under free field conditions (i.e. ~100% absorption).

Results

For the free field measurement an average RT of 0.05 s was obtained. The RT values per octave band are given in Table 1 for measurement condition 2 with 80% absorption and measurement condition 3 with 60% absorption. The RT averaged over all frequencies determined by the model using ODEON was 0.35 s in the condition with 80% absorption and 0.51 s with 60% absorption. The measurements with the simulation system in the (nearly) anechoic room resulted in a reverberation time of 0.39 s at 80% absorption and 0.58 s at 60% absorption. The greatest deviations were found in the lower frequency bands, especially in the range and below the cut-off frequency of the low-reflective room (~200 Hz).

Table 2 shows the D50 measurement per octave frequency as an example for measurement condition 2 (80% absorption) and measurement condition 3 (60% absorption). In the model, the averaged D50 values were 95% and 91% over all frequencies (condition 2 and 3 accordingly). The measurements showed 96% in condition 2 and 87% in condition 3. Similar to the RT values, the largest deviations are in the lower frequency range. Using the HRIRs recorded on the head and torso simulator, ILDs were calculated for all three conditions. Figure 4 shows the ILDs as a function of frequency. In all conditions the highest ILD is at about 4 kHz. Up to 4 kHz the ILDs rise steeply with increasing frequency, from about 5 kHz they decrease slightly and remain almost constant up to 10 kHz. The stronger the reverberation, the lower the ILD. Between condition 1 (free field condition) and condition

3 (strong reverberation) there is an ILD difference of up to 10 dB between 3,000 Hz and 5,000 Hz.

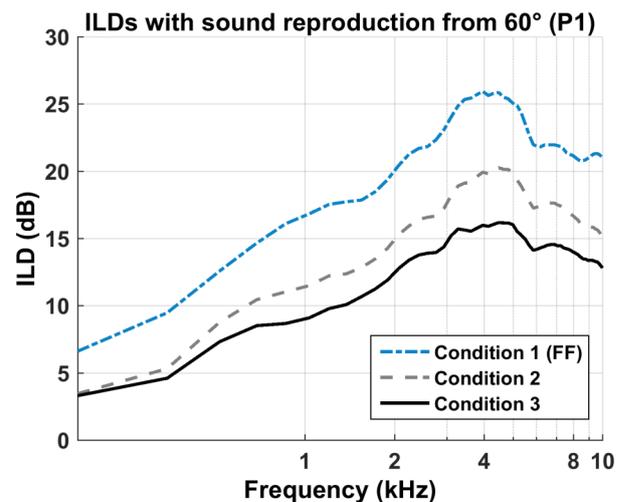


Figure 4: ILDs (dB) as a function of frequency for all three measurement conditions: condition 1: free field, condition 2: 80% absorption, condition 3: 60% absorption

Speech perception tests

Material and method

Speech perception in noise was determined in normal hearing (NH) with the Oldenburg sentence test (OLSA, [15], [16], [17] in the source arrangements SON0 (signal at 0°, noise at 0°) and SON60 (signal at 0°, noise at ±60°).

The distance between the subject and the loudspeaker at 0° was 1 m, to the loudspeakers at 60° the distance was 1.80 m. The direct sound of the speech signal was set at 60 dB SPL for all conditions. The level of the speech

Table 1: Reverberation time RT per octave band; indicated is the octave center frequency.

f (Hz)	63	125	250	500	1,000	2,000	4,000	8,000	∅
RT Model (s)	0.37	0.40	0.34	0.33	0.33	0.33	0.32	0.30	0.35
RT Meas. (s)	0.36	0.55	0.48	0.40	0.35	0.37	0.34	0.34	0.39

f (Hz)	63	125	250	500	1,000	2,000	4,000	8,000	∅
RT Model (s)	0.55	0.57	0.53	0.51	0.50	0.50	0.47	0.40	0.51
RT Meas. (s)	0.52	0.77	0.60	0.59	0.43	0.54	0.48	0.44	0.58

∅: average reverberation time calculated over the frequency bands

Upper table: condition 2, absorption coefficient 80%; lower table: condition 3, absorption coefficient 60%

Table 2: Degree of clarity D50 per octave band (indicating the octave centre frequency) and averaged over the frequency bands

f (Hz)	63	125	250	500	1,000	2,000	4,000	8,000	∅
D ₅₀ Model (%)	97	96	98	98	98	98	98	99	98
D ₅₀ Meas. (s)	98	92	95	92	97	98	98	99	96

f (Hz)	63	125	250	500	1,000	2,000	4,000	8,000	∅
D ₅₀ Model (%)	89	88	90	91	91	91	92	94	91
D ₅₀ Meas. (s)	82	86	72	87	90	92	94	93	87

