Evaluierung und Verbesserung eines psychoakustischen Modells für musikalische Härte

Isabella Czedik-Eysenberg[1], Denis Knauf[2], Christoph Reuter[1]
[1]Institut für Musikwissenschaft, Universität Wien; [2]Technische Universität Wien

Jahrestagung 2017
Universität Hamburg
11.-13. September

Hintergrund

„Harte Musik“ wird oft mit Heavy Metal gleichgesetzt (Reyes, 2008), wodurch die Klassifizierung von musikalischer Härte qua definitionem an die Genrezugehörigkeit gekoppelt wird. Ob ein Musikstück dem Höreindruck nach allerdings tatsächlich (genreunabhängig) als „hart“ empfunden wird, und von welchen klanglichen Kriterien das abhängt, war Thema einer vorangegangenen Untersuchung (Czedik-Eysenberg et al., 2017):
In dieser Vorstudie waren 62 Musikausschnitte unterschiedlicher Genres 40 Probanden dargeboten worden, welche jene Stimuli hinsichtlich ihres subjektiven Härteeindrucks bewerteten. Nach signalanalytischen Untersuchungen konnte eine Reihe von relevanten klanglichen Eigenschaften als härtebestimmend identifiziert werden: So zeigte sich besonders die Intensität **perkussiver Signalkomponenten** in Zusammenhang mit der Härteempfindung als aussagekräftiger akustischer Deskriptor. Auch die starke Ausprägung **hoher Frequenzbereiche** (vgl. auch Reyes, 2008; Berger & Fales, 2005) - insbesondere zwischen 2 und 4 kHz - schien eine Rolle zu spielen. Zudem ließ sich eine **flache dynamische Hüllkurve** (möglicherweise u.a. in Folge von Klangverzerrungen, siehe auch Berger & Fales, 2005) mit insgesamt **geringerem Dynamikumfang** und **hoher Lautheit** (vgl. auch Weinstein, 1991; Walser, 1993) feststellen. Jene akustischen Merkmale zeigten sich im Wesentlichen (jedoch in unterschiedlichen Ausprägungen) auch als genreübergreifend. Mittels multipler linearer Regression wurde ein erstes Messmodell aufgestellt, welches eine Korrelation von r = 0,86 mit den subjektiven Bewertungen der Versuchspersonen erreichte.

Fragestellungen und Ziele

→ Evaluierung: Inwieweit erweist sich das bisher aufgestellte Modell als tauglich zur Vorhersage der Härtebewertungen anhand eines neuen Stimuluskorpus?

→ Verbesserung des Ergebnisses mithilfe neuer Hörversuchsergebnisse und verfeinerter Klanganalysen

→ Betrachtung genreabhängiger Faktoren

Methode

Zur Evaluierung wurden 150 Musikbeispiele aus 10 verschiedenen Genres auf Grundlage von LastFM-Hörerstatistiken selektiert.

Aus dem Refrain der Stücke wurden Ausschnitte von jeweils 10-15 Sekunden herangezogen und in ihrer Lautheit angeglichen (500 Sone).

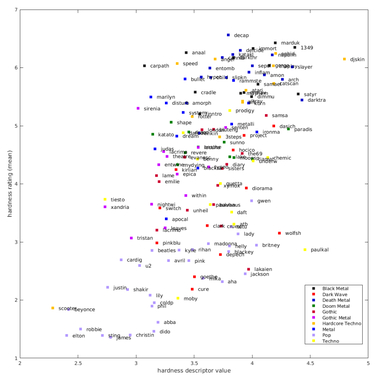
In einem neuerlichen Hörversuch wurden diese Stimuli wiederum von 40 Personen hinsichtlich ihrer Härte bewertet, um anhand dieser Ergebnisse die Präzision der Vorhersage durch das bestende Modell zu testen.

