

Ein Ton macht noch keine Harmonie -

Der Einfluss von Tonhaltigkeit und Dissonanz auf Präferenzurteile

Anna Rieger^{1,2}, Steven van de Par², Hans-Peter Rabl¹, Arne Oetjen

¹ OTH Regensburg, Labor für Entwicklungsanwendungen, 93053 Regensburg, Deutschland,

Email: anna.rieger@st.othr.de, anna.rieger@uol.de

² C/o Universität Oldenburg, Abt. Akustik, Cluster of Excellence "Hearing4All", 26129 Oldenburg, Deutschland,

Email: steven.van.de.par@uol.de

Einleitung

Die zuverlässige Auswertung und Optimierung der Geräuschqualität ist wesentlicher Bestandteil der Entwicklungsarbeiten an verschiedenen technischen Produkten. Am Beispiel elektrifizierter Fahrzeuge wird deutlich, wie sich NVH-Anforderungen kontinuierlich ändern. Sowohl Antriebskomponenten wie Elektromotoren, Inverter und Getriebe als auch künstlich erzeugte Geräusche wie beispielsweise AVAS (Acoustic Vehicle Alerting System) verleihen den Fahrzeugen in Summe ein sehr tonhaftes Spektrum, bei dem Ordnungen verschiedener Bauteile in gewissen Frequenzverhältnissen zueinander stehen können [1].

Zwicker und Fastl beschreiben die Lästigkeit von Geräuschen als Größe, die von der Lautheit, der Klangfarbe und zeitlichen Struktur abhängt [2]. Aus diversen Studien geht hervor, dass das Vorhandensein von Tönen die Lästigkeit steigert. Durch vorhergehende Arbeiten der Autoren konnte außerdem der Einfluss von Frequenzintervallen zwischen Tönen (Konsonanz- und Dissonanz) auf die Geräuschqualität nachgewiesen werden [3, 4]. Neben der Tonhaltigkeit als Maß für die Prominenz von Einzeltönen erlaubt die Konsonanz bzw. Dissonanz die Bewertung der Güte eines Klangs.

Aufbauend auf diesen Untersuchungen zeigt dieser Beitrag den Einfluss von Schärfe, Tonhaltigkeit und Dissonanz auf Präferenzurteile. Dazu werden ausschnittsweise Daten aus zwei unterschiedlichen Hörexperimenten mit zwei unterschiedlichen Versuchskollektiven gezeigt und diskutiert. Ein Ansatz für die Modellierung von Präferenzurteilen basierend auf Subjektivgrößen wird vorgestellt, sowie ein Ausblick auf eine rein modellbasierte Modellierung gegeben.

Methodik und Experimente

In [3, 4] wurde ein Hörexperiment beschrieben, wodurch ein psychoakustisches Konzept der Dissonanz entwickelt und der Einfluss der Dissonanz auf Präferenzurteile nachgewiesen werden konnte. Alle Geräusche wiesen eine ähnliche Tonhaltigkeit auf, weshalb diese bei der Präferenzvorhersage vernachlässigt werden konnte. Die Wechselwirkung aus Tonhaltigkeit und Dissonanz ist von spezieller Natur, da für die Existenz von Konsonanz bzw. Dissonanz zwingend ein tonhaltiges Geräusch benötigt wird. In Bezug auf die Metrikentwicklung für reale Geräusche sollte demnach auch diese Wechselwirkung untersucht werden.