Upper table: condition 2, degree of absorption 80%; lower table: condition 3, absorption coefficient 60%

signal remained constant, while the noise level was adaptively changed. The initial signal to noise ratio (SNR) was set to +5 dB. All subjects completed a training list. Subsequently, measurements with 100%, 80% and 60% absorption were performed in randomized order with the SONO and SON60 setups. As noise signal the Oldenburg noise (OINoise) was used. One run consisted of 20 sentences. The subject was placed in the center of the test room and responded via a touch screen monitor.

17 NH subjects (age: 26.7±8.0 years, 3 female, 14 male) participated in the measurements. The results were examined for normal distribution using the Kolmogorov-Smirnov test. Since the test confirmed the assumption of a normal distribution, statistical evaluations were carried out using the T-test. In multiple comparisons, the significance level was corrected using the Bonferroni method.

Results

Figure 5 shows boxplots of the speech perception thresholds (SRTs) for the three conditions in the two source arrangements SONO (Signal 0°, Noise 0°) and SON60 (Signal 0°, Noise ±60°).

For speech perception with spatially superimposed speech and noise playback from the front (SONO), the free field condition resulted in a median SRT of -7.4 dB SNR. In condition 2 this is reduced to -5.8 dB SNR. The results are significantly different ($p=0.024$). In condition 3 the SRT is still -4.7 dB SNR. This is significantly worse than the FF condition ($p<0.001$) and than in the condition with 80% absorption ($p=0.01$).

With spatial separation of the speech and noise sources (SON60), the SRT in the free field condition improved to -13.9 dB SNR. At an absorption coefficient of 80%, the SRT is reduced to -8.8 dB SNR, at an absorption coefficient of 60% it is still -6 dB SNR. All results differ significantly ($p<0.001$).

The benefit in terms of SRT due to spatial separation (SRM) was determined by subtracting the SRT of conditions SON60 and SONO. SRM results are shown in Figure 6. In the free field, the advantage of spatial unmasking is 6.8 dB. With 80% absorption, the SRM is reduced to 2.7 dB; with 60% absorption it is only 1.3 dB. The results of conditions with reverberation differ significantly from free field condition ($p<0.001$) and the SRM at 60% absorption is significantly lower than at 80% absorption ($p=0.016$).

Discussion and outlook

The aim of this work was the extension of our custom free field reproduction system by a room simulation procedure and the subsequent evaluation of the setup. For this purpose, room simulations were performed for a model auditorium and the acoustic parameters RT and D50 were compared between the predictions of the model and the laboratory reproduction. In addition, the effect of two different reverberation times on speech perception in noise and SRM was tested by introducing two different absorption levels.

Similar to other works, a room simulation based on reflectograms from a room model was realized [2], [13], [14] and supplemented by the diffuse reverberation using a

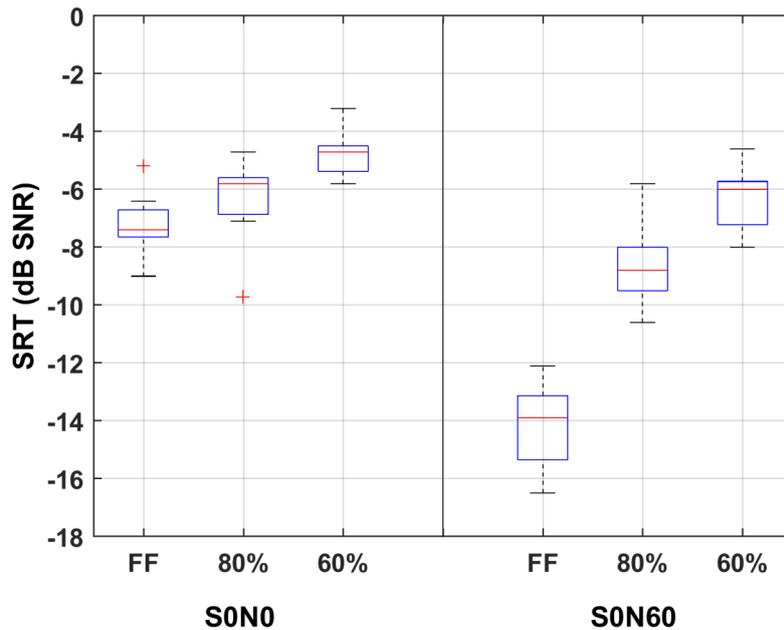


Figure 5: Speech reception threshold (SRT), for three measurement conditions: 1. free field, 2. RT 390 ms (80% absorption), 3. RT 580 ms (60% absorption); measurements for speech reproduction from 0° and noise reproduction from 0° and 60° (SON0 and SON60)

