

# Validierung einer Methode zur gleichzeitigen Bewertung von Rauigkeit, Schärfe, Tonhaltigkeit, Lautheit und Lästigkeit

Lena Schell-Majoor<sup>1</sup>, Jan Rennies<sup>1</sup>, Stephan D. Ewert<sup>2</sup>, Birger Kollmeier<sup>2</sup>

<sup>1</sup> Fraunhofer IDMT, Hör-, Sprach-, und Audiotechnologie & Cluster of Excellence Hearing4All, 26129 Oldenburg, E-Mail: [Lena.Schell-Majoor@idmt.fraunhofer.de](mailto:Lena.Schell-Majoor@idmt.fraunhofer.de); [Jan.Rennies@idmt.fraunhofer.de](mailto:Jan.Rennies@idmt.fraunhofer.de)

<sup>2</sup> Medizinische Physik & Cluster of Excellence Hearing4All, Universität Oldenburg, 26111 Oldenburg, E-Mail: [Stephan.Ewert@uni-oldenburg.de](mailto:Stephan.Ewert@uni-oldenburg.de); [Birger.Kollmeier@uni-oldenburg.de](mailto:Birger.Kollmeier@uni-oldenburg.de)

## Einleitung

Die Bewertung der Qualität von Geräuschen ist sowohl im Entwicklungsprozess als auch in der Anwendung von vielen technischen Geräten und Produkten von Bedeutung, z.B. für Haushaltsgeräte oder in Fahrzeugen. Modelle für die Vorhersage der Geräuschqualität ermöglichen eine schnelle und kostengünstige Abschätzung der zu erwartenden Probandenurteile. Eine gängige Methode zur Modellierung der Geräuschqualität ist die Wichtung und Kombination psychoakustischer Parameter, wie z.B. Lautheit, Schärfe und Rauigkeit zu einem Gesamtqualitätsurteil, wie z.B. in [1,2,3]. Für diese Methode wird die Qualität einer Auswahl von Geräusche im Experiment von Probanden beurteilt. Anschließend werden mittels Regression die Wichtungen für die einzelnen Parameter bestimmt. Die Parameter zur Beschreibung der Qualitätsbewertungen werden hierbei instrumentell ermittelt, d.h. mit Modellen, Standards oder Berechnungsvorschriften aus der Literatur. Diese instrumentellen Maße sind zwar in der Regel aus subjektiven Daten abgeleitet, aber nicht unbedingt für alle Geräusche validiert bzw. anwendbar. Für die weitere Erforschung grundlegender Mechanismen bei der Bewertung der Geräuschqualität und auch im Hinblick auf generalisierende Modellansätze sind daher Datensätze nützlich, in denen sowohl die Qualitätsbewertung als auch die Ausprägung der psychoakustischen Parameter experimentell ermittelt wurden. In dieser Studie stellen wir eine Methode zur Erhebung dieser Daten vor und überprüfen die Zuverlässigkeit der Ergebnisse.

## Methode

Es wurden zwei Experimente mit der gleichen Methode, aber unterschiedlichen Sets von Geräuschen durchgeführt. An jedem Experiment nahmen 30 normalhörende Probanden teil. Die Geräusche wurden dichotisch über Kopfhörer in einer Hörkabine dargeboten. Die Reihenfolge der Geräusche wurde für jeden Probanden randomisiert. Jedes Experiment bestand aus einer Test- und einer Retestsitzung, die mindestens eine Woche auseinander lagen. Im ersten Experiment wurden drei Geräusche in jeder Session zweimal dargeboten und einige der Geräusche aus Experiment 1 wurden ebenfalls in Experiment 2 dargeboten. Vor den eigentlichen Bewertungen wurden den Probanden neun der im jeweiligen Experiment vorkommenden Geräusche dargeboten, um ihnen einen Eindruck der zu erwartenden

Geräusche zu vermitteln. Für alle Geräusche wurde die Rauigkeit, Schärfe, Tonhaltigkeit, Lautheit und Lästigkeit auf einer Skala von 0 bis 50 mit einem Schieberegler bewertet. Als zusätzliche Orientierung waren fünf verbale Marken [4] an der Skala angegeben. Die Probanden konnten die Geräusche beliebig oft wiederholen und so die Bewertungen für die verschiedenen Attribute einstellen.

