

VI Raum- und Bauakustik

In der Raumakustik beschäftigt man sich mit der Ausbreitung von Schall in Räumen und den Eigenschaften der dabei auftretenden Schallfelder. Dieses Gebiet ist wichtig, um die Geometrie und Ausstattung eines Raumes an die jeweiligen Erfordernisse anzupassen, z. B. für die Nutzung als Konzertsaal, als Konferenzraum oder als Sprecherstudio. In der Bauakustik versucht man dagegen, die akustische Isolierung von Räumen in einem Gebäude gegeneinander zu beschreiben und zu verbessern und die Auswirkung von baulichen Gegebenheiten auf die Ausbreitung von Schallwellen im Gebäude und im jeweiligen Raum zu beherrschen.

VI.1 Wellenakustik in Räumen

Bei der Ausbreitung von Schallwellen in begrenzten Räumen treten aufgrund der Reflexionen an den Wänden stehende Wellen auf, die bei bestimmten Frequenzen zu einer resonanzförmigen Anregung des Raumes führen können (d. h. wenn beispielsweise ein ganzzahliges Vielfaches von halben Wellenlängen zwischen zwei parallel zueinander stehenden Wänden „paßt“). Diese speziellen Schwingungsformen im Raum werden als **Eigenmoden** des Raumes bezeichnet, die bei bestimmten Resonanzfrequenzen auftauchen. Beispielsweise treten bei einem quaderförmigen Hohlraum mit schallharten Wänden und den drei Dimensionen l_1 , l_2 , l_3 („Schuhkarton“) die folgenden Eigenmoden auf:

$$\varphi_n = \cos \frac{n_1 \cdot \pi \cdot x}{l_1} \cos \frac{n_2 \cdot \pi \cdot y}{l_2} \cos \frac{n_3 \cdot \pi \cdot z}{l_3} \quad (\text{VI.1})$$

$$f_n = \frac{c}{2} \cdot \sqrt{\left(\frac{n_1}{l_1}\right)^2 + \left(\frac{n_2}{l_2}\right)^2 + \left(\frac{n_3}{l_3}\right)^2} \quad (\text{VI.2})$$

φ_n bezeichnet dabei das Schallpotential, aus dem Schalldruck und -schnelle durch Differentiation bzw. Gradientenbildung hervorgehen; n_1 , n_2 und n_3 bezeichnen die Kennziffern der Mode, d. h. die Anzahl der Knotenlinien, die in x-, y-, bzw. z-Richtung in dem Raum auftreten. Die Resonanz-Frequenzen f_n sind bei tiefen Frequenzen relativ selten, ihre Dichte nimmt jedoch bei hohen Frequenzen stark zu. Je weniger absorbierend die Wände und die Luft in dem Raum sind, desto stärker ausgeprägt sind die Raumresonanzen und umso entscheidender ist es für die Ausbreitung einer bestimmten Schallwelle, ob sie genau auf einer Resonanz-Frequenz liegt oder nicht. Diese exakte Bestimmung des Schallfeldes durch die Randbedingungen macht eine Analogie zur Quantenmechanik (Teilchen

in einem abgeschlossenen Käfing mit quantisierten Energiezuständen) möglich, die hier aber nicht weiter vertieft werden soll.

In der **Realität** tritt diese Modenstruktur jedoch nicht so ausgeprägt auf, weil schon kleinste Abweichungen von der idealen „Geometrie“ und die zusätzliche Einführung von kleinsten Störungen im Schallfeld bereits zu einer starken Veränderung der Modenstruktur führen. Damit lassen sich die auftretenden Moden nicht mehr deterministisch berechnen, sondern es läßt sich nur noch Statistik über die Modendichte und die Art der Moden führen („statistische Raumakustik“). Beispielsweise beträgt die Modendichte ΔN im Frequenzbereich Δf für eine beliebige Raumgeometrie

$$\Delta N = V \cdot f^2 \cdot \frac{4\pi}{c^3} \cdot \Delta f \quad (\text{VI.3})$$

Die statistische Anwesenheit von Moden führt zu einer stark schwankenden Raumübertragungsfunktion als Funktion der Frequenz und als Funktion des Ortes. Sie schwankt typischerweise um etwa 10 dB und der Abstand benachbarter Maxima Δf beträgt etwa:

$$\langle \Delta f \rangle \approx \frac{4}{T} \quad (\text{VI.4})$$

Wobei T die Nachhallzeit des Raumes bezeichnet, d. h. die Zeit, in der die im Raum eingespeiste Schallenergie um 60 dB abgeklungen ist.

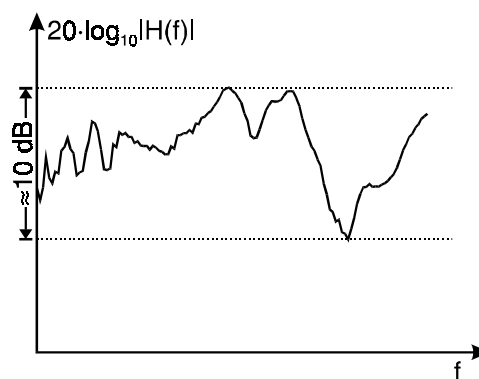


Abbildung 6.1: Schema einer im Mittel um ca. 10 dB schwankenden Raumübertragungsfunktion

VI.2 Statistische Raumakustik

Neben der statistisch schwankenden Raumübertragungsfunktion sind auch die zeitlichen Eigenschaften der Schallausbreitung in Räumen von

Bedeutung, die ebenfalls einer statistischen Betrachtung bedürfen. Für die Berechnung der Raumimpulsantwort (d. h. des Schalldruckverlaufs an einem bestimmten Ort im Raum als Funktion der Zeit, wenn an einem anderen Ort im Raum eine δ -förmige akustische Welle ausgesandt wird) kann z. B. die sogenannte **Spiegelschallquellenmethode** verwendet werden. Bei ihr wird die Reflexion eines „Schallstrahls“ an einer Wand dadurch beschrieben, daß der Raum an dieser Wand gespiegelt wird und die reflektierte Schallwelle scheinbar von einer „Spiegelschallquelle“ in diesem reflektierten Raum zustande kommt, von der aus eine geradlinige Schallausbreitung auf den Empfänger hin erfolgt.

