

Ein Algorithmus zur Vorhersage des für transparente Auralisierung nutzbaren Dynamikbereichs rauschbehafteter Impulsantworten

Florian Völk, Tobias Riesenweber, Hugo Fastl

AG Technische Akustik, MMK, Technische Universität München, 80333 München, Deutschland, florian.voelk@mytum.de

Einleitung

Da die Impulsantworten (impulse responses, IRs) linearer, zeitinvarianter akustischer Systeme eine einfache und zugleich vollständige Beschreibung dieser Systeme darstellen, werden sie häufig für technische Anwendungen, die eine solche Systembeschreibung erfordern, eingesetzt. Auralisierung, binaurale Synthese oder Faltungshall seien hier als typische Anwendungen genannt. Werden IRs durch Messung bestimmt, überlagert sich der aufgenommenen IR selbst unter idealen Messbedingungen zumindest das inhärente Rauschen des Messsystems und -verfahrens. Unter nicht idealen Bedingungen können weitere Störanteile auftreten. In diesem Beitrag wird eine Berechnungsvorschrift entwickelt, die es erlaubt, basierend auf bekannten Hörbarkeitsschwellen den maximal zur Auralisierung nutzbaren Dynamikbereich einer beliebigen IR zu prognostizieren. Ein Vergleich der Modell-Prognosen mit in psychoakustischen Experimenten ermittelten Daten zeigt Abweichungen zwischen den interindividuellen Medianen der psychoakustischen Messwerte und den entsprechenden Vorhersagen kleiner als ± 2 dB.

Dynamikbereich von Impulsantworten

Eine ideale akustische IR fällt über der Zeit im Pegelmaß abhängig vom Aufnahmeraum unterschiedlich schnell unendlich weit ab. In allen realen Fällen überlagert sich der idealen, hier ohne Einschränkung der Allgemeinheit digital angenommenen IR zumindest das durch Aufnahmeequipment und -methode verursachte additive Rauschen. Die Differenz zwischen Maximalpegel $L_{\text{IR,MAX}}$ der IR und RMS-Pegel des Rauschens L_N wird hier als Signal-Rauschverhältnis (signal to noise ratio, $\text{SNR} = L_{\text{IR,MAX}} - L_N$) bezeichnet. Da der Rauschteppich keine Nutzinformation enthält, wird die rauschbehaftete IR üblicherweise zeitlich begrenzt, nachdem der Momentanpegel $L_{\text{IR}}[n]$ des informationstragenden Teils in den Bereich von L_N abgefallen ist. Hier wird dafür ein gaussförmiges Fenster verwendet, das sein Maximum zum Abtastzeitpunkt n_0 erreicht, zu dem $L_{\text{IR}}[n]$ zum ersten mal kleiner als $L_N + 4$ dB wird (vgl. Einblendung in Abb. 1).

Nach Müller und Massarani (2001) sind für Auralisierung mindestens 90 dB SNR in IRs nötig. Zur Bestimmung des nutzbaren Dynamikbereichs einer IR ist das SNR alleine allerdings nur wenig aussagekräftig, da unbekannt ist, ab welchem Wiedergabepiegel Differenzen zur idealen IR hörbar werden. Hörbare Abweichungen können sich durch die notwendige zeitliche Begrenzung und das additive Rauschen ergeben.

Methode und Vorgehen

In der Wiedergabesituation wird ein (für vorliegende Betrachtung ohne Einschränkung der Allgemeinheit digital angenommenes) Audiosignal $s[n]$ mit der betrachteten IR gefaltet. Das Faltungsprodukt setzt sich also als additive Überlagerung zeitlich verschobener, gewichteter Kopien der IR zusammen. Die zeitliche Begrenzung der IR wird in vielen Fällen von nachfolgenden Signalanteilen verdeckt. Die Wahrscheinlichkeit, dass das additive Rauschen hörbar ist, nimmt mit zunehmendem zeitlichen Abstand vom Beginn der IR zu, da der Momentanpegel der idealen IR über der Zeit abnimmt. Auch hier kann also Verdeckung durch nachfolgende Signalanteile auftreten. Somit ist die Hörbarkeitsschwelle von Übertragungsfehlern für ein Klick-Signale mit hohem Pegel (also für großen Gewichtungsfaktor) ohne nachfolgende Signalanteile zu erwarten. Da das additive Rauschen üblicherweise breitbandig ist und typische akustische IRs frequenzabhängig unterschiedlich schnell abfallen (vgl. z. B. die Nachhallzeit eines typischen Raumes), ist anzunehmen, dass das additive Rauschen während des Nutzanteils der IR, also vor der zeitlichen Begrenzung, hörbar wird. Im Folgenden wird diese Annahme bestätigt und anschließend die Hörbarkeitsschwelle des Rauschens verschiedener IRs ermittelt.

