

Modellierung der psychoakustischen Rauigkeit

Roland Sottek¹, Julian Becker²

¹ HEAD acoustics GmbH, 52134 Herzogenrath, E-Mail: roland.sottek@head-acoustics.de

² HEAD acoustics GmbH, 52134 Herzogenrath, E-Mail: julian.becker@head-acoustics.de

Einleitung

Die Rauigkeitsempfindung eines akustischen Signals steht seit Jahren im Mittelpunkt des Sounddesigns. In der Praxis entstehen Rauigkeiten häufig bei rotierenden Bauteilen (Motoren, Getriebe, Lüfter). Ein raues Geräusch kann sowohl zu einem sportlichen Charakter eines Motors beitragen als auch zu einem sehr lästigen Störfaktor werden. Die Herausforderungen im Bereich der Klanggestaltung nehmen ständig zu. Es wird ein Klang angestrebt, der zum Produkt passt. Dieses Ziel soll möglichst im frühen Entwicklungsprozess erreichbar sein. Dazu müssen perzeptive Bewertungen in Kombination mit Modellrechnungen und Simulationswerkzeugen herangezogen werden. Existierende Berechnungsmodelle der Rauigkeit eignen sich gut für synthetische Signale wie modulierte Töne oder Rauschsignale. Die Vorhersage der Rauigkeit technischer Schalle ist jedoch wegen der komplexeren spektralen und zeitlichen Muster deutlich anspruchsvoller.

Obwohl Rauigkeiten in der Praxis sehr häufig auftreten, existiert bisher kein standardisiertes Verfahren zur Berechnung der Rauigkeit. In diesem Beitrag wird ein solches Verfahren beschrieben, das auf einem Modell des menschlichen Gehörs nach Sottek [1] basiert. Es wurde hinsichtlich der nichtlinearen Verarbeitung und der Gewichtung der Modulationsspektren optimiert. Darüber hinaus wird ein Vorschlag zur Berechnung eines binauralen Einzahlwertes der Rauigkeit gegeben. Dadurch lässt sich die tatsächlich wahrgenommene Rauigkeit sehr gut nachbilden. Eine Standardisierung dieses Verfahrens ist sowohl als DIN-Norm als auch im Rahmen von ECMA-74 geplant.

Beschreibung des Modells

Die Berechnung der psychoakustischen Rauigkeit basiert auf einem Gehörmodell, das die Verarbeitungsschritte des menschlichen Gehörs signaltheoretisch abbildet. Dieses Gehörmodell wird auch in dem bereits in ECMA-74 standardisierten Verfahren zur Berechnung der Tonhaltigkeit [2,3] als Grundlage verwendet. Die Struktur der Rauigkeitsberechnung ist in Abbildung 1 dargestellt. Als Zwischenergebnis wird dabei auch die spezifische Lautheit $N'_z(l)$ nach dem Gehörmodell berechnet. Die Größe z bezeichnet hierbei die Nummer des Bandes der Frequenzgruppe und l den Zeitindex nach der Blockverarbeitung. Weiterhin wird die Einhüllende des Signals ermittelt und daraus, zusammen mit der ermittelten Lautheit, die Rauigkeit berechnet. Im Folgenden werden zunächst die verschiedenen Schritte des Gehörmodells erklärt. Das zeitdiskrete Eingangssignal $p(n)$ enthält dabei die gemessenen Schalldruckwerte. Es wird eine Abtastfrequenz von $f_a = 48$ kHz angenommen.