Zusätzlich wurden insgesamt 230 Signaleigenschaften über die spektrale und zeitliche Verteilung der betrachteten Stimuli unter Verwendung von Essentia (Bogdanov et al., 2013), MIRtoolbox (Lartillot & Toiviainen, 2007), Loudness Toolbox (Genesis, 2009) und TSM Toolbox (Driedger & Müller, 2014) erhoben. Diese Daten wurden einer weiteren explorativen Analyse unterzogen und mittels Feature-Selektions und -Regressionsverfahren unter 5-facher Kreuzvalidierung neue Modelle gebildet.

Evaluierung des ersten Modells

Die Vorhersage durch den ersten Härtedeskriptor zeigt eine mittlere Korrelation ($r = 0.566$, $p < 0.001$) mit den tatsächlichen Härtebewertungen der Testklänge.

Lautheitsabhängigkeit als Schwäche: Angleichung wirkt sich auf die absolute Intensität der perkussiven Signalkomponenten aus.

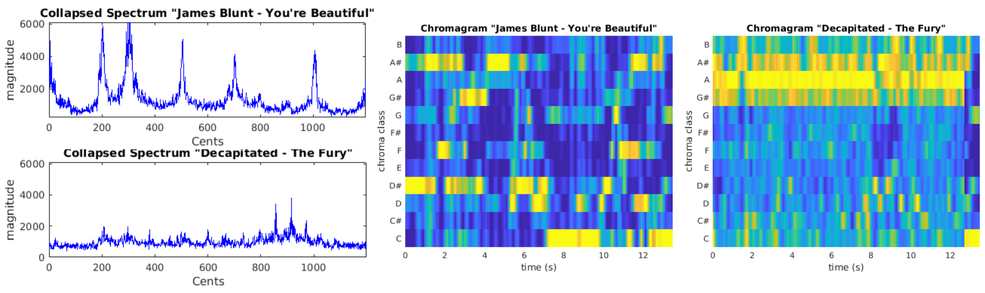


Model Feature	r	p
2 - 4 kHz Energy	0,465	< 0,001
Low Centroid Rate	-0,320	< 0,001
Percussive Energy	-0,086	0,299
Harmonic/Percussive Ratio	*-0,552*	*< 0,001*

Robuster lässt sich dieser Zusammenhang anhand des Verhältnisses perkussiver gegenüber harmonischer Signalkomponenten messen.

Relevante Klangeigenschaften

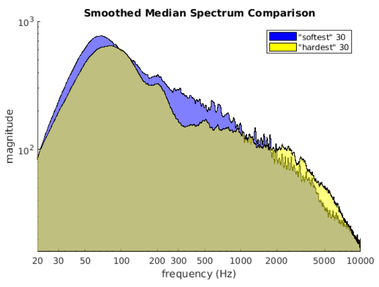
1. Tonklarheit / Harmonizität



„wenig Melodie“, unklare Tonalität / harmonische Dissonanzen (siehe auch Berger, 1999)

Feature	r	p
HPCP Entropy	0,781	< 0,001
Dissonance	0,814	< 0,001
Diatonic Strength	-0,340	< 0,001

2. Spektrale Verteilung



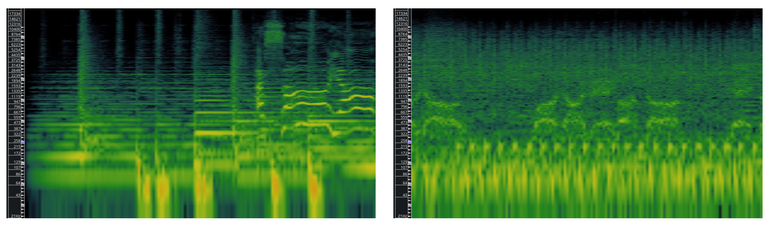
Bei Betrachtung der spektralen Verteilung zeigt sich:
Die härtesten 30 Klangbeispiele weisen bei gleicher Lautheit signifikant geringere Intensität im Mittenbereich zwischen 150 und 1000 Hz (Gammatone-Band 3 & 4) und höhere oberhalb von 2 kHz (Band 6-8) auf als die 30 am wenigsten harten.