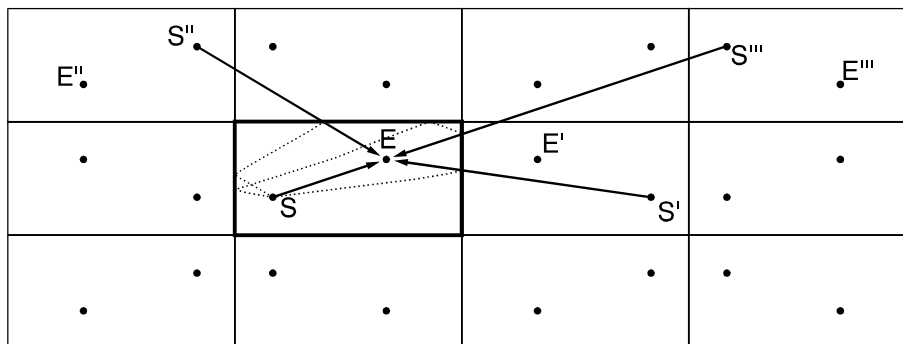


Abbildung 6.2: Spiegelschallquellenmethode (schematisch): Bei Empfänger E trifft neben dem Direktschall auch an den Wänden reflektierter Schall ein, dessen Weg durch Spiegelung des Raumes an den Grenzflächen gut beschrieben werden kann.

Die n -te Reflexion wird daher durch Spiegelsender aus dem n -fach gespiegelten Raum repräsentiert. Unter der Annahme eines quaderförmigen Raums, der an allen sechs Wänden eine Spiegelung zuläßt (d. h. nicht schallabsorbierend ausgekleidet ist), werden die ersten Reflexionen durch Spiegelschallquellen in den sechs dem ursprünglichen Quader benachbarten Quader repräsentiert. Zur Repräsentation der zweifachen Spiegelung müssen wiederum die Räume als Spiegelräume angeführt werden, die aus Spiegelungen der sechs Räume der ersten Generation hervorgehen. Wenn man diese Konstruktion fortsetzt, passen sämtliche Spiegelschallquellen, von denen zum Zeitpunkt t Reflexionen an den Empfänger einfallen, in eine Kugel mit dem Radius $R = c \cdot t$ (c : Schallgeschwindigkeit). Die Energiedichte E_n des zur Zeit t eintreffenden Impulses (ohne Berücksichtigung von Verlusten) müßte demnach betragen:

$$E_n = \frac{1}{4 \cdot \pi \cdot R^2} \cdot \frac{P}{c} \quad (\text{VI.5})$$

P bezeichnet dabei die gesamte abgestrahlte Leistung. Die Anzahl N der Impulse, die bis zum Zeitpunkt t beim Empfänger eintreffen, berechnet sich dabei aus dem Volumen einer Kugel mit dem Radius $R = c \cdot t$ geteilt durch das Volumen des Raumes V , also:

$$N = \frac{\frac{4}{3} \cdot \pi \cdot (ct)^3}{V} \quad (\text{VI.6})$$

Die Impulsdichte, d. h. die Anzahl der Impulse ΔN , die im Zeitintervall Δt beim Empfänger eintreffen, berechnet sich aus der Ableitung von VI.6 zu:

$$\Delta N = \frac{4 \cdot \pi \cdot (ct)^2}{V} \cdot c \cdot \Delta t \quad (\text{VI.7})$$

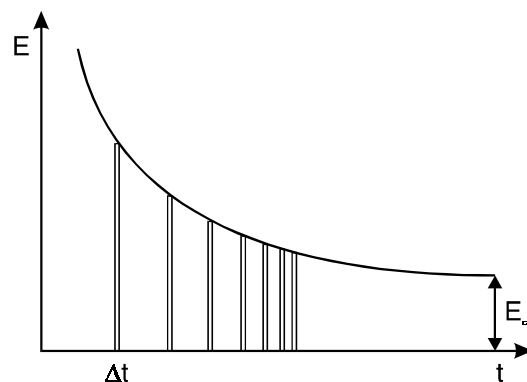


Abbildung 6.3: Schematischer Verlauf der Raumimpulsantwort nach der Spiegelschallquellenmethode

Als Asymptote E_∞ ergibt sich dann nach unendlich langer Zeit das Produkt aus Anzahldichte und Energie pro Impuls zu:

$$E_\infty = E_n \cdot \Delta N = \frac{P / c}{4\pi \cdot (ct)^2} \cdot \frac{4 \cdot \pi \cdot (ct)^2}{V} \cdot c \cdot \Delta t = \frac{P}{V} \cdot \Delta t \quad (\text{VI.8})$$

Dieses Ergebnis spiegelt die Annahme wieder, daß keine Energie verloren geht und beschreibt die Tatsache, daß gegen Ende des Nachhallvorgangs sämtliche einmal eingestrahlte Schallenergie in statistischer Weise auf einen Raum mit dem Volumen V verteilt ist, so daß keine weitere Abnahme (oder etwa Zunahme) der Energiedichte am Empfängerort zu erwarten ist. Obwohl dieser Fall hypothetisch ist, beschreibt er relativ gut den Endzustand eines „realen“ Nachhallprozesses, bei dem nach dem Eintreffen

der ersten (deterministischen) Reflexionen (von den verschiedenen Wänden und den ersten Spiegelschallquellen) der Nachhall eher durch eine statistische Anordnung von weiteren Reflexionen gekennzeichnet ist, die in den Rauschuntergrund übergehen.