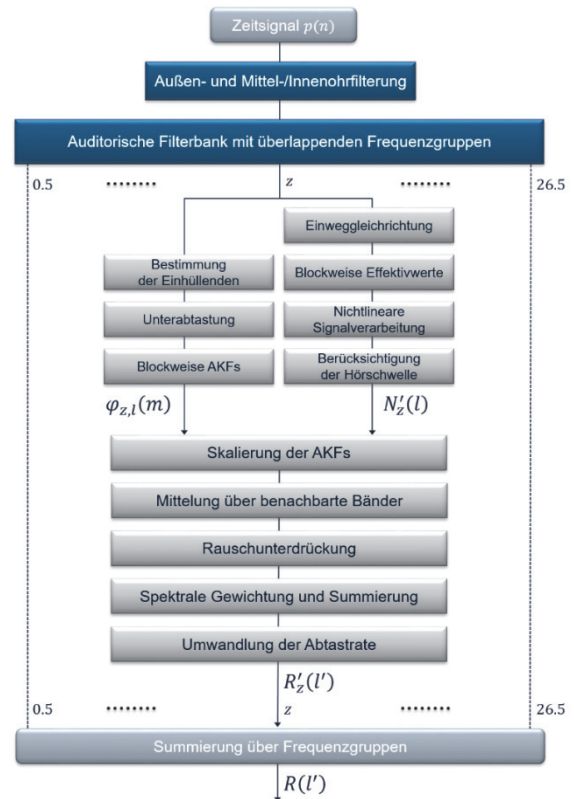


Abbildung 1: Grundlegende Struktur der Rauigkeitsberechnung

Außen- und Mittel-/Innenohrfilterung

In einem ersten Vorverarbeitungsschritt wird das Eingangssignal $p(n)$ entsprechend des Übertragungsverhaltens von Außen- und Mittel-/Innenohr gefiltert (Abbildung 2). Die Übertragungsfunktion des Filters wird so gewählt, dass die Filterung zusammen mit der spezifischen Lautheitsschwelle (Abbildung 3) die Kurven gleicher Lautheit nachbildet.

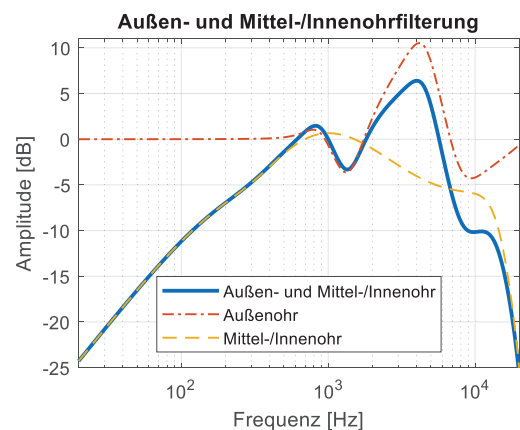


Abbildung 2: Betrag der Übertragungsfunktion des Außen- und Mittel-/Innenohrfilters

Das Filter wurde auf die Kurven gleicher Lautheit entsprechend des Standards ISO 226:2003 optimiert. Die Hörschwellenkurve wurde entsprechend ISO 226:2003 verwendet. Diese Kurve entspricht auch den Daten der Hörschwelle die in ISO 389-7:2005 veröffentlicht wurden.

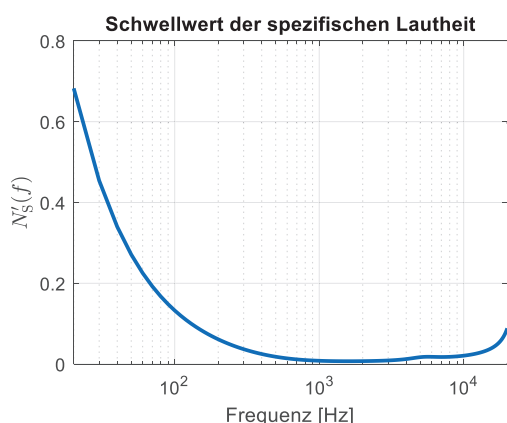


Abbildung 3: Schwellwert der spezifischen Lautheit $N'_s(f)$ zur Berücksichtigung der Hörschwelle