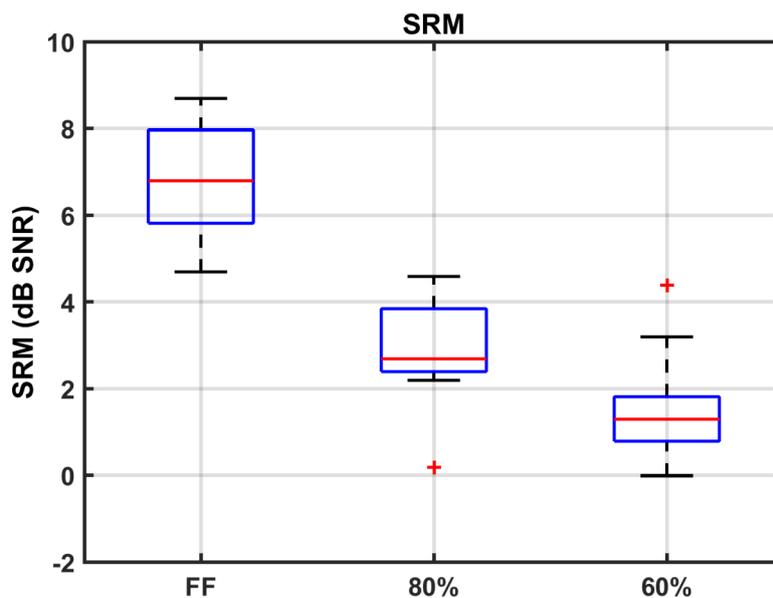


Figure 6: Spatial release from masking (SRM) from SON0 and SON60 of the three measurement conditions: 1. free field, 2. RT 390 ms (80% absorption), and 3. RT 580 ms (60% absorption)

feedback delay network. For the reproduction of the initial reflections, the approach of the nearest loudspeaker was chosen. This method was used because the integration into the already existing test setup with 128 loudspeakers could be done without technical modifications of our local reproduction room. Furthermore, the small diameter of the individual loudspeakers (and the resulting small loudspeaker distance) allow an extremely good angular resolution in the horizontal plane. Other systems with fewer loudspeakers, which have already been described in the acoustic room simulation literature, apply “Higher Order Ambisonics” (HOA, e.g. [2]), Vector Base Amplitude Panning (VBAP, [18]) or Wave Field Synthesis (WFS, [19]).

Methods such as HOA and WFS have the disadvantage, that due to the limited number of loudspeakers or due to the not infinitely small distance between two loudspeakers, a more or less pronounced spatial subsampling (so-called “spatial aliasing”) occurs, i.e. the synthesized sound field is no longer physically correct above the alias frequency. With WFS, the alias frequency is usually around 1–4 kHz, depending on the arrangement and speaker distance. The method of the nearest loudspeaker used here was favored, since it always allows physically correct reproduction of the individual reflections in the entire relevant frequency range. The disadvantage of the system described here is that only a playback in the horizontal

plane is possible. To enable a three-dimensional room simulation, the use of HOA or VBAP is recommended. To accomplish this, loudspeakers with elevated positions must be installed. If the listener position is held constant in the so-called “sweet spot” (center of the playback array), a sufficiently correct spatial resolution for hearing research can also be achieved when using HOA with the appropriate playback setup. In the three-dimensional implementation, the diffuse reverberation can also be generated with 1st order ambisonics based on the envelope [2]. Instead of using a room model, three-dimensional microphone recordings (i.e. measured impulse responses) of real rooms can also be used for room simulation with the HOA reproduction method.

While the method used here cannot generate dynamically moving sound sources, this is possible with other methods. Depending on the problem and the desired listening situation, the choice of the reproduction method used must therefore be considered individually with the respective advantages and disadvantages. The studies of Grimm et al. [20] and Ahrens et al. [21] provide a good overview of which methods are best suited for which problem.