Versuch 1-2: Stimuli

In zunächst synthetisch generierten Stimuli wurde der komplexe Klangcharakter eines bspw. Elektrofahrzeuges etwas vereinfacht nachempfunden. Grundsätzlich wird zwischen Test- und Referenzgeräuschen unterschieden, wobei für alle $f_0 = 1046.5$ Hz gilt. Jedes Testgeräusch enthielt zwei harmonische Tonkomplexe ($f_0, 2 \cdot f_0, 3 \cdot f_0$), die in einem bestimmten Frequenzintervall $x = 2^{n/12}$ der gleichstufigen Stimmung zueinander standen, sowie ein bandpassgefiltertes Hintergrundrauschen besaßen. Daneben gab es vier Referenzgeräusche, alle davon mit demselben Hintergrundrauschen wie beschrieben, wovon zwei Sinuskonditionen mit f_0 und verschiedenem SNR (Ref. 1 niedriger, Ref. 2 hoher SNR), eines ein harmonischer Tonkomplex mit $f_0, 2 \cdot f_0, 3 \cdot f_0$ (Ref. 3) und eines ein konsonanter Zweiklang (Ref. 4) war (aufgebaut wie die Testgeräusche). Als SNR ist für die Geräusche das Verhältnis von Toninformation zum Hintergrundrauschen spezifiziert, wobei dieser für alle Geräusche mit Intervallen bei ungefähr 16 dB lag und diese daher eine ausgeprägte Tonhaltigkeit besaßen, was ebenso für die Referenzen 2, 3 galt.

Als Frequenzintervalle x wurden die aus der Musik bekannten Intervalle bzw. Halbtonschritte $n = \pm 10, \pm 6, \pm 5, \pm 1$ und $+4$ in den Geräuschen angewandt. Was zur Beschreibung der Intervalle als \pm Septime, \pm Tritonus, \pm Quarte, \pm Sekunde für Test- und $+Terz$ für das Referenzgeräusch 4 führt. Als Beispiel zeigt \pm Septime, dass die Intervalltöne einmal zu Frequenzen $< f_0$ und einmal $> f_0$ verschoben waren, was zusätzlich einen Einfluss der Schärfe in den Geräuschen erzeugte.

Alle Stimuli wurden zu Beginn und während der Versuche, wenn Änderung von Signaleigenschaften erfolgten, iterativ auf gleiche Lautheit nach DIN45631 [5] zu Referenzstimulus 4 (65 dB SPL) gebracht. Der Einfluss der Lautheit kann in diesen Untersuchungen daher vernachlässigt werden. Durch dieses Versuchsdesign wird es demnach möglich den Einfluss weiterer psychoakustischer Dimensionen, wie Schärfe, Tonhaltigkeit und Dissonanz auf Präferenzurteile genau zu untersuchen.

Versuch 1: PSE / Kollektiv 1

In einem adaptiven, automatisierten 2-Intervall Alternative Forced Choice Verfahren¹ wurde eine erste Gruppe aus $N = 26$ Proband:innen nach ihrer

¹Toolbox nach [6] <https://medi.uni-oldenburg.de/afc/>

Präferenz bei Geräuschvergleichen befragt. Dabei wurden Testgeräusche im Vergleich zu Referenzgeräuschen anhand der Frage *Which interval would you prefer as a vehicle sound?* präsentiert. Bei der Entscheidung für eines der Geräusche wurden beim anderen Geräusch im Programmhintergrund der SNR schrittweise reduziert. Mit jeder Entscheidung wurden somit die Tonpegelüberschüsse aller tonaler Komponenten relativ zum Hintergrundrauschen verändert, welche maßgeblich die empfundene Tonhaltigkeit von Geräuschen beeinflussen [7, 8]. Die Lautheit wurde iterativ für jedes Geräusch auf den Referenzwert gebracht. Das Ergebnis für einen der Geräuschvergleiche ist der sog. *Point of Subjective Equality* (PSE), hier also eine benötigte Änderung des SNR eines Geräusches, um gleich beurteilte Präferenz mit dem jeweiligen Vergleichsgeräusch zu erreichen.

Versuch 2: Kategorial / Kollektiv 2

Ein Kollektiv 2 bestehend aus $N = 24$ Proband:innen wurde aufgefordert alle Versuchsstimuli (Test und Referenzen) mit verschiedenen diskreten Einstellungen der jeweiligen SNR-Verringerung auf teilweise gelabelten Kategorialskalen bezüglich der empfundenen Schärfe, Tonhaltigkeit und Dissonanz zu bewerten. Jede Skala reichte dabei von 1 *nicht scharf, tonhaltig, dissonant* bis 9 *sehr scharf, tonhaltig, dissonant*.

