„This is (not) a Love Song“ – Über die Beziehung zwischen Themengebiet und Audiosignalmerkmalen in populärer Musik.

Virtuelle Postertagung
3.–6. September
2020

Isabella Czedik-Eysenberg[1], Oliver Wieczorek[2], Christoph Reuter[1]

[1]Institut für Musikwissenschaft, Universität Wien; [2]Lehrstuhl für Soziologie, insbesondere soziologische Theorie, Universität Bamberg

Hintergrund

Werden in besonders tanzbaren Musikstücken vermehrt Party-Themen besungen? Deutet ein verzerrter, rauer Sound darauf hin, dass der Text von einem Rock'n'Roll-Lebensstil oder rauen und vulgären Themen handelt?

Liedtexte stellen komplementär zu den klanglichen Eigenschaften eine weitere Ebene von Musik dar, deren Erforschung in Verknüpfung mit der ersteren Ebene Aufschluss über die multimodale Entstehung emotionaler Stimmungen oder der Interpretation von Musik geben kann (siehe etwa Hu et al., 2009; Choi 2016). Am Beispiel des Metal-Genres konnte gezeigt werden, dass textuelle Inhalte, die sich etwa auf *dystopische* Themen, *brutale* Todesschilderungen und *Satanismus* beziehen, mit musikalischen Eigenschaften assoziiert sind, die auf einen besonders *harten* und *düsteren* Klang der Musik hindeuten (Czedik-Eysenberg et al., 2019).

Ziele & Fragestellungen

In diesem Beitrag generalisieren wir diesen Ansatz unter Einbeziehung verschiedener populärer Musikgenres, um:

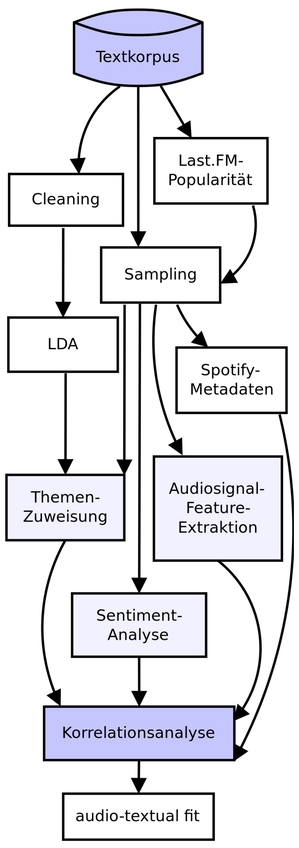
1. einen Überblick über die **textuellen Themen** zu geben, welche innerhalb populärer Musik zu finden sind, und
2. zu untersuchen, inwieweit spezifische textuelle Themen (z.B. *Liebe* oder *Politik*) typischerweise mit bestimmten **klanglichen Musikdimensionen** (z.B. *Tanzbarkeit* oder *Melodiösität*) einhergehen, und
3. inwieweit Audio und Text „zusammenpassen“. (Gehen *positive* Texte mit hoher *Valenz* der Musik einher?)

Daten & Methode (Überblick)

Textanalyse/Topic Models

Über die API von *genius.com* wurde ein Datensatz von 944.553 Songtexten aus 22 unterschiedlichen Genres erstellt. Nach **Bereinigungsschritten** (Einschränkung auf englischsprachige Texte, Entfernen von Stoppwörtern, Stemming), lagen 771.663 Texte vor.

Auf diesen Textkorpus wurde eine **Latent Dirichlet Allocation** (LDA, Blei et al., 2003) angewandt, um ein probabilistisches **Topic-Modell** zu erstellen. Unter Berücksichtigung von Log-Perplexity, Log-U-Mass Coherence sowie inhaltlicher Interpretierbarkeit wurde ein Modell aus 22 Themen gewählt.
Mittels `TextBlob` (Loria et al., 2014) und VADER (Hutto & Gilbert, 2014) wurde eine **Sentiment-Analyse** der Texte durchgeführt.



