

Stärken und Schwächen von Pegelmaßen und Lautheitsmodellen bei der Bewertung von Musik

Florian Schmidt, Birger Kollmeier, Stefan Uppenkamp

Medizinische Physik, Universität Oldenburg, 26111 Oldenburg

E-Mail: florian.schmidt@uni-oldenburg.de

Einleitung

Es ist eine seltsam anmutende Beobachtung, dass viele Jahrzehnte nach der Entwicklung des ersten Lautheitsmodells in der Praxis zur Bestimmung der Lautheitswahrnehmung von Musik hauptsächlich frequenzgewichtete Pegelmaße benutzt und sogar empfohlen werden [1], [2]. Dabei sind Pegelmaße eigentlich nur grobe Schätzer der Lautheit, da sie längst nicht alle physiologisch relevanten Lautheitseffekte berücksichtigen wie bspw. spektrale Lautheitssummutation oder temporale Lautheitsintegration [3]. Sollten nicht die Lautheitsmodelle den Pegelmaßen überlegen sein? Wenn nicht: worin liegen die Stärken und worin die Schwächen von Pegelmaßen und Lautheitsmodellen bei der Bewertung von Musik? Und gibt es vielleicht vernachlässigte Einflussgrößen, die helfen könnten die Lautheitsmodelle für diese Aufgabe zu modifizieren?

Methode

Es gibt verschiedene Möglichkeiten Modelle qualitativ miteinander zu vergleichen. Eine Methode besteht darin, eine Stichprobe unterschiedlich skaliert Messobjekte durch die Modelle schätzen zu lassen, und aus den gewonnen Skalenwerten eine Modellskala zu bilden, und diese mit einer experimentell gewonnenen Skala zu vergleichen. Es gibt verschiedene psychometrische Methoden eine solche Skala zu gewinnen. Für schwer erfassbare Messgrößen gilt es, den Schwierigkeitsgrad der Aufgabenstellung im Experiment möglichst gering zu halten ohne dabei allzu viel Informationsgehalt der Messdaten preiszugeben. Der Paarvergleich ermöglicht zum Einen denkbar einfachste Aufgabenstellungen und liefert zum Anderen unter gewissen Umständen einen hohen Gehalt an Skaleninformation. Letzteres lässt sich für Paarvergleiche mit binomialer Antwortmöglichkeit mithilfe des Bradley-Terry-Luce-Verfahren (BTL) bewerkstelligen [4].

Beim BTL-Verfahren wird das Verhältnis zweier Bewertungsgrößen zueinander aus den Informationen mehrfach durchgeführter Vergleiche geschätzt. Basierend auf der Annahme, dass der Paarvergleich einem Bernoulli-Prozess folgt, können aus den Gewinnwahrscheinlichkeiten jeweiliger Bewertungsgrößen eigene Skalenwerte abgeleitet werden. Das Skalenniveau dieser Werte entspricht dann dem einer Verhältnis-Skala.

Die Modellskalen der in dieser Studie untersuchten Lautheitsmaße liefern Informationsgehalt auf Intervallniveau. Die Qualität dieser Skalen kann also ermittelt werden, indem geprüft wird, wie gut Aussagen über die Abstände zwischen den Bewertungsgrößen getroffen werden können. Dies geschieht in dieser Studie mithilfe einer kreuzvalidierten Korrelationsanalyse. Das hieraus gewonnene Bestimmtheitsmaß R^2 gibt darüber Auskunft, wie ähnlich die Modelle gegenüber der experimentellen Methode die Bewertungsgrößen skalieren können.

Ergebnisse

Aus fünf verschiedenen Musik-Genres: Klassik, Hip-hop, Jazz, Rock und Pop, wurden jeweils zwei Musikstücke gewählt, aus denen wiederum zwei unterschiedliche 10s-Ausschnitte für den Paarvergleich ausgewählt wurden - d.h. 5x2x2 Stimuli. Die Ausschnitte wurden so gewählt, dass der Dynamikbereich gleiche Pegelbreite hat. Der Dynamikbereich wird hier über die Pegeldifferenz des 20. und des 95. Perzentils definiert und wurde auf 2 dB festgelegt. Der Dauerschallpegel der jeweiligen Ausschnitte wurde unterschiedlich eingestellt und liegt zwischen 60 und 95 dB SPL. Es wurden 29 normalhörende Versuchspersonen für das Experiment ausgewählt.