Beide Kollektive wurden jeweils nicht über die Aufgabe des anderen Kollektives informiert.

Ergebnisse PSE: Präferenzurteile

In Abbildung 1 ist der, für jeweils gleiche Präferenz mit dem jeweiligen Vergleichsgeräusch, ermittelte SNR über das Intervall des Testgeräusches aufgetragen. Positiv aufgetragene Werte bedeuten, dass der SNR im Testgeräusch verringert werden musste. Negativ aufgetragene Werte sprechen folglich für die Verringerung des SNR im Referenzgeräusch. Für alle Testgeräusche war die Wahrnehmungsschwelle der tonalen Komponenten bei ungefähr 30 dB Verringerung im SNR (vgl. Maximum Ordinate).

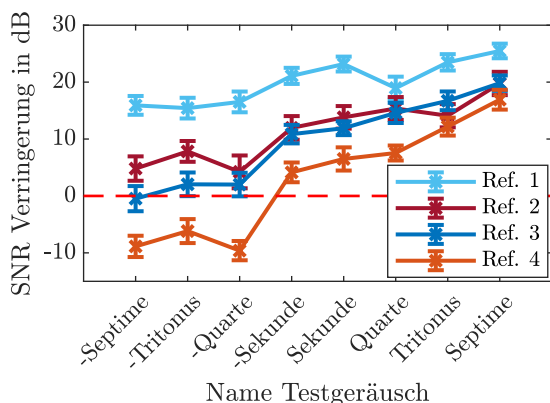


Abbildung 1: Ergebnisse der PSE-Messung als Mittelwerte und Standardfehler für $N = 26$ Proband:innen, aufgetragen als Verringerung im SNR in dB über das jeweilige Testgeräusch.

In Bezug auf Referenz 1 mussten die SNRs in allen Testgeräuschen relativ stark reduziert werden, um Punkte gleich bewerteter Präferenz zu erreichen. Für Referenz

2, die äquivalente Sinuskondition wie Referenz 1 nur mit höherem SNR, ist beobachtbar, dass die SNRs der Testgeräusche weniger stark reduziert werden mussten. Hier liese sich direkt der Einfluss der unterschiedlichen Tonhaltigkeit aufgrund der unterschiedlichen SNRs der Referenzgeräusche beobachten. Für den harmonischen Tonkomplex, Referenz 3, ist wenig Änderung im Vergleich zu Referenz 2 beobachtbar, außer für die Konditionen *-Septime, -Tritonus*. Diese Unterschiede könnten durch den Einfluss der Obertöne im harmonischen Tonkomplex im Vergleich zur Sinuskondition erklärt werden. Die Ergebnisse für Referenz 4, der konsonante Zweiklang, zeigen den Einfluss von zusätzlichen Frequenzintervallen zwischen Tönen. Hier gab es einige Testgeräusche *-Septime, -Tritonus, -Quarte* für welche das Referenzgeräusch im SNR reduziert werden musste, um die Punkte gleich bewerteter Präferenz zu erreichen.

Wie auch in vorherigen Ergebnissen [3, 4] sichtbar war, werden Geräusche mit gleichen Frequenzintervallen, z.B. \pm *Septime, \pm Tritonus, \pm Quarte*² sehr unterschiedlich bezüglich der Präferenz bewertet, was vermutlich vor allem an den unterschiedlichen Platzierung der spektralen Komponenten, also der unterschiedlichen Schärfe der Geräusche liegt. Ebenso kann in den Daten wieder der Einfluss der Dissonanz auf Präferenzurteile beobachtet werden, beispielsweise direkt am sprunghaften Anstieg der Kurven von *-Quarte* auf *-Sekunde*.