## Stimuli

In Experiment 1 wurden 74 verschiedene Geräusche dargeboten. Vier davon waren reale, aufgenommene Geräusche, die anderen synthetische Geräusche wie Bandpassrauschen, reine und amplitudenmodulierte (AM) Töne, sowie weißes Rauschen. Die Geräusche hatten eine berechnete Lautheit von 4 sone, bis auf das weiße Rauschen (10 sone). In Experiment 2 wurden 53 aufgenommene Geräusche technischer Geräte, z.B. Rasierer und Staubsauger, mit einer berechneten Lautheit von 12 sone dargeboten. Zusätzlich wurden ein reiner Ton und 23 AM-Töne, wiederum mit einer berechneten Lautheit von 4 sone, sowie weißes Rauschen mit einer Lautheit von 10 sone dargeboten.

## Ergebnisse

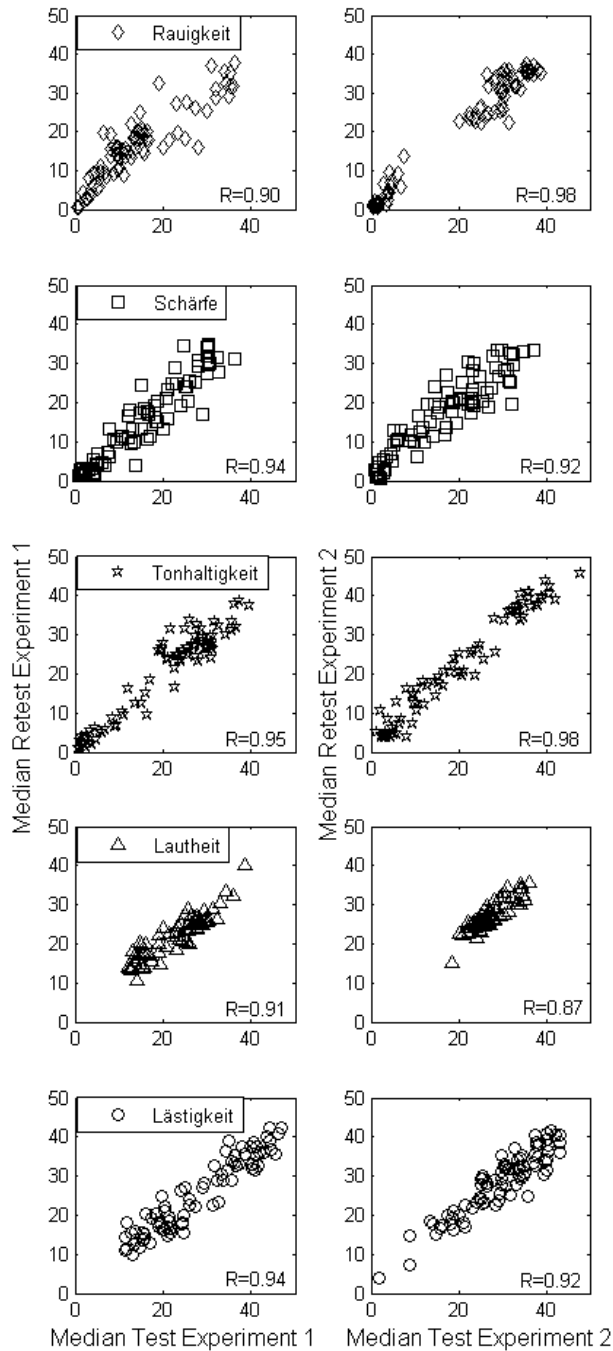
### Vergleich der Bewertungen von gleichen Geräuschen innerhalb einer Sitzung

Für die Prüfung, ob die Probanden grundsätzlich in der Lage sind ihre Bewertungen zu reproduzieren, wurden in jeder Sitzung des ersten Experiments drei Geräusche doppelt dargeboten. Die Differenzen zwischen den Medianen der ersten und zweiten Bewertung liegen zwischen +3 und -6 Skaleneinheiten. Diese Abweichungen sind gering und daher ist davon auszugehen, dass die Probanden stabile Bewertungen innerhalb einer Sitzung abgeben.