Die hier vorgestellten Experimente wurden in einer abgedunkelten, schallisolierten Hörkabine durchgeführt. Da die Präsentation der Schalle diotisch und kalibriert über Beyerd DT 48 Kopfhörer mit vorgeschaltetem Freifeldentzerrer nach Fastl und Zwicker (2007) erfolgte, sind die angegebenen Pegel als freifeldäquivalente Schalldruckpegel zu verstehen. Mittels eines 3-AFC Verfahrens kombiniert mit einer 2up/1down-Regel variierten die Probanden den Pegel der Testschalle. Dafür bekamen sie (visuell oder akustisch) drei zeitliche Intervalle in Folge vorgegeben (0.5 s Pause) und beurteilten mittels Tastatendruck, welches der Intervalle sich von den beiden anderen unterschied. Über die Präsentationsreihenfolge und ein- oder zweimalige Testschallpräsentation wurde zufällig entschieden. Vor dem Versuch wurden die Probanden trainiert, bis sie drei korrekte Urteile in Folge abgegeben hatten.

Hörversuche und Ergebnisse

Um die Hypothese zu prüfen, dass das additive Rauschen und nicht die zeitliche Begrenzung die Hörbarkeit von Übertragungsfehlern in IRs bestimmt, wurde eine IR mit SNRs zwischen 40 und 100 dB in 0.5 dB Schritten aufgenommen. Die IR mit dem höchsten SNR diente als Referenz IR_{ref} . Im Experiment, dessen Ergebnisse Abb. 1

zeigt, wurden die Unterscheidbarkeitsschwellen der wie üblich verarbeiteten IR (MR) und der bei entsprechenden Pegeln gefensterten IR_{ref} (OR) zu IR_{ref} bei zwei Wiedergabepegeln (Index) bestimmt (9 Versuchspersonen, im Mittel 25 Jahre, mittlere Dauer 8.9 Minuten). Dargestellt sind die interindividuellen Mediane und Interquartilbereiche der intraindividuellen Mittelwerte.

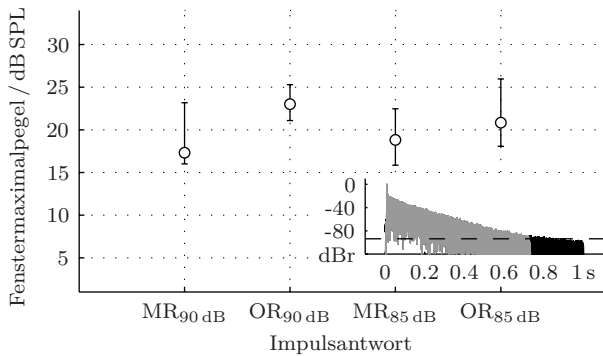


Abbildung 1: Unterscheidbarkeitsschwelle zwischen idealer und realer Impulsantwort. MR - mit Rauschen, Fensterung 4 dB über Rauschpegel (vgl. Einblendung), OR - ohne Rauschen, Index - Maximalpegel.

Es zeigt sich, dass das Rauschen unabhängig vom Wiedergabepegel bei hochsignifikant geringeren Pegeln (zweifaktorielle ANOVA, Faktor Rauschen [F(1,8)=13.54; p=0.006], Faktor Wiedergabepegel [F(1,8)=0.08; p=0.791], Interaktion [F(1,8)=0.95; p=0.357]) als die zeitliche Begrenzung der selben Impulsantwort hörbar wird. Die Hörbarkeitsschwelle des Rauschens liegt in diesem Fall bei etwa 15 dB (4 dB unterhalb des Fenstermaximums). Basierend auf diesen Ergebnissen wird die Hypothese, dass Rauschen in IRs bei der betrachteten Vorgehensweise vor deren zeitlicher Begrenzung hörbar wird, als bestätigt erachtet. Damit verbleibt das additive Rauschen als bestimmender Faktor für die Hörbarkeitsschwelle der betrachteten Abweichungen akustischer IRs. Die Ruhehörschwelle des additiven Rauschens gibt also den Pegel $L_{N,MAX}$ an, den das additive Rauschen nicht überschreiten darf, damit Abweichungen der verwendeten zur idealen IR unhörbar bleiben. Die obere Grenze des verwendbaren Dynamikbereichs DYN der betrachteten IR ergibt sich dann als $DYN = L_{N,MAX} + SNR$. Da sich die Spektralverteilungen der additiven Rauschen von mit verschiedenen Verfahren oder unter unterschiedlichen Bedingungen aufgenommenen IRs unterscheiden, ist zu erwarten, dass sich auch die entsprechenden Hörschwellen unterscheiden. Da dies bedeuten würde, dass zur Prognose des nutzbaren Dynamikbereichs für jede gemessene IR die Hörschwelle des Rauschens bestimmt werden muss, wurden die Ruhehörschwellen der additiven Rauschen verschiedener IR-Messungen im Hörversuch ermittelt. Die IRs wurden in drei verschiedenen Räumen unter Verwendung exponentieller Sinussweeps (ESS, vgl. Müller und Massarani 2001), in einem Raum mit ESS und Maximum-Length-Sequences (MLS, vgl. Rife und Vanderkooy 1989), aufgenommen. Die für das 3-AFC Verfahren notwendigen Intervalle wurden visuell über