Auditorische Filterbank

Nach der Außen- und Mittel-/Innenohrfilterung wird das Signal in einer auditorischen Filterbank weiterverarbeitet, die die Frequenzgruppen des menschlichen Gehörs nachbildet. Die Form der auditorischen Filter entspricht den Gammaton-Filtern. Die Amplitude wurde so gewählt, dass das Filter bei der Mittenfrequenz $F(z)$ einen Verstärkungsfaktor von 0 dB hat. Die Bandbreite wurde so gewählt, dass sie der äquivalenten Rechteckbandbreite entspricht. Die Mittenfrequenzen und Bandbreiten der Filter werden entsprechend [4; Tabelle II, Formel 2] für Frequenzbänder z , von 0,5 bis 26,5 Bark mit einer Schrittweite von $\Delta z = 0,5$ Bark, berechnet. Diese Funktionen ergeben eine bessere Anpassung an die Bark Tabelle von Zwicker [5] als andere existierende Formeln, wie in [4] im Detail dokumentiert ist.

Das Ergebnis der auditorischen Filterbank sind 53 Bandpasssignale $p_{FG,z}(n)$ mit z von 0,5 bis 26,5 Bark, wodurch eine Erweiterung der Bark-Skala auf 27 Frequenzgruppen mit einer Überlappung von 50 % erreicht wird.

Berechnung der spezifischen Lautheit

Zur Berechnung der Lautheit wird anschließend eine Einweggleichrichtung durchgeführt, da die Hörnerve nur bei Auslenkung der Basilarmembran in eine bestimmte Richtung aktiviert werden.

Für die weitere Verarbeitung müssen blockweise Effektivwerte berechnet werden. Die Blöcke werden mit einer Blockgröße von $s_b = 16384$ Samples und einem Vorschub von $s_v = 4096$ Samples berechnet (75 % zeitliche Überlappung).

Anschließend muss die nichtlineare Beziehung der wahrgenommenen Lautheit zum Schalldruckpegel beachtet werden, die durch die nichtlineare Kompression des Gehörs entsteht. Dieser nichtlineare Zusammenhang zwischen spezifischer Lautheit und Schalldruck wurde für das Gehörmodell auf

Grundlage der Ergebnisse vieler Hörversuche überprüft [6]. Weitere Verbesserungen für höhere Pegel über etwa 80 dB wurden durch eine neue nichtlineare Funktion entsprechend [4, Gleichung 29] erreicht. Diese Gleichung resultiert in einer steileren Steigung für höhere Pegel, in Übereinstimmung mit den Ergebnissen der Experimente von Buus et al. [7] und Epstein et al. [8].

Abschließend wird in jedem Band z der bandabhängige Schwellwert $N'_s(f)$ subtrahiert. Negative Werte werden zu null gesetzt. Dieser Schritt entspricht der Berücksichtigung der Hörschwelle des menschlichen Hörens. Das Ergebnis dieses Verarbeitungsschrittes ist die spezifische Lautheit $N'_z(l)$. Die Gesamtlautheit wird als Summe aller spezifischen Lautheitswerte in Schritten von $\Delta z = 0,5$ Bark berechnet:

$$N(l) = \sum_{z=0,5 \text{ Bark}'}^{26,5 \text{ Bark}'} N'_z(l) \cdot \Delta z. \quad (1)$$

Autokorrelationsfunktion der Einhüllenden

Zur Berechnung der tieffrequenten Modulationshüllkurve aus $p_{FG,z}(n)$ wird die Hilbert-Transformation eingesetzt. Die Modulationshüllkurve wird dabei als Betrag des analytischen Signals angenommen. Die tieffrequenten Hüllkurven werden für eine effizientere Berechnung mit einem Faktor von 16 unterabgetastet (Abtastfrequenz: $f'_a = 3$ kHz).

Die Hüllkurven werden blockweise weiterverarbeitet. Hierbei wird eine Blockgröße von $s'_b = 1024$ und ein Vorschub von $s'_v = 256$ Samples verwendet¹. Für jeden Block wird die Autokorrelationsfunktion (AKF) $\varphi_{z,l}(m)$ berechnet.