Musikalische Audiosignalanalyse

Aus dem Gesamtkorpus wurde eine **Teilstichprobe** von 1070 Songs (balanciert nach Popularität auf Last.FM) gezogen. Für diese wurden über die *Spotify*-API eine Reihe von **Audiomerkmalen** erhoben (*acousticness* - *valence*).

Zusätzlich zu diesen wurden weitere High-Level-Musikeigenschaften mittels **Prädiktionsmodellen** extrahiert, welche zuvor mittels **Hörversuchsdaten** und verfügbaren Datensätzen trainiert worden waren (Czedik-Eysenberg et al., 2017; 2018; 2020; Aljanaki 2018).
Jene Modelle basierten auf Merkmalen aus den folgenden Paketen:
- *LibROSA* (McFee et al., 2015)
- *Essentia* (Bogdanov et al., 2013)
- *AudioCommons timbral models* (Pearce et al., 2017)

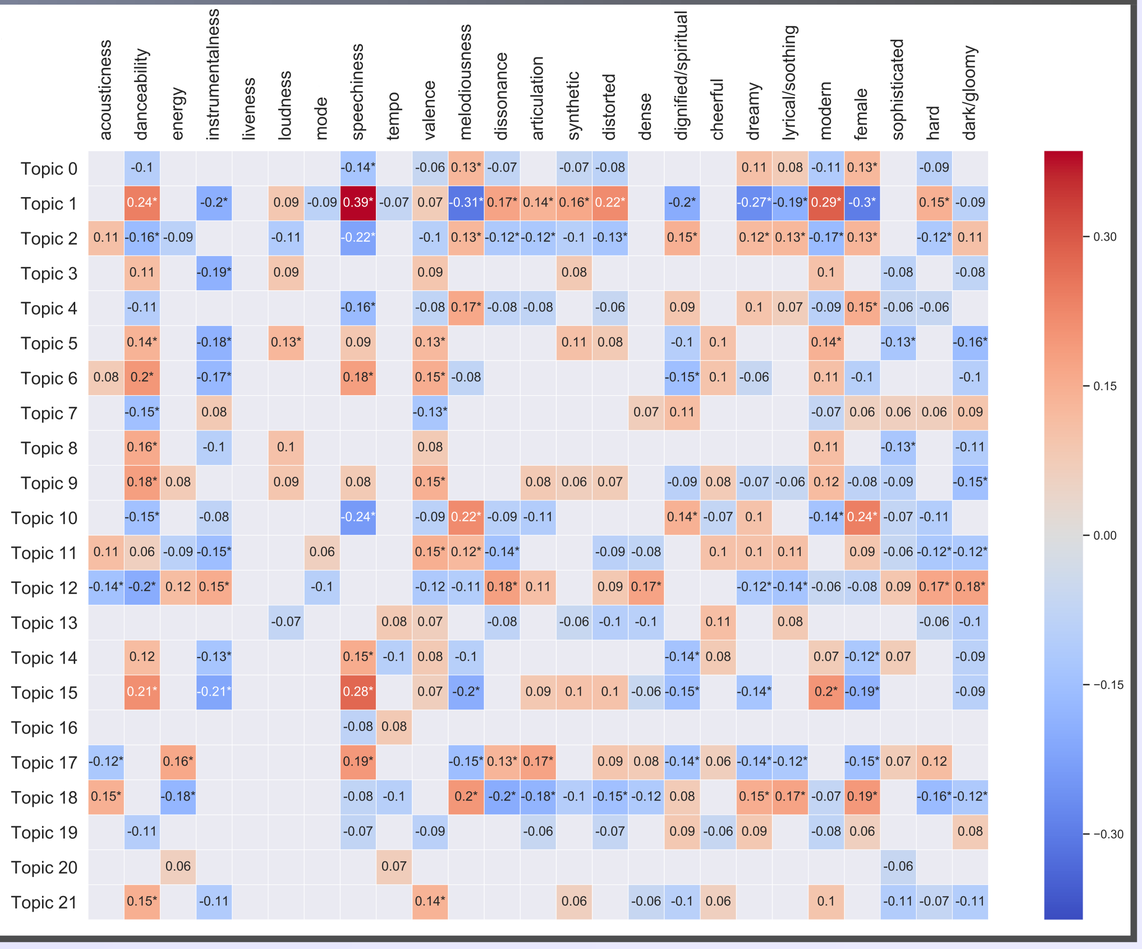
Statistische Auswertung

Mittels Spearman's ρ wurden **Korrelationen** zwischen den errechneten Themen und High-Level-Audiosignalmerkmalen ermittelt. Zudem wurde ein Maß für Audio-Textuellen Fit berechnet.