### Vergleich der Bewertungen in Test- und Retestsitzung

Der Vergleich der Ergebnisse aus den Test- und Retestsitzungen ermöglicht die Untersuchung der Frage, ob die Bewertungen der Probanden auch über die Zeit stabil bleiben. In Abbildung 1 sind die Mediane der Bewertungen aus der Test- (x-Achse) und der Retestsitzung (y-Achse) für die einzelnen Attribute und separat dargestellt. Die Ergebnisse zeigen, dass für fast alle Attribute in beiden Experimenten Korrelationen  $R > 0.9$  erreicht werden. Die einzige Ausnahme ist die Lautheit in Experiment 2

( $R = 0.87$ ). Hier ist die reduzierte Korrelation auf den eingeschränkten Wertebereich der Bewertungen zurückzuführen. Insgesamt zeigen die Ergebnisse also, dass die Bewertungen auch über die Zeit stabil bleiben.



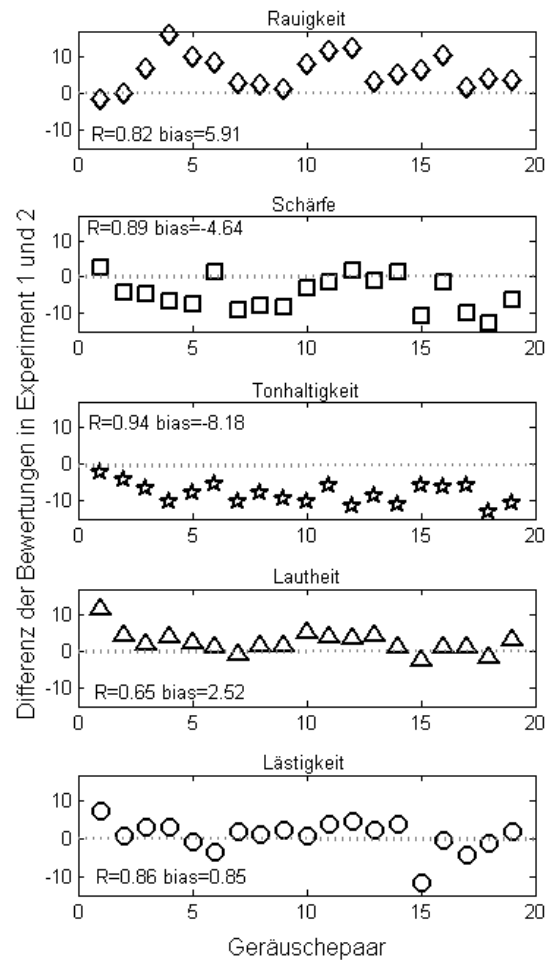
**Abbildung 1:** Vergleich der Mediane der Bewertungen aus Test- (x-Achse) und Retestsitzung (y-Achse) links für Experiment 1 und rechts für Experiment 2, mit den jeweiligen Korrelationskoeffizienten R für die verschiedenen Attribute.

**Vergleich der Bewertungen von gleichen Geräuschen in Experiment 1 und Experiment 2**

Um zu untersuchen, inwiefern der Kontext, hier insbesondere die Auswahl der Geräusche, einen Einfluss auf die Bewertungen hat, wurden 19 Geräusche in beiden Experimenten dargeboten. Abbildung 2 zeigt die

Differenzen zwischen den Medianen der Bewertung in Experiment 1 und 2. Darüber hinaus wurde der Bias berechnet, der angibt, um welchen Wert ein Datensatz aus Experiment 2 verschoben werden muss, um die Root-Mean-Square-Abweichungen zum Datensatz aus Experiment 1 zu minimieren. Für Schärfe und Tonhaltigkeit ist der Bias deutlich negativ (-4,6 und -8,2), was bedeutet, dass die gleichen Geräusche in Experiment 2 schärfer und tonhaltiger beurteilt wurde. Die Rauigkeit wurde hingegen in Experiment 2 niedriger bewertet (Bias = 5.9) und auch für die Lautheit ist eine leichte Verschiebung zu niedrigeren Bewertungen in Experiment 2 zu erkennen (Bias = 2.5).