Skalierung der Autokorrelationsfunktion

Die Autokorrelationsfunktionen werden zunächst so gewichtet, dass die Energie jedes Blocks (also $\varphi_{z,l}(0)$) der zuvor berechneten spezifischen Lautheit $N'_z(l)$ des jeweiligen Blocks mit einer zusätzlichen Gewichtung unter Verwendung der maximalen spezifischen Lautheit $N'_{max}(l)$ über alle Bänder des entsprechenden Blocks entspricht:

$$\varphi'_{z,l}(m) = \frac{\varphi_{z,l}(m)}{\varphi_{z,l}(0)} \cdot \frac{N'_z(l)}{N'_{max}(l)^{0,45}} \quad (2)$$

mit

$$N'_{max}(l) = \max_z(N'_z(l)). \quad (3)$$

Damit wird beachtet, dass sich der Rauigkeitseindruck nichtlinear mit der Lautheit verändert.

Mittlung der AKFs benachbarter Bänder

Nun werden die AKFs $\varphi'_{z,l}(m)$ benachbarter Bänder gemittelt um Rauscheffekte zu verringern. Die Mittlung wird über 9 Bänder durchgeführt. Das heißt, jedes Band wird mit vier höheren und mit vier niedrigeren Bändern gemittelt. Für Bänder, die nur weniger als vier Nachbarbänder besitzen, wird die Mittlung symmetrisch über eine entsprechend reduzierte Anzahl von Bändern durchgeführt. Für das zweite Band wird also beispielsweise über drei Bänder gemittelt (ein niedrigeres und ein höheres Band). Das Ergebnis ist $\bar{\varphi}_{z,l}(m)$.

das nicht unterabgestastete Signal angewendeten Blockgrößen von $s_b = 16384$ und $s_v = 4096$ Samples.

¹ Aufgrund der Unterabtastung mit dem Faktor 16 entsprechen diese Blockgrößen s'_b und s'_v den zur Berechnung der Effektivwerte auf

Rauschunterdrückung / Verringerung des Leck-Effekts

Um weitere Rauschanteile zu unterdrücken werden Anteile der Autokorrelationsfunktion bei niedrigen Verschiebungszeiten abgeschwächt, da in diesem Bereich ein Großteil der Rauschenergie vorhanden ist. Hierzu wird zunächst ein Gewichtungsfenster für $m < m_N$ bestimmt. Der Zeitindex m_N entspricht einer Zeit von 8 ms.

$$w_{R,z,l}(m) = E_{p,z,l} + 0.5 \cdot (1 - E_{p,z,l}) \cdot \left(1 - \cos\left(\pi \frac{m}{m_N}\right)\right) \quad (4)$$

Wobei die Energie $E_{p,z,l}$ des periodischen Teils des Signals geschätzt wird als

$$E_{p,z,l} = \frac{\max_{m_N < m < \frac{s'_b}{2}} (\bar{\varphi}_{z,l}(m))}{\bar{\varphi}_{z,l}(0)} \quad (5)$$

Für $m \geq m_N$ gilt $w_{R,z,l}(m) = 1$. Um den Leck-Effekt durch die Blockverarbeitung zu verringern wird zusätzlich ein Fenster, ähnlich dem von-Hann-Fenster, definiert:

$$w_L(m) = 0.5 \cdot \left(1 + \cos\left(\pi \frac{m}{s'_b}\right)\right) \quad (6)$$

Die Rauschunterdrückung und Verringerung des Leck-Effekts erfolgt durch Multiplikation mit den Gewichtungsfenstern:

$$\bar{\varphi}_{z,l}(m) = w_{R,z,l}(m) \cdot w_L(m) \cdot \bar{\varphi}_{z,l}(m) \quad (7)$$

Spektrale Gewichtung

Zunächst wird eine diskrete Fourier-Transformation von $\bar{\varphi}_{z,l}(m)$ durchgeführt. Es werden nur Verschiebungszeiten bis zur halben Blocklänge $m = s'_b/2$ beachtet, da die AKF für höhere Verschiebungszeiten nicht zuverlässig ist. Höhere Indizes werden durch eine symmetrische Ergänzung ersetzt, sodass eine FFT durchgeführt werden kann, die in einem reellwertigen Ergebnis resultiert. Das daraus resultierende Spektrum wird anschließend mit einer bandabhängigen Gewichtungsfunktion $G_z(k)$ multipliziert:

$$G_z(k) = \frac{r_{\max}(z)}{\left(1 + \left(\left(\frac{f'_a \cdot k}{f_{\max}(z)s'_b} - \frac{f_{\max}(z)s'_b}{f'_a \cdot k}\right) \cdot q_1(z)\right)^2\right)^{q_2(z)}} \quad (8)$$