Da wie beschrieben die Lautheit der Stimuli ausgeglichen war und nach dieser Analyse die hauptsächlichsten psychoakustischen Einflussfaktoren für die Geräusche bei Schärfe, Tonhaltigkeit und Dissonanz liegen, könnte ein Präferenzindikator nach folgender, relativ simplen Annahme gebildet werden:

$$P - \text{Indikator} \sim S^a + b \cdot T^c + d \cdot D^e \quad (1)$$

Zur Anwendung bei der Bewertung der Geräuschqualität sollten die Terme S, T und D durch die Berechnung von Algorithmen zu Schärfe, Tonhaltigkeit und Dissonanz, vergleiche z.B. [3, 9, 10], substituiert werden. Für die hier gezeigten Untersuchungen sollten jedoch zunächst Subjektivdaten zu Schärfe, Tonhaltigkeit und Dissonanz gewonnen werden, um daraufhin zuverlässig die Performance des Präferenz-Indikators sowie der psychoakustischen Modelle zu bewerten.

Ergebnisse: Kategorialurteile

Abbildung 2 zeigt die mittleren Kategorialbeurteilungen mit Standardfehlern von Kollektiv 2 ($N = 24$ Proband:innen) für von oben nach unten: Schärfe, Tonhaltigkeit und Dissonanz für alle Testgeräusche.³ Die maximalen Standardfehler sind in der Größenordnung einer Kategorie. Dadurch, dass alle Test- und Referenzgeräusche mit unterschiedlichen diskreten Verringerungen des SNR (von 0 bis zu 30 dB) bewertet wurden, gibt es für jede psychoakustische Dimension eine Art Kennfeld.

Für die Schärfe kann beobachtet werden, dass sich bei einer Verringerung des SNR grundsätzlich die Schärfe ver-

²beachte, dass bspw. für Septimen gilt: $(2^{-10/12})^{-1} = 2^{10/12}$

³Nur ein Teil der Daten ist hier gezeigt.

ringert. Der charakteristische Anstieg der Schärfe liegt im Unterschied der spektralen Komponenten in den Geräuschen. Dabei hat die *-Septime* insgesamt die niedrigsten spektralen Komponenten, wohingegen die *+Septime* die insgesamt Höchsten hat, vgl. die beschriebenen Intervalle in *Methodik und Experimente*. Die Tonhaltigkeit nimmt wie erwartet mit einer Verringerung des SNR ab. Grundsätzlich haben Geräusche mit gleichem SNR eine sehr ähnlich empfundene Tonhaltigkeit. Im Kennfeld der Dissonanz ist zu sehen, dass diese maßgeblich vom Frequenzintervall in den Geräuschen abhängt. Ebenso wie in anderen Studien kann grundsätzlich die allgemein aus der Musik bekannte Intervalltheorie beobachtet werden.