Ergebnisse: Themen und Audio-Themen-Beziehungen

Die Tabelle unterhalb zeigt die resultierenden **Textthemen**, rechts: **Korrelationen** mit berechneten Audiomerkmalen (alle dargestellten: $p < 0.5$; *: sig. nach Bonferroni-Korrektur).

Thema	Interpretation	Saliente Begriffe/Wortstämme (Top 10)
Topic 0	Places & Motion	away, way, run, home, gone, long, alon, walk, come, stay
Topic 1	Vulgar Language	nigga, fuck, bitch, shit, like, know, ass, caus, real, hit
Topic 2	Landscape/Visual Beauty	rain, sun, eye, sky, blue, wind, water, see, cold, fall
Topic 3	Smalltalk	like, hey, feel, okay, hello, bit, cool, say, look, act
Topic 4	Night (& Day)	time, night, day, light, wait, dream, tonight, last, shine, sleep
Topic 5	Exclamation	yeah, ayi, woah, huh, nobodi, woo, mmm, hoo, whoo, hmm
Topic 6	Abbreviated Language	goin, bout, keep, nothin, lookin, gettin, comin, wit, caus, talkin
Topic 7	Murder & Gory Topics	dead, kill, lie, blood, eye, breath, face, pain, head, hate
Topic 8	Contentment	know, babi, gon, wan, need, make, right, give, caus, tell
Topic 9	Party	danc, stop, rock, move, new, play, roll, parti, beat, bodi
Topic 10	Longing	never, know, see, life, feel, time, tri, live, could, noth
Topic 11	Talking/Dialogue	say, know, think, said, well, thing, friend, tell, would, could
Topic 12	Death/Sinister Topics	death, dark, world, rise, soul, fear, fight, end, life, burn
Topic 13	Regional & Seasonal	black, white, wild, town, countri, west, ohh, south, citi, sunday
Topic 14	Freestyle/Language Games	like, make, caus, kid, rap, rhyme, big, dog, see, kick
Topic 15	Money & Wealth	money, work, big, like, diamond, top, buy, bag, pull, car
Topic 16	Body/Physical Sensation	feel, keep, around, hold, turn, hand, hear, head, round, sound
Topic 17	Politics & Society	state, world, human, new, time, gener, right, law, self, natur
Topic 18	Love	love, heart, yuh, lover, true, nuh, nah, gyal, say, inna
Topic 19	Religion & Worship	god, lord, sing, heaven, name, jesu, soul, king, save, pray
Topic 20	Action	take, high, fire, ride, burn, fli, slow, low, side, drive
Topic 21	Flirt & Nicknames	girl, boy, pleas, bad, look, kiss, sweet, woman, crazi, ladi



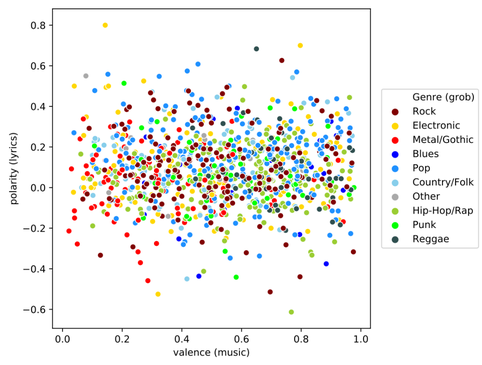
Audio-Textueller Fit & Sentiment-Analyse

Basierend auf den identifizierten Korrelationen wurde ein **Audio-Textueller Fit** (*ATF*) für jedes der Stücke berechnet:

$$ATF = \sum_{i=0}^{N} \sum_{j=0}^{M} topic_i * corr_{ij} * audio_j$$

$\Longrightarrow$ Der durchschnittliche *ATF* über alle Songs betrug 0,42 (SD = 0,66). Die höchsten Werte ließen sich in den Genres *Death/Black Metal* und *(Gangster) Rap* feststellen.