Dabei ist r_{\max} der bandabhängige maximale Gewichtungswert, der bei der Frequenz f_{\max} auftritt. Diese beiden Werte berechnen sich als

$$r_{\max}(z) = 1 + \left(r_1(z) + r_2(z) \log_2\left(\frac{F(z)}{1 \text{ kHz}}\right)\right) \log_2\left(\frac{F(z)}{1 \text{ kHz}}\right) \quad (9)$$

und

$$\frac{f_{\max}(z)}{\text{Hz}} = 70 + \left(f_1(z) + f_2(z) \log_2\left(\frac{F(z)}{1 \text{ kHz}}\right)\right) \log_2\left(\frac{F(z)}{1 \text{ kHz}}\right) \quad (10)$$

mit

	$F(z) < 1 \text{ kHz}$	$F(z) \geq 1 \text{ kHz}$
$r_1(z)$	0.4325	-0.5165
$r_2(z)$	0.1304	0.1025
$f_1(z)$	16.1711	0.0000
$f_2(z)$	1.4809	0.0000

Die Parameter $q_1(z)$ und $q_2(z)$ werden gewählt zu

$$q_1(z) = \begin{cases} 0.9132, & f'_a \cdot k < f_{\max}(z)s'_b \\ 0.8503, & f'_a \cdot k \geq f_{\max}(z)s'_b \end{cases} \quad (11)$$

$$q_2(z) = \begin{cases} 0.7585, & f'_a \cdot k < f_{\max}(z)s'_b \\ 0.5478, & f'_a \cdot k \geq f_{\max}(z)s'_b \end{cases} \quad (12)$$

Das resultierende Spektrum wird im Folgenden mit $\Phi_{z,l}(k)$ bezeichnet.

Berechnung der Rauigkeit

Zunächst wird für jeden Zeitblock jedes Bandes die Summe über den Frequenzindex k gebildet und mit einem Kalibrierfaktor $c_R = 0.00152089$ gewichtet. So ergibt sich die zeitabhängige spezifische Rauigkeit $R'_z(l)$:

$$R'_z(l) = c_R \sum_k (\Phi_{z,l}(k)) \quad (13)$$

Somit wird für das Kalibriersignal (Cosinusförmig modulierter Sinuston mit Trägerfrequenz 1 kHz, Modulationsfrequenz 70 Hz, 60 dB SPL und Modulationsgrad 1) eine Rauigkeit von 1 asper berechnet. Anschließend findet eine Interpolation der Abtastwerte auf eine Abtastrate von 50 Hz statt. Für die Interpolation wird ein stückweise kubisch hermitescher Spline verwendet. Daraus resultiert eine zeitliche Auflösung von 20 ms. Der neue Zeitindex wird mit l' bezeichnet.

Aus den spezifischen Rauigkeiten $R'_z(l')$ wird nun ein Einzahlwert $R(l')$ für jeden Zeitindex l' sowie die mittlere spezifische Rauigkeit $R'(z)$ berechnet. Zur Berechnung von $R(l')$ wird für jeden Zeitindex die Summe über alle Bänder z unter Berücksichtigung der Überlappung der Bänder gebildet:

$$R(l') = \sum_z (R'_z(l') \cdot \Delta z) \quad (14)$$

Die Berechnung der spezifischen Rauigkeit $R'(z)$ erfolgt als Mittelwert von $R'_z(l')$ über den Zeitindex l' für jedes Band z . Hierbei werden alle Zeitblöcke, die zu einer Zeit kleiner 0,3 s korrespondieren ignoriert, da diese Zeit als Dauer des Einschwingvorgangs betrachtet wird.