The evaluation of the room simulation presented here showed a good agreement between measurements and model data. Both RT and D50 showed larger deviations only below the cut-off frequency of the local anechoic chamber. The auditorium simulated in this study had a volume of 3,520 m³ and according to the measurements a maximum RT of 580 ms.

The reverberation times accomplished in the present study were well below the recommended maximum average reverberation time of 900 ms for teaching/communication (1.2 s for speech/lecture) as required by the DIN standard [7]. Nevertheless, speech intelligibility was already severely impaired.

The results on speech perception in source arrangement SONO condition 3 (60% absorption, RT 580 ms) showed a deterioration of 2.7 dB SNR against FF situation. In condition 2 (80% absorption, RT 390 ms), the deterioration was 1.6 dB SNR. Due to early reflections, the emitted sound reaches the ear several times with small delay, additionally late reflections cause the sound level to decrease slowly. In speech listening tasks, this leads to a masking of speech components by sound reflections. Individual phonemes may be masked, which leads to reduced speech perception. The RT and D50 also express this: while in condition 1 (free field condition) there is a low RT and 100% of the sound energy is integrated within 50 ms, in condition 3 an average of 13% of the sound energy occurred after the first 50 ms after direct sound incidence. This suggests an increased masking of speech by reverberation.

In the SON60 source setup at 60% absorption, the SRT deteriorated by 7.9 dB SNR compared to the FF condition. This was particularly reflected in the SRM. Only 1.3 dB SNR improvement could be achieved by spatial separation of noise and target sources. Figure 4 shows the ILDs of the three measured acoustic conditions as a function of frequency. Between condition 1 (free field) and condition

3 (60% absorption), reduced ILDs are shown over the entire frequency spectrum. Between 3,000 Hz and 5,000 Hz the ILD difference is at its maximum with about 10 dB. This ILD reduction explains the significantly reduced SRM under reverberation, the shadowing of the sound by the head is reduced in a diffuse sound field. The SNR improvement in the contralateral ear caused by the head shadow effect is therefore significantly lower and results in a reduced effect. Kidd and co-authors [18] demonstrated the effect of reverberation on the SRM in a room lined with plexiglass in one condition and foam in the other. The benefit of spatial separation of target sound and noise by 90° decreased from 8 dB SNR to 2 dB SNR in the plexiglass condition and is in line with the results shown here.

The results confirm the long known observation of the impact of room acoustics on speech perception. Room acoustics with very small reverberation time promote good speech perception. A long reverberation time, on the other hand, leads to increased masking of the direct speech signal by additional reverberation components of the background noise and the target signal itself. Long reverberation times therefore have a negative effect on speech perception. The influence of room acoustics is currently not taken into account in clinical audiological test procedures for speech perception. The determination of hearing performance by means of a room simulation system is a useful supplement to established audiological measurement procedures. In the future, hearing tests with hearing aid and cochlear implant users should be performed in simulated acoustic listening environments. This will provide additional information on the impact of the type of hearing aid and the effect of signal pre-processing on speech perception. A supplement to the measurement conditions is planned and should lead to further insights into the effect of room acoustic parameters on speech perception. Especially the investigation of classrooms with different room acoustic properties and different distances between sound source (teacher) and listening position (student) are of interest. The influence of room acoustics on directional hearing will also be investigated. Thanks to room simulation, listening situations can now be reconstructed in the laboratory under realistic but controlled acoustic conditions.

Notes

Competing interests

The authors declare that they have no competing interests.

Acknowledgement

The authors thank the company Cochlear Deutschland GmbH & Co. KG and the Moessner Foundation Frankfurt am Main for supporting the study.