Feature	r	p
RMS Gammatone 3	-0,527	< 0,001
RMS Gammatone 4	-0,476	< 0,001
RMS Gammatone 6	0,303	< 0,001
RMS Gammatone 7	0,339	< 0,001
RMS Gammatone 8	0,422	< 0,001
Spectral Skewness	-0,650	< 0,001
Spectral Centroid	0,631	< 0,001
Spectral Rolloff	0,473	< 0,001
MFCC 1	-0,618	< 0,001
High Freq. Content	0,515	< 0,001

3. Tempo / rhythmische Dichte / perkussive Komponenten

→ dichteres Klangbild mit gleichmäßigen perkussiven Einsätzen (rechtes Sonagramm)

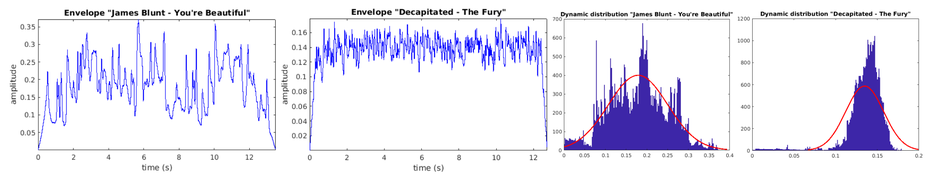
Feature	r	p
harmonicPercussiveRatio	-0,552	< 0,001
Event Density	0,491	< 0,001
Tempo (Onsets)	0,487	< 0,001
Rythmic Irregularity	-0,461	< 0,001



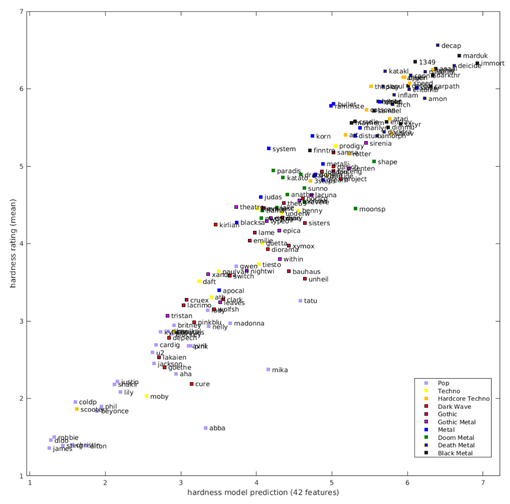
4. Dynamische Bandbreite

→ geringere dynamische Bandbreite mit weniger Variation

Feature	r	p
Dynamic Complexity	-0,523	< 0,001
Envelope Quantile Range	-0,383	< 0,001



Verbessertes Regressionsmodell



Es zeigten sich 120 - nach Bonferroni-Korrektur 65 - signifikante Korrelationen mit gemessenen Signaleigenschaften.

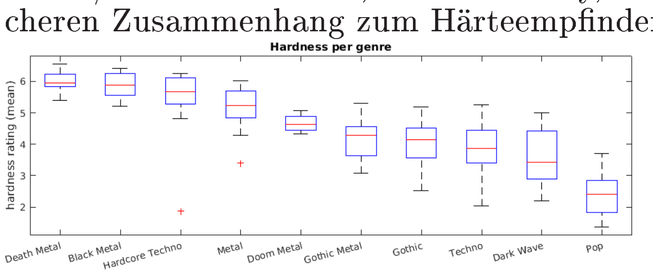
Zur Selektion eines optimalen Featuresets wurden verschiedene Methoden zur Reduktion der Dimensionalität (Sequential Forward Selection, LASSO, Partial-Least-Squares-Regression) unter 5-facher Kreuzvalidierung eingesetzt.