Bei der Auswertung der Daten mittels des BTL-Verfahrens wurde das MATLAB-Skript von F. Wickelmaier et al. [5] verwendet. Für die Erzeugung der experimentellen Skala mittels des BTL-Verfahrens wurden nur bewertete Musikausschnitte berücksichtigt, die zu ihrem etwas lauterem und etwas leiserem Nachbarn auf der Skala ein Verhältnis aufwiesen, d.h. im direkten Vergleich nicht einseitig beurteilt wurden. Sechs der zwanzig Stimuli wurden daher für die weitere Analyse nicht berücksichtigt. Des Weiteren ergab die BTL-Skalierung eine mögliche Aufteilung des Datensatzes in zwei Gruppen – G1: leiser und G2: lauter – entlang der Lautheitsskala. Der Diskriminationspunkt lag zwischen den beiden Musikausschnitten, die aufgrund zu vieler einseitiger Verhältnisse zu den darüber hinausliegenden Nachbarn durch das BTL-Verfahren am unsichersten bestimmt werden konnten. Dadurch umfasste die Gruppe G1 sechs und G2 acht Musikausschnitte.

Als Pegelmaße wurden die RMS-Werte des Schalldruckpegels (dB SPL), des A-gewichteten Schalldruckpegels (dB(A)) und die Empfehlung R 128 der Europäischen Rundfunkunion (EBU-R128) verwendet. Die untersuchten Lautheitsmodelle waren: die DIN45631 von

Zwicker sowie dessen Erweiterung für instationäre Signale DIN45631A1, das Dynamische Lautheitsmodell von Fastl & Chalupper und dessen Erweiterung von Rannies & Verhey (DLM_ext), sowie das Lautheitsmodell von Moore & Glasberg (MG02). Alle diese Modelle wurden mithilfe der Fraunhofer IDMT Sound Quality and Speech Intelligibility Prediction Toolbox (SIP-Toolbox) benutzt.

Über den in diesem Experiment gesamt betrachteten Lautheitsbereich konnten die Pegelmaße und Lautheitsmodelle sowohl Reihenfolge der Musikausschnitte als auch Lautheitsabstände zueinander befriedigend einschätzen (für alle gilt: $R^2 > 0,8$, s. Abb. 1). In der Qualität jedoch leicht herausragend zeigten sich die Pegelmaße dB(A) und EBU-R128, was die weitere Korrelationsanalyse der jeweiligen Gruppen G1 und G2 bestätigte. Bei den leiseren Musikausschnitten in G1 konnte dB(A) beinahe varianzlos und qualitativ mit großem Abstand gegenüber den anderen Lautheitsmaßen die Lautheit schätzen ($R^2 = 0,97$), bei den lauterer Musikausschnitten war es die EBU-R128 Empfehlung, die mit Abstand die besten Ergebnisse lieferte ($R^2 = 0,74$).

Um zu prüfen, ob es Einflussgrößen gibt, die für die abweichenden Schätzungen der Lautheitsmodelle verantwortlich sind, wurde eine lineare Diskriminanzanalyse durchgeführt. Unter der Annahme, dass die Lautheitsmodelle die Lautheit von Musik linear skalieren, wurde der lineare Fit aus Modell und Paarvergleichsdatensatz als Diskriminanz-Gerade verwendet, um vom Modell über- und unterschätzte Musikausschnitte zu unterscheiden. Als Einflussgrößen wurden modellierte psychoakustische Größen verwendet (entnommen aus der SIP-Toolbox, s.o.): Schärfe (DIN45692), Rauigkeit (in asper) und Tonhaltigkeit (nach Aures). Bei dieser Analyse hat sich gezeigt, dass alle Lautheitsmodelle deutlich durch die Parameter Schärfe und Tonhaltigkeit bei der Über- und Unterschätzung der Lautheit beeinflusst werden (exemplarisch: Abb.2).