In der **Realität** tritt jedoch Absorption auf, die zu einem endlichen Nachhallvorgang mit einem exponentiellen Abfall der Gesamt-Energie führt:

$$E = E_0 \cdot 10^{-6t/T} \quad (\text{VI.9})$$

Dabei bezeichnet T die Nachhallzeit d. h. die Zeit, in der die Schallenergie um 60 dB abgeklungen ist. Diese zunächst willkürlich erscheinende Definition beruht auf den historischen Untersuchungen von W. C. Sabine, der schon ca. 1900 raumakustische Messungen durchgeführt hat. Er definierte die **Nachhalldauer** als die Zeit, in der ein von einem Knall ausgelöster Nachhallvorgang in einem Raum noch hörbar ist. Aufgrund der Dynamik des Ohres und der begrenzten maximalen Energie, die mit Knallquellen in einem Raum erzeugt werden kann (z. B. mechanische Klappe, Knallfunkengenerator, Pistolenschuß) beträgt der Dynamikbereich zwischen der Schallanregung und der Hörschwelle etwa 60 dB, so daß die von Sabine definierte Nachhalldauer sehr gut mit der exakt definierten Nachhallzeit übereinstimmt. Die von Sabine empirisch angegebene Formel für die Nachhallzeit beschreibt ihre Abhängigkeit vom Raumvolumen V und der effektiven Absorptionsfläche A folgendermaßen:

$$T = \frac{24 \cdot \ln 10}{c} \cdot \frac{V}{A} = \frac{0,163 V [\text{m}^3]}{A [\text{m}^2]} \quad [\text{s}] \quad (\text{VI.10})$$

Die Größe A wird als „äquivalente offene Fensterfläche“ bezeichnet, d. h. als die Fläche, auf die man sämtliche absorbierenden Flächen unterbringen könnte, wenn sie den Absorptionsgrad $\alpha = 1$ aufweisen würden, d. h.

$$A = \sum S_i \cdot \bar{\alpha}_i \quad (\text{VI.11})$$

(S_i = Größe der i-ten Fläche des Raumes mit der Absorption α_i). Eine heuristische Herleitung dieser Nachhallformel folgt aus der Annahme, daß die über den gesamten Raum verteilte Energiedichte E mit dem Exponenten $2 \cdot \delta_\alpha$ abnimmt also:

$$-\frac{d}{dt}(VE) = 2 \cdot \delta_{\alpha}(VE) \quad , \quad \text{mit } E = E_0 \cdot e^{-2\delta_{\alpha}t} \quad (\text{VI.12})$$

Für diesen Abklingkoeffizienten ergibt sich durch Integration über den Raumwinkelbereich, aus dem die Energie auf die Fläche S_i einfällt, ein Wert von (vgl. Cremer, Hubert, Kap. 3.2):

$$2\delta_{\alpha} = \frac{c \cdot A}{4V} \quad (\text{VI.13})$$

Wenn noch zusätzlich die Dissipation von Schall in Luft mit der Abklingkonstante μ angenommen wird, gilt also für die reine Dissipation in Luft:

$$E = E_0 \cdot e^{-2\mu \cdot c \cdot t} \quad (\text{VI.14})$$

so daß für die Gesamt-Abklingkonstante 2δ gilt:

$$2\delta = c \cdot \left(\frac{A}{4V} + 2\mu \right) \quad (\text{VI.15})$$

$$T = \frac{0,163V}{A + 8\mu V} \quad (\text{VI.16})$$

Diese Formel wurde von V. O. Knudsen in die Raumakustik eingeführt und stellt eine bessere Näherung für realistische Räume dar als die Formel nach Sabine. In beiden Formeln ist jedoch eine gleichmäßige Verteilung der Absorption über den gesamten Raum vorausgesetzt und eine Gleichverteilung der Energie auf alle Raumeinfallrichtungen („diffuses“ Schallfeld“). In der Realität können jedoch erhebliche Abweichungen auftreten, wenn beispielsweise nur eine Fläche absorbierend ausgekleidet ist und die anderen Flächen nicht absorbierend oder wenn die gleiche Absorberfläche im ersten Fall nur auf einer Wand und im anderen Fall auf sämtliche Wände verteilt ist. Dabei können erhebliche Abweichungen von den (empirischen) Nachhallformeln von Knudsen und Sabine auftreten, die sich z. B. durch numerische Simulationen der Raumakustik erst gut erklären lassen. Durch verschiedene Techniken des „Ray-Tracing“ ist es heute möglich, die Raumakustik (insbesondere die Nachhallzeit) bei vorgegebener Raumgeometrie und vorgegebenen Absorptionsgraden der einzelnen Flächen sehr gut vorherzusagen (vgl. Kuttruff, Roomacoustics). Ein weiterer wichtiger Parameter zur Beschreibung der Eigenschaften des

Schallfeldes in einem realen Raum ist der **Hallradius**, d. h. der Radius des Umkreises um eine Schallquelle, in dem der Direktschallanteil größer ist als der Anteil des Nachhall-Schallfeldes. Innerhalb des Hallradius überwiegt damit der Eindruck des Direktschalls (z. B. mit einer hohen Sprachverständlichkeit) während außerhalb des Hallradius das diffuse Nachhallfeld dominiert, das aufgrund der Verwürfelung der Zeitfunktion und der Korrelation des ursprünglichen Signals beispielsweise eine schlechtere Verständlichkeit des Quellensignals bewirkt. Der Nachhallradius in Metern beträgt (gemäß Cremer, Hubert, 1992)

$$r_H = 0,057 \sqrt{\frac{V[\text{m}^3]}{T[\text{s}]}} [\text{m}] \quad (\text{VI.17})$$

Genau wie oben bezeichnet V dabei das Volumen des Raumes und T die Nachhallzeit (in Sekunden).