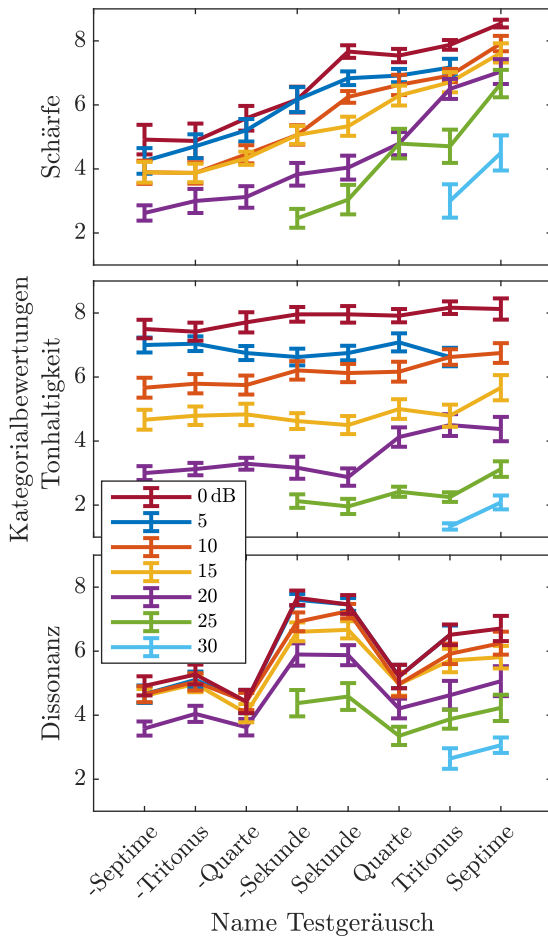


Abbildung 2: Ergebnisse des Kategorialverfahrens (1-9) als Mittelwerte und Standardfehler für $N = 24$ Proband:innen. Von oben nach unten sind die Dimensionen Schärfe, Tonhaltigkeit und Dissonanz über das jeweilige Testgeräusch aufgetragen.

In-the-loop Subjektiv-Modellierung

Mit den Daten aus den Kennfeldern (Gruppe 2) für Schärfe, Tonhaltigkeit und Dissonanz kann nun versucht werden die Präferenzurteile (Gruppe 1) mit Anwendung der Vorschrift aus Gleichung 1 nachzubilden.

Um die Modellperformance abzutesten wird der Präferenz-Indikator *in-the-loop* in den PSE Hörversuch eingebettet. Statt einem Versuchsteilnehmer, der in der Programmoberfläche die Geräuschvergleiche bewertet, wird der numerische Output des Präferenzindikators zur

Bewertung der Präferenz der Geräusche herangezogen. Abbildung 3 zeigt den Präferenzindikator für die Test- und Referenzgeräusche als Balken, woraus auch der Anteil der Schärfe, Tonhaltigkeit und Dissonanz am Gesamturteil ersichtlich wird.

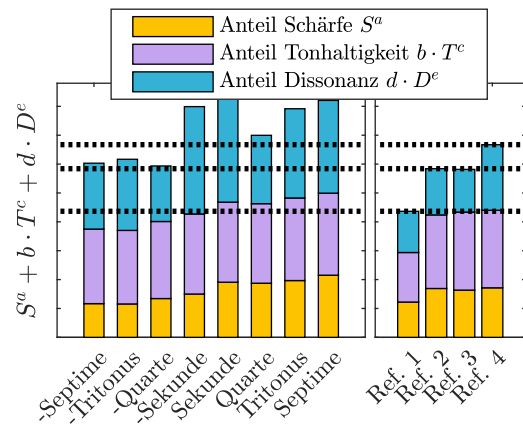


Abbildung 3: Präferenz-Indikator für Test- und Referenzgeräusche in der Kondition mit 0 dB SNR Verringerung.

Nach jeder Bewertung eines Geräuschvergleiches durch den Präferenz-Indikator, wird in einem der Geräusche der SNR reduziert, wobei für jede Einstellung des SNR ein Wert für Schärfe, Tonhaltigkeit und Dissonanz aus den Kennfeldern interpoliert werden kann. Die kleinste zu bewertende Änderung im SNR der Geräusche war 0,125 dB, was die Vorhersage im Vergleich zur realen Versuchsperson sehr genau macht. Die Abbildung zeigt für alle Geräusche die Kondition mit 0 dB Verringerung im SNR. Die gepunkteten Linien zeigen die Unterschiede im Präferenz-Indikator der Testgeräusche (links) zu Referenzgeräuschen (rechts) auf. Niedrigere Werte im Präferenz-Indikator bedeuten, dass dieser Stimulus präferiert werden würde und der SNR im anderen Vergleichsgeräusch reduziert werden müsste. Als Beispiel würden die Referenzen 1-3 gegenüber dem *-Tritonus* präferiert und daher der SNR im Testgeräusch *-Tritonus* reduziert. Gegenüber der *Referenz 4* würde hingegen der *-Tritonus* präferiert, weshalb der SNR in Referenz 4 reduziert werden müsste bis die Punkte gleich bewerteter Präferenz durch den Präferenz-Indikator (gleicher numerischer Output) erreicht werden.