Lautheitsmaße	R ²	R ² (G1)	R ² (G2)
dB SPL	0,82	0,01	0,52
dB(A)	0,93	0,97	0,51
EBU-R128	0,92	0,50	0,74
DIN45631	0,85	0,69	0,36
DIN45631A1	0,90	0,63	0,55
MG02	0,86	0,29	0,57
DLM	0,89	0,50	0,65
DLM_ext	0,80	0,39	0,31

■ 0-0,4 ■ 0,4-0,7 ■ 0,7-1

Abbildung 1: Pegelmaße und Lautheitsmodelle werden bewertet in ihrer Qualität die Abstände zueinander zu schätzen für den gesamten Skalenbereich des Paarvergleichs und für den in zwei Gruppen aufgeteilten Skalenbereich G1: leiser und G2: lauter

Diskussion

Die Ergebnisse dieser Studie deuten darauf hin, dass die Lautheitswahrnehmung von Musik zwar maßgeblich vom

Pegel dominiert wird, aber in deren Feinauflösung Lautheitseffekte wie die spektrale und temporale Lautheitsintegration eine untergeordnete Rolle spielen. So bringen Lautheitsmodelle gegenüber den Pegelmaßen eher eine Verschlechterung der Schätzung als eine Verbesserung. Stattdessen scheinen andere Größen, die vielleicht eher die Qualität oder die Lästigkeit des Stimulus betreffen, einen größeren Einfluss auf die Lautheitswahrnehmung von Musik zu haben wie bspw. Schärfe oder Tonhaltigkeit. Abhängig vom Präsentationspegel zeigte sich, dass durch eine Kombination aus A-gewichteten Schalldruckpegel und der EBU-R128 Empfehlung die Lautheit am besten geschätzt werden konnte.

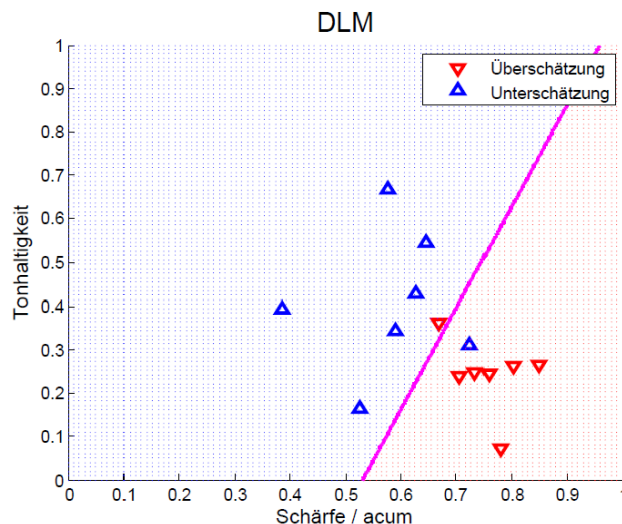


Abbildung 2: Lineare Diskriminanzanalyse für die in der Lautheit über- und unterschätzten Stimuli des Dynamischen Lautheitsmodells von Fastl&Chalupper (DLM) durch die Parameter Tonhaltigkeit und Schärfe.

Literatur

- [1] Bray, A.: Noise induced hearing loss in dance music disc jockeys and an examination of sound levels in nightclubs. The Journal of Laryngology & Otology Feb. 2004, Vol. 118, pp. 123-128.
- [2] International Telecommunication Union: Recommendation ITU-R BS.1770-2 - Algorithms to measure audio programme loudness and true-peak audio level. Electronic Publication, Geneva, 2011.
- [3] Rannies, J.: Comparison of loudness models for time-varying sounds. Acta Acustica united with Acustica 96 (2010), 383-396.
- [4] Tsukida, K.: How to Analyze Paired Comparison Data. UWEE Technical Report Number UWEETR-2011-0004 (2011).
- [5] Wickelmaier, F.: A Matlab function to estimate choice model parameters from paired-comparison data. Behavior Research Methods, Instruments, & Computers 2004, 36 (1), 29-40.