Die „**optimale**“ **Nachhallzeit** eines Raumes hängt von seiner gewünschten Eigenschaft ab: Ein **reflexionsarmer Raum** sollte eine Nachhallzeit von weniger als 0,01 sec. aufweisen, ein Konferenzraum, in dem eine gute Sprachübertragung möglich sein soll, sollte etwa Nachhallzeiten zwischen 0,5 und 1,5 sec. aufweisen, und Konzertsäle sollten je nach der verwendeten Musik Nachhallzeiten zwischen 1,5 und 2,5 sec. (klassische Musik) bzw. 2-10 sec. (geistliche Musik) haben. Speziell für die in Kirchen auftretenden Nachhallzeiten von mehreren Sekunden ist die geistliche Musik angepaßt, bei der langanhaltende, mit geringen Modulationsfrequenzen versehene Töne dominieren (z. B. gregorianische Gesänge). Eine extreme Nachhallzeit tritt bei einem **Hallraum** auf, d. h. ein Raum mit möglichst geringgradig absorbierenden Wänden, der eine Nachhallzeit von größer als 10 sec. aufweisen sollte. Ein derartiger Hallraum wird u. a. für die Messung des Absorptionsgrades von Schallabsorbern verwendet, weil das in ihm auftretende statistische Wellenfeld eine Mittelung über sämtliche mögliche Einfallsrichtungen des Schalls ermöglicht.

Neben der Nachhallzeit sind für den (subjektiven) Wohlklang eines Raumes auch andere akustische Faktoren wichtig, deren Bedeutung in umfangreichen Studien (u. a. durch Simulation verschiedener Konzertsäle, Kunstkopftechnik) eingehend untersucht wurde. Gemäß den Untersuchungen von M. R. Schroeder ist ein wichtiger subjektiver Faktor die interaurale (d.h. zwischen den Ohren auftretende) Ungleichheit der ersten den Zuhörer nach dem Direktschall erreichenden Reflexionen. D. h. die erste Reflexion sollte möglichst nicht von der Decke, sondern beispielsweise von der Seitenwand auf den Zuhörer einfallen, weil bei der Deckenreflexi-

on der gleiche Schall auf beide Ohren gelangt, während bei einer seitlichen Reflexion der Schall zeitversetzt die beiden Ohren des Zuhörers erreicht. Ein Maß für die Ähnlichkeit der an beiden Ohren anliegenden Schallsignale ist die interaurale Kreuzkorrelation, die möglichst gering sein soll, um einen möglichst guten räumlichen Eindruck des Konzertsaals zu vermitteln. Diese Forderung steht allerdings im krassen Widerspruch zu der gängigen Praxis, große Konzertsäle möglichst breit und mit möglichst niedriger Decke zu bauen, um bei vorgegebenem Bauvolumen eine möglichst große Anzahl von Personen unterbringen zu können. Dies führt zu einem subjektiv schlechten Klang von modernen, großen Konzertsälen mit niedriger Decke, die im Vergleich zu historischen Sälen mit relativ hoher Decke („Schuhkastenform“, z. B. der Wiener Musikvereinssaal), wesentlich schlechter ausfallen. Eine Abhilfe bieten die von M. R. Schroeder vorgesehenen Phasendiffusoren auf der Basis der quadratischen Reste. Bei diesen als **Schroeder-Diffusoren** bezeichneten quadratischen Rest-Diffusoren wird beispielsweise an der Decke eines Konzertsaals reflektierendes Material mit einer Oberflächenstruktur angebracht, die eine Phasendrehung der einfallenden Schallwelle als Funktion des Ortes bewirkt, die gerade proportional zum Quadrat des Ortes (modulo einer Primzahl P) ist:

$$r_n = e^{2\pi i n^2 / P} \quad (\text{VI.18})$$

Dabei bezeichnet n die räumliche Position des Reflektorblocks und P die verwendete Primzahl.

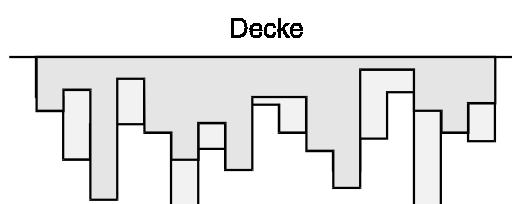


Abbildung 6.4: Schematische Bauform eines quadratischen-Restklasse Diffusors

Eine derartiger Diffusor zeigt ein breites Streuverhalten ohne Absorption der einfallenden Schallwelle. Ein weiterer Vorteil ist, daß die „spiegelnde“ Reflexion eliminiert wird, d. h. der auf die Decke fallende Schall wird in alle Richtungen zerstreut und nicht in die der geometrischen Akustik gehorchenden reflektierte Richtung geworfen.

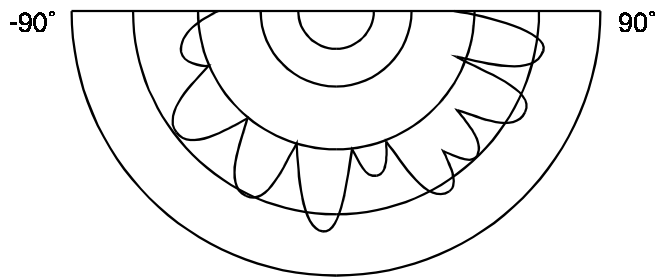


Abbildung 6.5: Richtdiagramm (schematisch) der Reflexion des in Abbildung 6.4 dargestellten Diffusors bei Schalleinfall aus 0-Grad-Richtung