Für die Lästigkeitsbewertungen zeigt sich kein systematischer Unterschied (Bias = 0.9). Als Anhaltspunkt für die Systematik dieser Abweichungen wurde der Korrelationskoeffizient berechnet. Außer für die Lautheit zeigen die Daten aus Experiment 1 und 2 jeweils hohe Korrelationen von  $R > 0.8$ . Das bedeutet, dass die Lage der Bewertungen zueinander sich nur wenig ändert. Die reduzierte Korrelation bei der Lautheit ( $R = 0.65$ ) ist vermutlich wieder auf den reduzierten Wertebereich zurück zu führen. Es gibt demnach einen Einfluss der Auswahl der Geräusche auf die Bewertungen, der in seiner Ausprägung auch vom jeweiligen Attribut abhängt.



**Abbildung 2:** Abweichungen zwischen den Bewertungen in Experiment 1 und 2 mit Korrelationskoeffizient R und dem Bias für die Geräusche, die in beiden Experimenten dargeboten wurden.

## Diskussion

Insgesamt zeigen die Ergebnisse, dass die Methode zu verlässlichen Bewertungen für alle Attribute bei einer festen Auswahl von Geräuschen führt. Die Test-Retest-Korrelationen liegen alle über 0,8, was von Guski [5] als Reliabilitätsmerkmal beschrieben wurde. Für die Differenzen zwischen den doppelt bewerteten Geräuschen gibt es keine Vergleichswerte in der Literatur. Betrachtet man die Abweichungen von bis zu 6 Einheiten im Verhältnis zur 50-stufigen Gesamtskala betragen sie etwa 10% und sind als gering zu bewerten. Bei den in beiden Experimenten bewerteten Geräuschen sind die Einzelabweichungen für manche Attribute mit fast 15 Einheiten mehr als doppelt so groß. Darüber hinaus weichen die Bewertungen hier meist systematisch in die gleiche Richtung ab und zeigen so, dass die Geräuschauswahl einen Einfluss auf die Bewertungen hat. Dieser Einfluss ist unterschiedlich groß für die unterschiedlichen Attribute. Für eine genauere Quantifizierung dieses Einflusses müssten jedoch weitere Untersuchungen durchgeführt werden, da die Anzahl der Geräusche, die in beiden Experimenten dargeboten wurden, mit 19 eher klein war und es sich hierbei ausschließlich um amplitudenmodulierte Töne handelte. Für die Anwendbarkeit in der Industrie sollte die Methode eine schnelle und einfache Bewertung von vielen Geräuschen ermöglichen. Die Experimente haben gezeigt, dass die Bewertung von ca. 80 Geräuschen mit einer Dauer von jeweils 2-4 Sekunden etwa eine Stunde dauert. Die Probanden waren untrainiert, d.h. sie hatten keine Erfahrungen mit Hörversuchen und die Attribute wurden den Probanden nicht erklärt oder anhand von Hörbeispielen verdeutlicht.

## Schlussfolgerungen und Ausblick

Die hier vorgestellte Methode ermöglicht eine schnelle Bewertung vieler Geräusche mit untrainierten Probanden. Für eine feste Auswahl an Geräuschen führt diese Methode zu verlässlichen Ergebnissen hinsichtlich der Mediane. Der gewonnene Datensatz wird hinsichtlich weiterer Fragestellungen weiter ausgewertet. So soll anhand der gewonnenen Daten auch untersucht werden, welchen Beitrag die einzelnen psychoakustischen Attribute zur Bewertung der Lästigkeit liefern.

## Literatur

- [1] Ellermeier, W., Kattner, F., Kurtze, L., Bös, J.: Psychoacoustic characterization of the noise produced by photovoltaic inverters. *Acta Acustica united with Acustica* 100 (2014), 1120-1128.
- [2] Kim, E.Y., Shin, T.J., Lee, S.K.: Sound quality index for assessment of sound quality of laser printers based on a combination of sound metrics. *Noise Control Engr. J.* 61 (6) (2013), 534-546.
- [3] Kuwano, S., Seiichiro, N.: Subjective impression of copy machine noises: An examination of physical metrics for the evaluation of sound quality. *Proceedings of Inter-noise 2009* (2009).
- [4] Rohrmann, B.: Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie* Bd. 9 (1978), 222-245.
- [5] Guski, R.: Psychological methods for evaluating sound quality and assessing acoustic information. *Acta Acustica united with Acustica* 83 (1997), 765-774.