Um einen Einzahlwert für das gesamte Signal zu erhalten wird anschließend das 90. Perzentil über alle Zeitblöcke l' berechnet. Auch hier werden alle Zeitblöcke, die zu einer Zeit kleiner als 0,3 s korrespondieren, ignoriert.

Berechnung von Einzahlwerten für binaurale Signale

Bei binauralen Signalen werden monaurale Rauigkeitswerte R_L und R_R des linken und rechten Kanals entsprechend des beschriebenen Verfahrens für jeden Kanal getrennt berechnet (unter der Annahme diotischer Signale). Anschließend wird ein binauraler Rauigkeitswert R als quadratischer Mittelwert der Rauigkeitswerte für die beiden Kanäle gebildet:

$$R = \sqrt{\frac{R_L^2 + R_R^2}{2}} \quad (15)$$

Gleichung 15 entspricht näherungsweise der Formel für die binaurale Inhibition aus dem binauralen Lautheitsmodell von Moore/Glasberg (ISO 532-2, siehe auch [9]).

Für den Fall, dass der Rauigkeitswert eines Kanals vernachlässigbar ist, ergibt sich nach Gleichung 15, verglichen mit der diotischen Darbietung, eine um den Faktor $\sqrt{0.5}$ geringere Rauigkeit.

Ergebnisse

Zur Evaluierung der psychoakustischen Rauigkeit werden die berechneten Rauigkeitswerte mit in psychoakustischen Experimenten ermittelten Werten der Rauigkeit verglichen.

Zunächst werden die Rauigkeiten amplitudenmodulierter Sinustöne verschiedener Modulationsfrequenzen betrachtet. Die Ergebnisse des psychoakustischen Berechnungsverfahrens werden dabei mit den von Fastl und Zwicker [5] veröffentlichten Daten verglichen. In Abbildung 4 ist zu erkennen, dass die berechneten Werte den veröffentlichten Kurven sehr ähnlich sind.

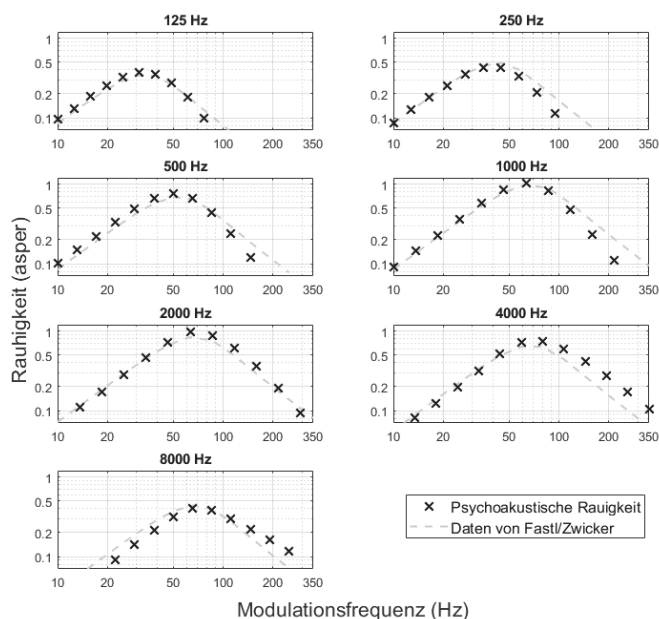


Abbildung 4: Berechnete Rauigkeiten amplitudenmodulierter Sinustöne für verschiedene Trägerfrequenzen und verschiedene Modulationsfrequenzen im Vergleich zu den von Fastl und Zwicker in [5] (Abbildung 11.2) veröffentlichten Daten

Um die Anwendbarkeit des Verfahrens auf technische Geräusche zu untersuchen wurden Hörversuche mit den in Tabelle 1 beschriebenen technischen Geräuschen durchgeführt.