VI.3 Subjektive Raumakustik

Die subjektive Raumakustik beschäftigt sich mit der Optimierung der Raumakustik anhand von (subjektiven) Aussagen der Hörer über die empfundene Qualität der Akustik. Ein derartiger Qualitätsvergleich wird erst durch die Technik des direkten Vergleichs möglich (d.h. durch „Hineinhören“ in verschiedene Konzertsäle, Studios etc.). Zur technischen Realisierung dieses Vergleichs steht die **Kunstkopftechnik**, die **Modelltechnik** oder die **digitale Simulation** zur Verfügung. Bei der Kunstkopftechnik wird eine Aufnahme des Schallfeldes (z. B. eines Konzertes) im realen Raum mit einem Kunstkopf durchgeführt, d. h. einer Kopfnachbildung mit nachgebildeten Außenohren, bei der anstelle der Trommelfelle Mikrophone angeordnet sind, um die akustische Übertragung vom Freifeld zum wahrgenommenen akustischen Schall nachzubilden. Beim Abspielen dieser Kunstkopfaufnahmen über Kopfhörer entsteht der Eindruck, daß man sich an derselben Stelle in dem Raum befindet, wie bei der Aufnahme der Kunstkopf. Ein ähnlicher Effekt kann durch die digitale Filterung eines „trockenen“ (d. h. in einem reflexionsarmen Raum aufgenommenen) Musiksignals mit der Raumimpulsantwort erreicht werden, die bei Abstrahlen eines δ -Impulses von der Bühne mit einem Kunstkopf aufgenommen wurde. Auf diese Weise kann dasselbe Musiksignal beim Abspielen in unterschiedlichen Räumen nachgebildet werden, so daß ein Vergleich der Räume unabhängig vom verwendeten Musiksignal möglich ist.

Bei der **Modelltechnik** (beispielsweise im Maßstab 1:10 oder 1:50 = $1/\gamma$) wird dagegen die Berechnung bzw. Messung einer Raumimpulsantwort nicht von realen, sondern nachgebildeten Räumen möglich, was beispielsweise für die Planungsphase eines Konzertsaals von großer Bedeutung ist. Dabei wird das akustische Signal (z. B. Musik) mit einem Frequenztransformator im Verhältnis $1:\gamma$ zu höheren Frequenzen transformiert und mit einem Miniatur-Lautsprecher im originalgetreu nachgebildeten Modellraum abgestrahlt. In diesem Modellraum befindet sich ein Modell-Kunstkopf (im selben Verkleinerungsmaßstab wie der Modellraum), mit

dem das Signal aufgenommen, im Frequenzverhältnis $\gamma:1$ heruntertransformiert und anschließend über Kopfhörer einer Beurteilungs-Person vorgespielt wird.

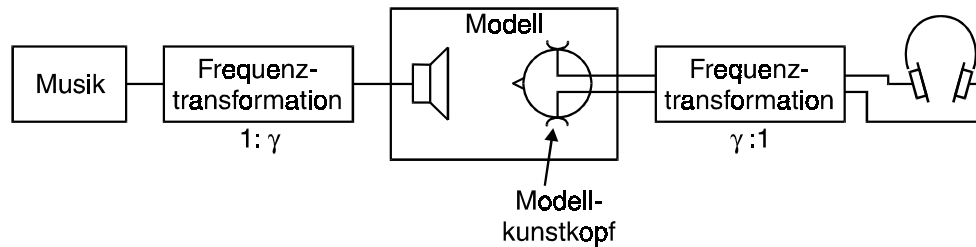


Abbildung 6.6: Beurteilung der Raumakustik durch die Modelltechnik

Der Vorteil dieser Methode liegt darin, daß ein akustisches Design bereits vor Fertigstellung des realen Gebäudes erfolgen kann und der Einfluß etwaiger Modifikationen des Baues direkt abgeschätzt werden kann. Der Nachteil dieser Methode ist der relativ hohe Aufwand bei der Erstellung des Modells, die zumeist falschen Reflexionseigenschaften der Materialien im verwendeten höheren Frequenzbereich und die falsche Dämpfung der Luft im Modellraum, die nicht im gleichen Maße transformiert werden kann wie die Frequenz. Daher verwendet man zumeist getrocknete Luft oder Stickstoff als Atmosphäre im Modellraum.

Als dritte Möglichkeit, mit der auch Räume in der Planungsphase akustisch simuliert werden können, bietet die **digitale Simulation** die Konstruktion der Impulsantwort zwischen dem Ort eines Schallsenders in einem virtuellen Raum und dem Empfänger im virtuellen (u. U. geplanten) Raum.

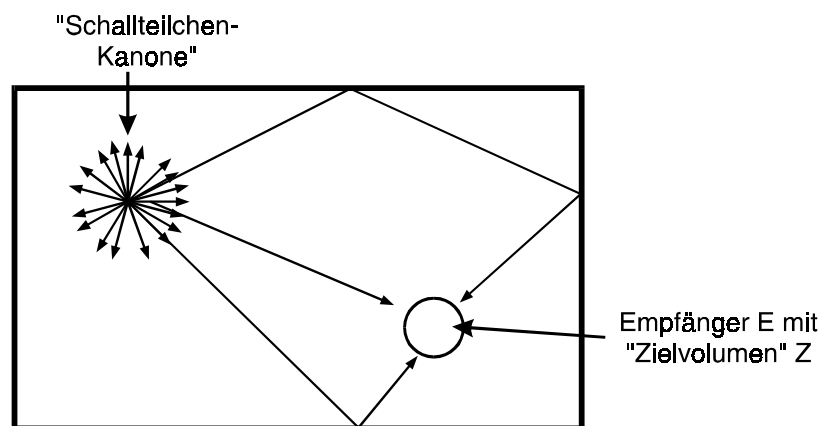


Abbildung 6.7: Berechnung einer Raumimpuls-Antwort nach dem Ray-tracing-Verfahren

Für jede beliebige, im Rechner angenommene Geometrie läßt sich so unter der Annahme einer „Schallteilchenkanone“ am Ort des Schallsenders und eines „Zielvolumens“ am Ort des Empfängers eine binaurale Raumimpulsantwort berechnen. Je mehr Details des in Planung befindlichen Raums in die digitale Simulation einbezogen werden können (z. B. zusätzliche reflektierende Flächen, effektiver Absorptionsgrad der einzelnen Flächen, Frequenzgang der Absorption), desto realistischer kann die Raumsimulation erfolgen. Eine typische damit errechnete Raumimpulsantwort sieht schematisch folgendermaßen aus:

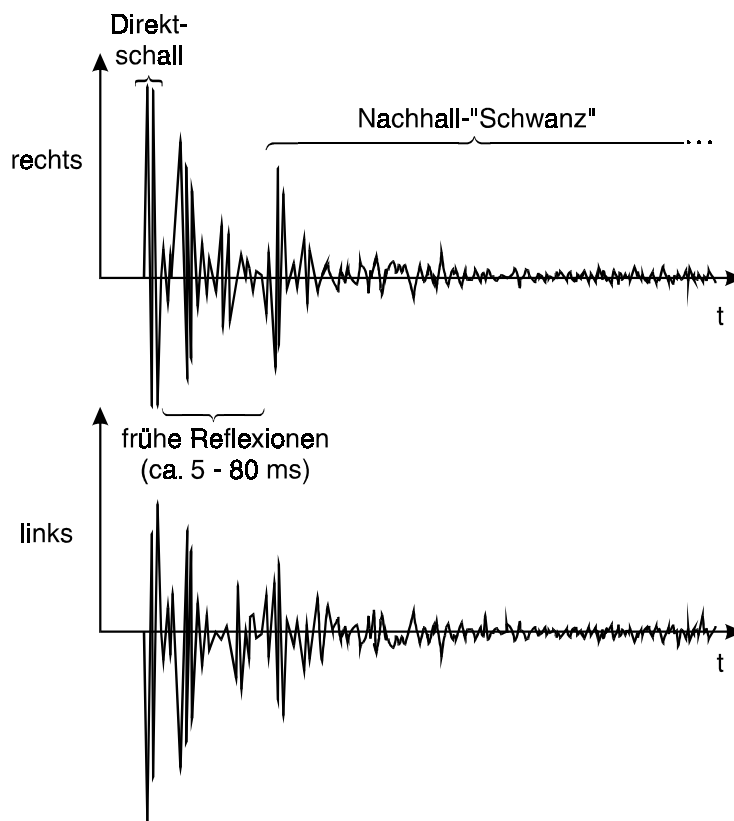


Abbildung 6.8: Schematischer Verlauf einer binauralen Raumimpulsantwort

Obwohl in diesen Raumimpulsantworten die ersten Reflexionen unter Nachhall unter Umständen wesentlich mehr Energie beinhalten als der erste Anteil der Impulsantwort, die durch den **Direktschall** hervorgerufen wird, wird die Schallquelle zumeist richtig in der Richtung geortet, aus der der Direktschall kommt.

Diese subjektive Unterdrückung der frühen Reflexion des Nachhalles auf die empfundene Lokalisation wird als **Gesetz der ersten Wellenfront** („Präzedenzeffekt“) bezeichnet. Er wird beispielsweise bei Beschallungsanlagen in guten Auditorien verwendet, bei denen die über Lautsprecher in rückwärtigen Anteilen des Raumes abgestrahlten, verstärkten Schallsi-

gnale zeitlich gegenüber dem Sprecher verzögert sind, so daß zunächst das (unverstärkte) Signal vom Sprecher direkt zum Zuhörer gelangt und erst nach einer gewissen Verzögerung das verstärkte Signal. Auf diese Weise wird der Lokalisationseindruck nicht gestört, obwohl zugleich der Sprecherpegel angehoben wird. Die Unterdrückung des ersten Echos bei der Wahrnehmung des Ortes der ersten Wellenfront wird im folgenden Bild dargestellt.

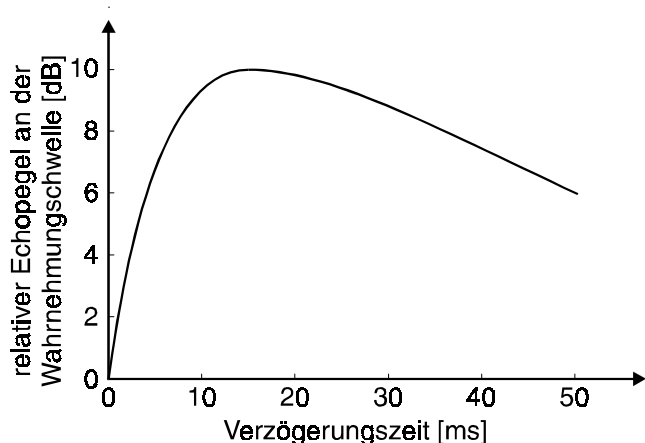


Abbildung 6.9: Unterdrückung wahrgenommener Echos als Funktion der Zeitverzögerung zwischen Direktschall und Echo

Bei sehr kleinen Verzögerungen überlagert sich das erste Echo mit dem Direktschall zur Bildung einer neuen Richtung (Pseudorichtung), d. h. Direktschall und erste Reflexionen werden zusammen zu einem neuen Lokalisationsort zusammengefaßt. Mit zunehmender Verzögerungszeit nimmt die Empfindlichkeit gegenüber dem zusätzlichen Echo ab, so daß die Wahrnehmungsschwelle für das Echosignal stark ansteigt. Bei weiter zunehmender Verzögerung kann man das Echo getrennt vom Direktschall wahrnehmen, so daß die relative Echo-Schwelle wieder sinkt.