Abbildung 4 zeigt die Ergebnisse des *Modell-in-the-loop* PSE-Versuches. Wie in Abbildung 1 gezeigt, entsprechen die durch Linien verbundenen Punkte den im Mittel nötigen Verringerungen im SNR, um Punkte gleich empfundener Präferenz zweier Geräusche zu finden (Kollektiv 1). Für alle Geräuschvergleiche sind im selben Farbcodex und Kreisen die Ergebnisse der Präferenzvorhersage mit den mittleren Kategorialurteilen von Kollektiv 2 (vgl. Abbildung 2) nach rechnerischer Ermittlung eines Präferenz-Indikators aus kategorial beurteilten Schärfe, Tonhaltigkeit und Dissonanz nach Gleichung 1 aufgetragen. Es kann beobachtet werden, dass nahezu alle Geräuschvergleiche, bzw. alle nötigen Verringerungen im SNR, die durch Kollektiv 1 bewertet wurden, vom Präferenz-Indikator, der aus dem Mit-

tel der Kategorialurteile von Kollektiv 2 gebildet wurde, innerhalb des Standardfehlers von Kollektiv 1 vorhergesagt werden können. Daraus kann schlussgefolgert werden, dass die allgemeine Präferenz oder Lästigkeit von Geräuschen grundsätzlich sehr gut aus einzelnen, relevanten psychoakustischen Dimensionen gebildet werden kann.

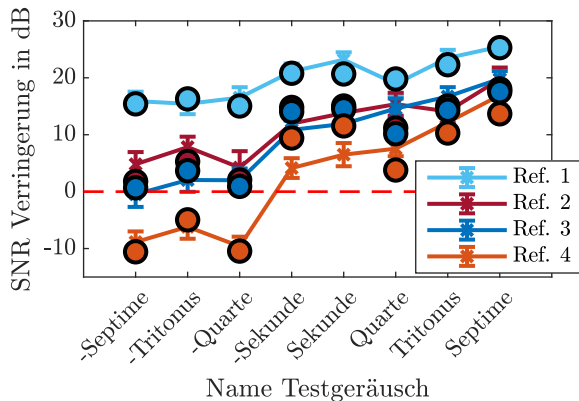


Abbildung 4: Linien: Ergebnisse der PSE-Messung als Mittelwerte und Standardfehler für Kollektiv 1 ($N = 26$), aufgetragen als Verringerung im SNR in dB über das jeweilige Testgeräusch. Kreise: Vorhersage der Präferenzurteile mittels Subjektivdaten von Kollektiv 2 ($N = 24$).

Zusammenfassung und Ausblick

In Anlehnung an die Geräusche von Fahrzeugen mit elektrifiziertem Antriebsstrang wurde der Einfluss der ausgewählten Dimensionen Schärfe, Tonhaltigkeit und Dissonanz auf allgemeine Präferenzurteile untersucht. Dazu wurden Daten aus zwei Hörversuchen gezeigt und diskutiert, die darauf abzielten, experimentell erhobene Präferenzurteile tonhaltiger Stimuli gleicher Lautheit mithilfe dieser ausgewählten Dimensionen zu modellieren. Dabei konnte gezeigt werden, dass ein Präferenz-Indikator nach Gleichung 1 mit Subjektivurteilen zu diesen Dimensionen die bewertete Präferenz von Geräuschen zuverlässig vorhersagen kann, was zusätzlich zu vorhergehenden Experimenten die Theorie stützt, dass die Angenehmheit allgemein negativ beeinflusst wird von Lautheit, Klangfarbe und zeitlicher Struktur eines Geräusches [2].