References

1. Cox RM, Alexander GC. Hearing aid benefit in everyday environments. *Ear Hear.* 1991 Apr;12(2):127-39. DOI: 10.1097/00003446-199104000-00009
2. Minnaar P, Favrot S, Buchholz JM. Improving hearing aids through listening tests in a virtual sound environment. *Hear J.* 2010;63(10):40-42. DOI: 10.1097/01.HJ.0000389926.64797.3e
3. Revit LJ, Killion MC, Compton-Conley CL. Developing and testing a laboratory sound system that yields accurate real-world results. *Hear Rev.* 2007;14(11):54.
4. Weissgerber T. Ein Wiedergabesystem mit Wellenfeldsynthese zur Simulation alltäglicher Hörumgebungen [A sound reproduction system using wave field synthesis to simulate everyday listening conditions]. *HNO.* 2019 Apr;67(4):265-271. DOI: 10.1007/s00106-019-0635-5
5. Kuttruff H. *Room Acoustics.* 6th Ed. Boca Raton: CRC Press; 2016. DOI: 10.1201/9781315372150
6. Bradley JS, Sato H, Picard M. On the importance of early reflections for speech in rooms. *J Acoust Soc Am.* 2003 Jun;113(6):3233-44. DOI: 10.1121/1.1570439
7. DIN 18041:2016-03. Hörsamkeit in Räumen - Anforderungen, Empfehlungen und Hinweise für die Planung. Berlin: Beuth Verlag; 2016.
8. Weissgerber T, Bandeira M, Brendel M, Stöver T, Baumann U. Impact of Microphone Configuration on Speech Perception of Cochlear Implant Users in Traffic Noise. *Otol Neurotol.* 2019 03;40(3):e198-e205. DOI: 10.1097/MAO.0000000000002135
9. Weissgerber T, Rader T, Baumann U. Impact of a moving noise masker on speech perception in cochlear implant users. *PLoS ONE.* 2015;10(5):e0126133. DOI: 10.1371/journal.pone.0126133
10. Weissgerber T, Rader T, Baumann U. Effectiveness of Directional Microphones in Bilateral/Bimodal Cochlear Implant Users-Impact of Spatial and Temporal Noise Characteristics. *Otol Neurotol.* 2017 12;38(10):e551-e557. DOI: 10.1097/MAO.0000000000001524
11. Weissgerber T, Neumayer HL, Baumann U. Sprachverständlichkeitsschwellen mit Cochlea Implantat und mit CI-Simulation in Abhängigkeit vom Pegelverhältnis zwischen Direktschall und Diffussschall. *Z Audiol.* 2016;55(1):14-9.
12. Christensen CL, Rindel JH. A new scattering method that combines roughness and diffraction effects. In: *Proceedings Forum Acusticum; 2005; Budapest, Hungary.* p. 344-52.
13. Seeber BU, Kerber S, Hafter ER. A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hear Res.* 2010 Feb;260(1-2):1-10. DOI: 10.1016/j.heares.2009.11.004
14. Favrot S, Buchholz, J. M. LoRA: A Loudspeaker-Based Room Auralization System. *Acta Acustica.* 2010;96(2):364-75. DOI: 3813/AAA.918285
15. Wagener K, Kühnel V, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil I: Design des Oldenburger Satztests. *Z Audiol.* 1999;38(1):4-15.
16. Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil II: Optimierung des Oldenburger Satztests. *Z Audiol.* 1999;38(2):44-56.
17. Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil III: Evaluation des Oldenburger Satztests. *Z Audiol.* 1999;38(3):86-95.
18. Pulkki V. Virtual sound source positioning using vector base amplitude panning. *J Audio Eng Soc.* 1997;45(6):456-66.
19. Berkhout AJ. A holographic approach to acoustic control. *J Audio Eng Soc.* 1988;36(12):977-95. DOI: 10.1111/j.1532-5415.1988.tb05800.x
20. Grimm G, Ewert S, Hohmann V. Multi-channel loudspeaker reproduction and virtual acoustic environments in the context of hearing aid evaluation. *J Acoust Soc Am.* 2016;140(4):2999. DOI: 10.1121/1.4969292
21. Ahrens A, Marschall M, Dau T. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. *Hear Res.* 2019 06;377:307-17. DOI: 10.1016/j.heares.2019.02.003
22. Kidd G, Mason CR, Brughera A, Hartmann WM. The Role of Reverberation in Release from Masking Due to Spatial Separation of Sources for Speech Identification. *Acta Acustica.* 2005;91(3):526-36.

Corresponding author:

Anja Eichenauer
Audiological Acoustics, Clinic for Otolaryngology, University Hospital Frankfurt am Main, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany, Phone: +49 69 6301 7804
anja.eichenauer@kgu.de

Please cite as

Eichenauer A, Baumann U, Weißgerber T. Implementierung und Evaluation einer akustischen Raumsimulation für audiologische Testungen. *GMS Z Audiol (Audiol Acoust).* 2020;2:Doc06. DOI: 10.3205/zaud000010, URN: urn:nbn:de:0183-zaud0000108

This article is freely available from

<https://www.egms.de/en/journals/zaud/2020-2/zaud000010.shtml>

Published: 2020-09-11

Copyright

©2020 Eichenauer et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.