Anhand der o. a. schematisch dargestellten Raum-Impulsantworten lassen sich verschiedene Qualitätskriterien für die Akustik eines Konzertsaals oder Auditoriums ableiten, die mit dem subjektiven Eindruck korrelieren. So wird der erste Anteil der Impulsantwort mit dem Direktschall und den frühen Reflexionen als „nützlich“, die Klarheit und die Verständlichkeit der damit zu übertragenden Sprach- und Musiksignale erhöhend angesehen. Dagegen wird der nach einer bestimmten Zeit t_2 auftretende „Schwanz“ der Raumimpulsantwort, der diffuse Nachhall, als „schädlich“ bzw. als Störschall angesehen. Gemäß diesen Vorstellungen können zwei Zeiten t_1 bzw. t_2 zur Abgrenzung von nützlichen und schädlichen Anteilen der Raumimpulsantwort definiert werden, so daß eine Gewichtsfunktion $a(t)$ den „gewünschten“ Anteil bezeichnet. Die in diesen beiden Anteilen lie-

gende Energie der Impulsantwort wird durch folgende Integrale ausgedrückt:

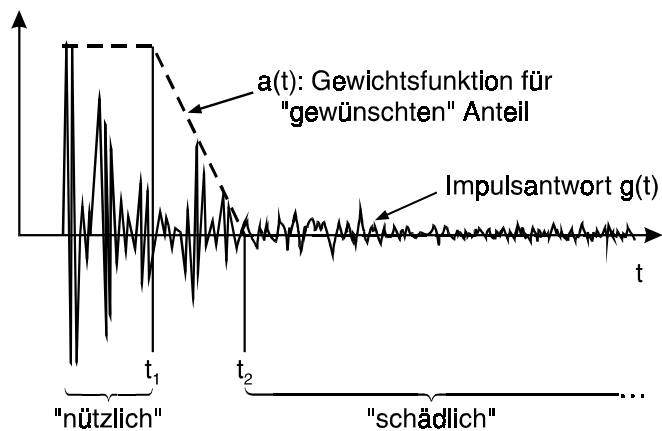


Abbildung 6.10: Schematische Darstellung einer Raum-Impulsantwort und Aufteilung in Nutz- und Störschallanteil

$$E_N = \int_0^{t_1} a(t)[g(t)]^2 dt, \quad E_S = \int_{t_2}^{\infty} [g(t)]^2 dt \quad (\text{VI.19})$$

Darauf aufbauend wurden in der Literatur die folgenden **subjektiven Qualitätskriterien** definiert:

„Deutlichkeit“ (Thiele)	$\frac{E_N}{E_N + E_S}$	$t_1 = t_2 = 50 \text{ ms}$
Klarheitsmaß (Reichardt)	$10 \log \frac{E_N}{E_S}$	$t_1 = t_2 = 80 \text{ ms}$
Echograd (Niese)	$\frac{E_N}{E_N + E'_S} \leftarrow \text{nur Nachhallspitzen}$	$t_1 = 17 \text{ ms}$ $t_2 = 33 \text{ ms}$
Störabstand (Lochner und Bergner)	$10 \log \frac{E_N}{E_S}$	$t_1 = 35 \text{ ms}$ $t_2 = 95 \text{ ms}$

Beim Echograd nach Niese bezeichnet E'_S das Integral über den Nachhall-„Schwanz“ bei dem nicht die Impulsantwort selbst, sondern nur die im Nachhall auftretenden Spitzen der Impulsantwort ausgewertet und linear interpoliert werden. Sämtliche dieser subjektiven Qualitätskriterien liefern eine ähnliche Aussage über das Verhältnis von „nützlichen“ Schallanteilen der Raumübertragung im Verhältnis zum diffusen Nachhallfeld. Eine Überprüfung dieser Kriterien durch den Vergleich mit empirisch erhobenen subjektiven Qualitätskriterien von Beobachtern ist jedoch nicht in allen Fällen erfolgt.

Für die Beurteilung von Konferenz- und Sprechräumen ist die **Sprachverständlichkeit in Räumen** von großer Bedeutung. Sie kann anhand von Messungen mit einer größeren Zahl von Versuchspersonen entweder direkt bestimmt werden (was allerdings einen großen Aufwand bedeutet) oder durch Berechnungsverfahren bestimmt werden, die aufgrund von empirischen Daten zur Abhängigkeit der Sprachverständlichkeit von verschiedenen akustischen Parametern aufgestellt wurden. Die wichtigste dabei zugrundeliegende Annahme ist, daß der gesamte Hör-Frequenzbereich in verschiedene Frequenzbänder (z. B. Oktav-, Terz- oder Frequenzgruppenbänder) aufgeteilt wird. Jedes dieser Frequenzbänder (mit dem Index k) trägt nun mit einem Gewichtungsfaktor g_k zur Verständlichkeit bei, wobei dieser Beitrag maximal ist, wenn im jeweiligen Frequenz-Band bloß das ungestörte Sprachsignal vorliegt. Im Fall einer kompletten Störung des Sprachsignals im Frequenzkanal k ist dieser Beitrag dagegen 0. Als Articulation-Index (AI) bzw. als Speech Transmission Index (STI) wird nun die Summe über sämtliche gewichtete Frequenz-Bänder der in diesen Bändern vorliegenden Signal-Rauschverhältnisse bezeichnet:

$$\left. \begin{array}{l} \text{AI} \\ \text{STI} \end{array} \right\} = \sum_k g_k \left(\frac{S}{N} \right)^{(k)} \quad (\text{VI.20})$$

Dabei bezeichnet $(S/N)^{(k)}$ den Signal-Rausch-Abstand in Band k , der auf den relevanten Pegelbereich (z. B. +15 dB bis -15 dB) begrenzt ist. Diese Begrenzung ist sinnvoll, weil bei Signal-Rausch-Abständen von 15 dB bereits eine maximale Sprachverständlichkeit erreicht wird und eine weitere Erhöhung keinen weiteren Gewinn bringt bzw. bei Signal-Rauschabständen von -15 dB bereits eine Sprachverständlichkeit von 0 vorliegt und eine weitere Verringerung des Signals-Rausch-Abstandes diese Situation nicht weiter verschlechtern kann. Die Gewichtungsfaktoren g_k stellen dabei eine vom Sprachmaterial abhängige Frequenzgewichtung dar. Beispielsweise weisen bei Sätzen und mehrsilbigen Wörtern die tiefen Frequenzen eine relativ höhere Wichtigkeit auf als bei einsilbigem Sprachmaterial, bei dem weniger Redundanz im Material zugrundeliegt. Für die vollständige Erkennung der Sprache sind zudem die hochfrequenten Anteile der Konsonanten von entscheidenderer Bedeutung als bei Sätzen. Der resultierende Index ist dann eine auf den Bereich von 0 bis +1 begrenzte Größe, die ein Maß für die Verständlichkeit darstellt und streng monoton mit der Sprachverständlichkeit für das jeweils verwendete Sprachmaterial zusammenhängt. Dieser Zusammenhang zwischen Articulation-Index bzw. Speech-Transmission-Index und Sprachverständlichkeit muß für jeden Sprachtest getrennt ermittelt werden, allerdings gibt es Anhaltswerte für bestimmte Sprachmaterialien. Die beiden Indizes unter-