Um Overfitting auch bei der Modellbildung zu verhindern wurden schließlich ebenfalls unter 5-facher Kreuzvalidierung ein Regressionsmodell konstruiert.

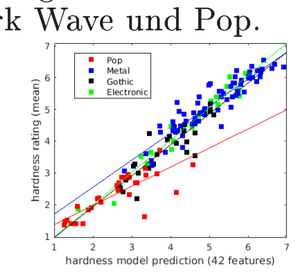
Regressionsmodell auf Basis von	RMSE	R-Squ.	MSE	MAE	r
LASSO → 18 Features	0,62	0,80	0,39	0,47	0,925
Sequential Feature Selection → 42 Features	0,57	0,84	0,33	0,44	0,957
Sequential Feature Selection Top 5 Features	0,64	0,80	0,40	0,49	0,900

Genreabhängigkeit

Innerhalb der Stimuli-Subgruppen konnten im Wesentlichen übereinstimmende Kriterien für Härte gefunden werden. Eine gewisse Ausnahme bildeten rhythmische/perkussive Eigenschaften (Harmonic/Percussive Ratio, Event Density, ...), welche im Fall von Metal-Subgenres einen deutlicheren Zusammenhang zum Härteempfinden aufwiesen als bei Gothic/Dark Wave und Pop.



Mit der geringsten Genauigkeit konnte die musikalische Härte daher für Klangbeispiele aus dem Genre Pop modelliert werden.



Zusammenfassung & Ausblick

→ Es war möglich, akustische Signaleigenschaften, welche in Zusammenhang mit der empfundenen „Härte“ stehen, zu identifizieren und mit deren Hilfe ein verbessertes (weitgehend) genreunabhängiges Härte-Modell aufzustellen.

→ „Musikalische Härte“ bietet sich als hilfreicher High-Level-Deskriptor bei der interaktiven Suche von Musikdateien an.

→ Das Härte-Modell wurde in eine Softwareanwendung integriert, mit deren Hilfe eine interaktive Suche von Klängen basierend auf akustischen High-Level-Deskriptoren möglich ist. Langfristiges Ziel ist u.a. auch die Entwicklung einer mobilen Applikation zur Detektion musikalischer Härte.

Literatur

Berger, H. (1999). Metal, Rock and Jazz: Perception and the Phenomenology of Musical Experience. Hanover, N.H.: Wesleyan University Press/University Press of New England | Berger, H. M. & Fales, C. (2005). 'Heaviness' in the Perception of Heavy Metal Guitar Timbres: The Match of Perceptual and Acoustic Features over Time. Wired for Sound: Engineering and Technologies in Sonic Cultures. Middletown, CT: Wesleyan University Press, 181-197. | Bogdanov, D., Wack N., Gómez E., Gulati S., Herrera P., Mayor O., et al. (2013). ESSENTIA: an Audio Analysis Library for Music Information Retrieval. International Society for Music Information Retrieval Conference (ISMIR'13). 493-498. | Czedik-Eysenberg, I.; Knauf, D. & Reuter, C. (2017). Was macht Musik "hart"? Klangliche Merkmale zur genre-übergreifenden Identifikation musikalischer Härte. Fortschritte der Akustik. DAGA 2017. Kiel 2017, 186-189. | Driedger, J. & Müller, M. (2014). TSM Toolbox: MATLAB Implementations of Time-Scale Modification Algorithms. In: Proc. of the International Conference on Digital Audio Effects. | Genesis (2009). Loudness toolbox, http://www.genesis-acoustics.com/en/index.php?page=32 | Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio, International Conference on Digital Audio Effects, Bordeaux, S. 237-244. | Reyes, I. (2008). Sound, Technology, and interpretation in Subcultures of Heavy Music Production. Doctoral Dissertation, University of Pittsburgh. | Walser, R. (1993). Running with the devil: Power, gender, and madness in heavy metal music. Wesleyan University Press. | Weinstein, D. (1991) Heavy Metal: A Cultural Sociology. New York, Lexington Books. |