einen Lichtpunkt vorgegeben, die Testschalldauer betrug 0.1 s. Am Versuch nahmen 14 Probanden im mittleren Alter von 26 Jahren teil, die mittlere Dauer ergab sich zu 8.9 Minuten. In Abb. 2 sind die Ergebnisse dieses Experiments als Kreise dargestellt.

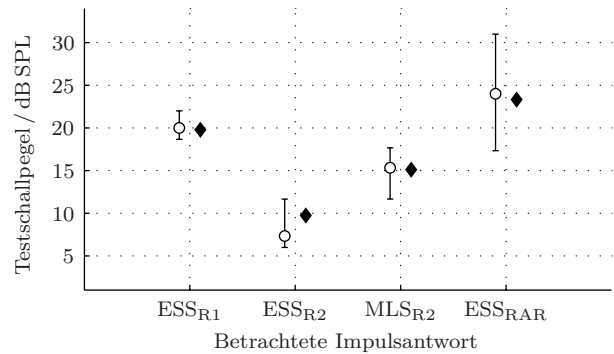


Abbildung 2: Ruhehörschwelle des Rauschens in Impulsantworten. Aufnahme: ESS - exponentieller Sinus-Sweep, MLS - Maximum Length Sequences. Räume: R1 - Seminarsaal, R2 - Laborraum, RAR - reflexionsarm (mit jeweils unterschiedlichen Aufbauten). Gefüllte Rhomben zeigen Modellprognosen.

Die Ergebnisse zeigen, dass sich die Hörschwellen der betrachteten Rauschsignale im Mittel um bis zu 20 dB unterscheiden. Um nicht für jede Messung die Schwellen im Hörversuch ermitteln zu müssen, wird im Folgenden ein Algorithmus zur Prognose der Schwellen und damit des nutzbaren Dynamikbereichs entwickelt und verifiziert.

Prognose des nutzbaren Dynamikbereichs

Grundlage des Verfahrens ist die Annahme, dass ein Rauschsignal hörbar wird, sobald sein RMS-Pegel L_G in einer Frequenzgruppe (FG, vgl. Fastl und Zwicker 2007) die entsprechende Ruhehörschwelle (RHS) übersteigt. Diese ergibt sich aus der RHS für Töne nach Berücksichtigung der unterschiedlichen Crest-Faktoren der Signale. Zur Ermittlung des nutzbaren Dynamikbereichs einer IR sind dann folgende Berechnungen notwendig:

1. $L_{IR,MAX}$, L_N und $SNR = L_{IR,MAX} - L_N$
2. $L_{G,N}$ in allen Frequenzgruppen
3. höchster $L_{N,MAX,model}$ bei dem alle $L_{G,N}$ unter der um 7 dB korrigierten RHS für Töne liegen
4. Dynamikbereich $DYN_{model} = L_{N,MAX,model} + SNR$

Wendet man das Verfahren auf die in Abb. 2 betrachteten IRs an, ergeben sich die als Rhomben gezeigten Prognosen.

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft (DFG) im Projekt FA 140/4 gefördert.

Literatur

- Fastl H., E. Zwicker: *Psychoacoustics - Facts and Models*. 3. Auflage (Springer, Berlin Heidelberg, 2007)
- Müller S., P. Massarani: Transfer-Function Measurement with Sweeps. *J. Audio Eng. Soc.* **49**, 443–471 (2001)
- Rife D. D., J. Vanderkooy: Transfer-Function Measurement with MLS. *J. Audio Eng. Soc.* **37**, 419–444 (1989)