scheiden sich in der Ermittlung des „effektiven“ Signal-Rauschabstandes in Band k (S/N)^(k). Beim Articulation Index muß das Leistungsspektrum sowohl vom Sprachsignal (mittleres Leistungsspektrum) als auch vom Störsignal vorliegen, so daß der Signal-Rausch-Abstand direkt berechnet werden kann. Dieses Verfahren ist daher nicht anwendbar, wenn beispielsweise nur das Nutzsignal und die Summe aus Nutz- und Störsignal vorliegt. Eleganter ist daher die beim Speech-Transmission-Index benutzte Methode, bei der der effektive Signal-Rauschabstand aus der **Modulations-Transfer-Funktion (MTF)** berechnet wird. Sie wird gemessen, indem man ein sinusförmig modulierte Signal (z. B. ein sinusförmig amplitudenmoduliertes Schmalbandrauschen) in das auszumessende System schickt und am Ausgang den verbleibenden Modulationsgrad mißt.

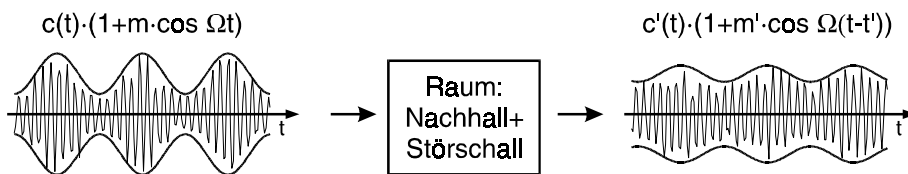


Abbildung 6.11: Schematische Auswirkung der Raumakustik auf Modulationen eines Eingangssignales (linkes Teilbild), die eine Verringerung der Modulationen im Ausgangssignal (rechtes Teilbild) bewirkt

Wenn wir den Träger des Eingangssignals mit $c(t)$, den Modulationsgrad mit m und die Modulationsfrequenz mit Ω bezeichnen, dann beträgt das Eingangssignal:

$$c(t) \cdot (1 + m \cdot \cos \Omega t) \quad (\text{VI.21})$$

und das Ausgangssignal:

$$c'(t) \cdot (1 + m' \cdot \cos \Omega(t - t')) \quad (\text{VI.22})$$

Wenn im Übertragungskanal (z.B. im Raum) nun ein stationäres Rauschen dem Eingangssignal aufaddiert wird oder ein Nachhallvorgang auftritt, werden die Modulationsmaxima und -minima einander angenähert, so daß ein „Auffüllen“ der Modulationstäler auftritt, das mit zunehmender Nachhallzeit und niedrigerem Signal-Rausch-Abstand stärker ausgeprägt ist. Auf diese Weise kann man aus der Abschwächung der Modulations-Übertragung auf den effektiven Signal-Rausch-Abstand im Ausgang des Systems schließen, wobei der Einfluß von Nachhall und Rauschen auf die gleiche Weise behandelt wird und zudem nur das Eingangssignal und das Ausgangssignal bekannt sein muß, nicht jedoch das Rauschen und das

Stör- und Eingangssignal getrennt. Die Modulationstransferfunktion ist nun definiert als:

$$\text{MTF}(\Omega) = \frac{m'}{m}(\Omega) \circ - \bullet g^2(t) \quad (\text{VI.23})$$

$g^2(t)$ bezeichnet die quadrierte Raum-Impulsantwort.

Daraus folgt

$$\left(\frac{S}{N}\right)^{(k)} = \left\langle \frac{m'}{m-m'} \right\rangle_{\Omega} = \left\langle \frac{\text{MTF}}{1-\text{MTF}} \right\rangle_{\Omega} \quad (\text{VI.24})$$

Die eckige Klammer bezeichnet dabei eine Mittlung über sämtliche für die Sprachwahrnehmung relevanten Modulationsfrequenzen Ω , die nach Houtgast und Steeneken (1972) im Bereich zwischen 1 Hz und 50 Hz liegen. Beim sogenannten **RASTI-Verfahren** (Rapid-STI) wird die Modulations-Transferfunktion nur bei 500 Hz und 2 kHz gemessen und auch nur bei den 4 Modulationsfrequenzen im Bereich zwischen 0,7 Hz und 11,2 Hz ausgewertet. Mit dieser Methode läßt sich sehr schnell die MTF in Räumen bestimmen und daraus der STI ableiten. Diese Methode hat sich aufgrund des einfachen Meßvorgangs und der relativ hohen Robustheit des Verfahrens für die Anwendung in der Raumakustik bewährt, so daß es kommerzielle Geräte auf dem Markt gibt, die diese Messungen automatisch durchführen. In jüngerer Zeit ist man dazu übergegangen, Elemente des AI und STI in einer neuen Norm zusammenzufassen und zu einem Speech Intelligibility Index (SII) zusammenzufassen. Es bleibt jedoch festzuhalten, daß die Anwendung von AI, STI und SII vorwiegend für additives Störgeräusch und Nachhall sowie für den über viele Versuchspersonen und Sprachtests gemittelten Verständlichkeitswert beschränkt sind, während die Aussagen des Verfahrens für andere Arten von Störungen (z. B. nichtlineare Verzerrungen, Dynamikkompression) und für kurze Sprachproben und eine kleinere Anzahl von Versuchspersonen nicht hinreichend gültig sind. (Kollmeier. 1992)