Tabelle 1: Technische Geräusche

Abkürzung	Beschreibung: englisch	Beschreibung: deutsch
ESA_02	Electrical Seat Adjuster	Elektrischer Sitzversteller
ETB_01	Electrical Toothbrush	Elektrische Zahnbürste
GEN_02	Generator	Generator
HDD_07	Hard Disk Drive	Festplatte
HDD_09	Hard Disk Drive	Festplatte
SCOOT	Pass-by of Scooter	Vorbeifahrt Motorroller
SINUS	Calibration Signal: Modulated Sinus Tone	Kalibriersignal: Modulierter Sinuston
TOF_03	Take Of (Airplane)	Start (Flugzeug)

Anschließend wurden Rauigkeitswerte berechnet. Die Ergebnisse sind in Abbildung 5 zusammen mit den Hörversuchsergebnissen (Mittelwerte und 95% Konfidenzintervalle) dargestellt. Es ist zu erkennen, dass alle berechneten Ergebnisse im 95% Konfidenzintervall der Hörversuchsdaten liegen.

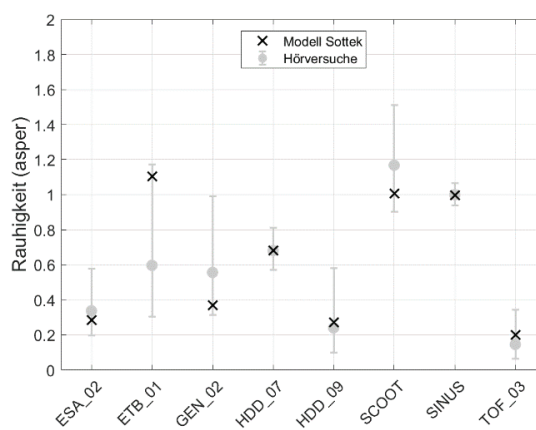


Abbildung 5: Ergebnisse für technische Geräusche

Fazit

In diesem Beitrag wurde ein psychoakustisches Verfahren zur Berechnung der Rauigkeit vorgestellt. Da diese Berechnungsmethode auf einem Modell des menschlichen Gehörs basiert werden viele psychoakustische Effekte beachtet und die menschliche Wahrnehmung kann sehr gut nachgebildet werden. Das Verfahren befindet sich zurzeit im Prozess für eine Standardisierung im Rahmen des DIN. Weiterhin ist eine Standardisierung im internationalen Standard ECMA-74 geplant.

Literatur

- [1] Sottek, R.: Modelle zur Signalverarbeitung im menschlichen Gehör, Dissertation, RWTH Aachen, 1993.
- [2] ECMA-74 15th Edition/July 2018, Annex F and Annex G, ECMA international, Rue du Rhône 114, CH-1204 Geneva, Switzerland, 2018,
- [3] Becker, J., Sottek, R.: Psychoacoustic Tonality Analysis, Proc. Inter-Noise 2018, Chicago, 2018.
- [4] Sottek, R.: A Hearing Model Approach to Time-Varying Loudness, Acta Acustica united with Acustica, vol. 102, no. 4, pp. 725-744, 2016.
- [5] Fastl, H., Zwicker, E.: Psychoacoustics. Facts and models. Springer Verlag, Berlin, Heidelberg, New York, 2006.
- [6] Bierbaums, T., Sottek, R.: Modellierung der zeitvarianten Lautheit mit einem Gehörmodell, DAGA 2012, Darmstadt, 591-592, 2012.
- [7] Buus, S., Florentine, M.: Modifications to the power function for loudness. – In: Fechner Day 2001. Proceedings of the 17th Annual Meeting of the International Society for Psychophysics. E. Summerfield, R. Kompuss, T. Lachmann (eds.). Pabst, Berlin, 236–241, 2001.
- [8] Epstein, M., Florentine, M.: A test of the equal-loudnessratio hypothesis using cross-modality matching functions. J. Acoust. Soc. Am. 118, 907–913, 2005.
- [9] Moore, B. C., Glasberg, B. R., Varathanathan, A., Schlittenlacher, J.: A loudness model for time-varying sounds incorporating binaural inhibition. Trends in hearing, 20, 2331216516682698, 2016.