Die erhobenen Daten ermöglichen nun weitere Studien zur Performance von Algorithmen für die Hörempfindungen Schärfe, Tonhaltigkeit und Dissonanz im Vergleich zu Subjektivdaten, vergleiche beispielsweise [3, 9, 10]. Als Ausblick zeigt Abbildung 5 die Korrelation zweier Modelle (Ordinaten) mit den Kategorialurteilen zu Tonhaltigkeit und Dissonanz (Abszissen). Es ist zu sehen, dass beispielsweise das Tonhaltigkeits-Modell nach Oetjen et al. [10], welches grundsätzlich eine Weiterentwicklung der Tonhaltigkeit nach DIN 45681 [7] für transiente Geräusche beschreibt, bereits 90% der Varianz in den Daten für die Testgeräusche erklären kann. Der Modellansatz für Dissonanz nach Rieger et al. [3], dessen Modelloutput bisher noch unskaliert ist, kann in der aktuellen Version bereits 70% der Varianz in den Daten der Testgeräusche erklären.

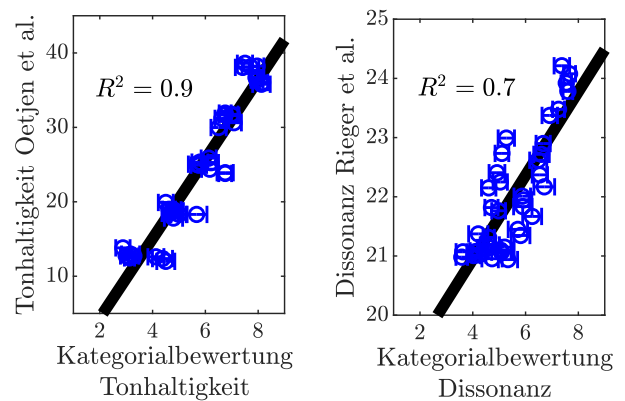


Abbildung 5: Korrelation der Kategorialbewertungen für Tonhaltigkeit und Dissonanz der Testgeräusche mit Modellberechnungen.

Literatur

- [1] Tschöke, H., Gutzmer, P. und Pfund, T. (Hrsg.): *Elektrifizierung des Antriebsstrangs: Grundlagen - vom Mikro-Hybrid zum vollelektrischen Antrieb*. Springer-Verlag Berlin-Heidelberg, 2019.
- [2] Fastl, H. und Zwicker, E. (Hrsg.): *Psychoacoustics: Facts and Models*. Springer-Verlag, Berlin und New York, 3. Ausgabe, 2007.
- [3] Rieger, A., Oetjen A., van de Par, S. und Rabl, H.-P.: Gehörbezogene Modellierung der Dissonanz. *Fortschritte der Akustik - DAGA 2021 Wien*, S. 1034-1037.
- [4] Rieger, A., Oetjen A., van de Par, S. und Rabl, H.-P.: Einfluss der Dissonanz auf die Wahrnehmung der Geräuschqualität. *Fortschritte der Akustik - DAGA 2022 Stuttgart*, S. 963-966.
- [5] Deutsches Institut für Normung e.V.: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum, DIN45631/A1:2010-03.
- [6] Ewert, S.: AFC - A modular framework for running psychoacoustic experiments and computational perception models. *Fortschritte der Akustik - AIA - DAGA 2013*, S. 1326-1329.
- [7] Deutsches Institut für Normung e.V.: Bestimmung der Tonhaltigkeit von Geräuschen und Ermittlung eines Tonzuschlages für die Beurteilung von Geräuschmissionen, DIN45681:2006-08.
- [8] Vormann, M.: *Untersuchungen zu psychoakustischen Mess- und Berechnungsverfahren der Tonhaltigkeit*. Dissertation, Oldenburg, 2011.
- [9] Deutsches Institut für Normung e.V.: Messtechnische Simulation der Hörempfindung Schärfe, DIN45692:2009-08.
- [10] Oetjen, A., Letens, U., van de Par, S.: Modellierung der Wahrnehmung von nichtstationären tonalen Komponenten. Motor- und Aggregate-Akustik. 10. Magdeburger Symposium, 2018, Tagungsband S. 62-74.