Gemäß **Sentiment-Analyse** der Texte ließ sich *kein* Zusammenhang zwischen der Polarität (negativ bis positiv) der Inhalte und der Valenz der Musik beobachten ($r = 0{,}00$; $p = 0{,}89$).
Das Auftreten von Begriffen mit negativer Polarität war jedoch mit musikalischer *Dissonanz* ($r = 0{,}22$; $p < 0{,}01$), *Härte* ($r = 0{,}20$; $p < 0{,}01$) und *Sprachhaftigkeit* ($r = 0{,}17$; $p = 0{,}01$) verknüpft, während es negative Zusammenhänge mit Klangeigenschaften gab, welche mit einem *beruhigenden* ($r = -0{,}20$; $p < 0{,}01$) und *verträumten* ($r = -0{,}18$; $p < 0{,}01$) Klang in Verbindung stehen.



Zusammenfassung

Im gemeinsamen Auftreten textueller Themen und musikalischer Charakteristika ließen sich einige signifikante Muster beobachten:

- *Vulgäre* Sprache und Texte, welche *Geld* oder *Politik* thematisieren, sind typischerweise durch eine hohe *Sprachhaftigkeit* charakterisiert.
- Von *Party*-Themen handelnde Lieder tendieren dazu, *tanzbar* zu sein und einen *hellen*, *fröhlichen* Klang (hohe *Valenz*) aufzuweisen.
- Texte, welche *Liebe* and *Sehnsucht* besingen, werden tendenziell von *melodiöser* und *ruhiger* Musik begleitet.
- Ein *dissonanter*, *harter* and *düsterer* Klang der Musik verweist auf *düstere* and *brutale* Textthemen wie etwa *Tod*.

Eine generelle Verknüpfung textueller und musikalischer Valenz ließ sich anhand der vorliegenden Daten *nicht* beobachten.

References

Aljanaki, A. (2018). *Mid-level perceptual musical features.* Retrieved from osf.io/5aupt | **Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003).** Latent dirichlet allocation. *Journal of machine Learning research*, 3: 993-1022. | **Bogdanov, D. et al. (2013).** Essentia: An audio analysis library for music information retrieval. In *ISMIR'13*, pp. 493-98. | **Choi, K., Lee, J. H., Hu, X., & Downie, J. S. (2016).** Music subject classification based on lyrics and user interpretations. In *Proceedings of the Association for Information Science and Technology*, 53(1):1-10. | **Czedik-Eysenberg, I., Knauf, D., & Reuter, C. (2017).** "Hardnessäs a semantic audio descriptor for music using automatic feature extraction. *INFORMATIK 2017.* | **Czedik-Eysenberg, I., Reuter, C., & Knauf, D. (2018).** Decoding the sound of "hardnessänd "darknessäs perceptual dimensions of music. *ICMPC-ESCOM: Book of Abstracts.* Graz, pp. 112-13. | **Czedik-Eysenberg, I., Wieczorek, O., & Reuter, C. (2019).** "Warriors of the Word Deciphering Lyrical Topics in Music and Their Connection to Audio Feature Dimensions Based on a Corpus of Over 100,000 Metal Songs. *arXiv* preprint arXiv:1911.04952. | **Czedik-Eysenberg, I.** (2020). *Music dimension ratings and factors dataset.* https://doi.org/10.17605/OSF.IO/3VG4E | **Hu, X., Downie, J. S., & Ehmann, A. F. (2009).** Lyric text mining in music mood classification. In *ISMIR'09*, pp. 159-68. | **Hutto, C.J. & Gilbert, E.E. (2014).** VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). | **Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., & Dempsey, E. (2014).** Textblob: simplified text processing. https://github.com/sloria/textblob | **Raschka, S. (2016).** Musicmood: Predicting the mood of music from song lyrics using machine learning. *arXiv* preprint arXiv:1611.00138.