

Die Verwendung hierarchischer linearer Modelle zur Audio-Feature-basierten Vorhersage kontinuierlicher emotionaler Reaktionen auf Musik

Christoph Graefe¹, Hauke Egermann²

¹ Technische Universität Berlin, Fachgebiet Audiokommunikation, Email: christoph.graefe@gmx.de

² Technische Universität Berlin, Fachgebiet Audiokommunikation, Email: hauke.egermann@tu-berlin.de

Einleitung

Bisherige Forschung konnte zeigen, dass Vorhersagen emotionaler Reaktionen auf Musik basierend auf linearen Regressionsmodellen und Zeitreihen psychoakustischer Audio-Features möglich sind [7]. Hierauf, auf Emotionstheorien [5, 6] und auf Ergebnisse von [1] aufbauend, möchte diese Studie die Entstehung musikalischer Emotionen weiter ausleuchten. Dazu wurde auf der Basis hierarchischer linearer Modelle (HLMs) untersucht, ob sich aus verschiedenen, nicht zwingend musikspezifischen psychologischen Mechanismen (*Brain Stem Reflex*, *Rhythmic Entrainment*, *Emotional Contagion* und *Musical Expectancy* [2]) Erklärungsansätze für den Einfluss von Low Level Features (*Spectral Centroid*, *Brightness*, *RMS*, *Roughness*), Mid Level Features (*Tempo*, *Melodic Contour*) und Higher Level Features (*Information Content* (IC) [4], *Unexpectedness* (UE)) auf emotionale Reaktionen ableiten lassen. Zusätzlich wurde untersucht, ob Low und Mid Level Features auf der einen und Higher Level Features auf der anderen Seite unabhängig voneinander Einfluss auf die emotionalen Reaktionen nehmen. Diese Reaktionen wurden als abhängige Variablen *Skin Conductance Response* (SCR), *Arousal* und *Valence* in die Analysen integriert. IC bezeichnet hier den mit Hilfe eines Computermodells berechneten Informationsgehalt der jeweiligen Notenwerte in Bit (aufgeteilt in High IC und Low IC), UE wurde als Gruppenmittelwert aus den kontinuierlichen Bewertungen von *Unexpectedness* innerhalb desselben Experiments berechnet (aufgeteilt in High UE und Low UE).

Methode

Während eines Live-Konzerts mit sechs Stücken klassischer Musik (S1-S6) für Solo-Querflöte und $N = 48$ Probanden wurde sowohl die periphere physiologische Erregung (SCR) gemessen als auch zeitkontinuierliche Bewertungen subjektiven Gefühls abgegeben (Valence und Arousal) [1]. Die Messungen wurden mit Hilfe eines eigens für die Studie entwickelten Messsystems vorgenommen, dessen Kern aus 48 Procomp Infiniti[®] Geräten für die physiologischen Messungen ($N = 48$) und 48 iPod touch[®] für die zeitkontinuierlichen subjektiven Bewertungen ($N = 24$ für Arousal/Valence, $N = 24$ für UE) bestand.

Extraktion der Features

Die oben genannten Low Level Features wurden mit Hilfe der MIRtoolbox 1.5 [3] extrahiert, beide Mid Level

Features hingegen wurden anhand von Audioaufnahmen des Konzerts und der Partituren manuell annotiert. Beide Higher Level Features wurden aus bestehenden Daten integriert (IC) bzw. berechnet (UE).

Statistische Analyse

Um die Einflüsse der Higher Level Features getrennt von allen anderen Features untersuchen zu können und Kollinearität der Prädiktoren zu eliminieren, wurde pro Stimulus eine Hauptkomponentenanalyse (PCA) mit allen o.g. Prädiktoren durchgeführt. Dabei wurde die Anzahl der Hauptkomponenten (PCs) schrittweise erhöht, bis die vier Higher Level Features jeweils in einer eigenen Komponente separiert waren; alle anderen Features fielen teilweise zusammen in eine Komponente oder waren ebenfalls einzeln separiert. Anschließend wurden 90 HLMs für wiederholte Messungen (repeated measurements) berechnet, in welche die in den PCAs entstandenen Hauptkomponenten als fixed effects und jeweils Arousal, Valence und SCR als abhängige Variablen einfließen konnten. Von jeder abhängigen Variable wurden zusätzlich



Abbildung 1: Links: iPod touch[®] Geräte zur Abgabe der kontinuierlichen Bewertungen; rechts: Procomp Infiniti[®] zur zeitkontinuierlichen Messung physiologischer Daten.

vier Duplikate erstellt, deren Werte um die Anzahl von Samples entsprechend 1, 2, 3 und 4 Sekunden vorgezogen wurden. Mit der sich ergebenden Struktur von Time Lags konnten verschiedene Reaktionszeiten der Probanden während der Bewertungen/Messungen in den Modellen berücksichtigt werden. Zur Verringerung von Autokorrelation wurde die Samplingfrequenz aller Zeitreihen zuvor auf $f_s = .5Hz$ reduziert. Darüber hinaus wurde zur Schätzung der Residuen eine autoregressive Kovarianzstruktur erster Ordnung mit heterogenen Varianzen (ARH1) angenommen. Diese ermöglicht es, Autokorrelation zu kontrollieren und somit das die zeitkontinuierlichen Messungen beeinflussende 'emotionale Gedächtnis'

mit zu modellieren [7].

Ergebnisse und Schlussfolgerungen

Es konnte gezeigt werden, dass Vorhersagen emotionaler Reaktionen auf Musik mit Hilfe der o.g. Prädiktoren erfolgreich möglich sind (vgl. beispielsweise Abb. 2). Von

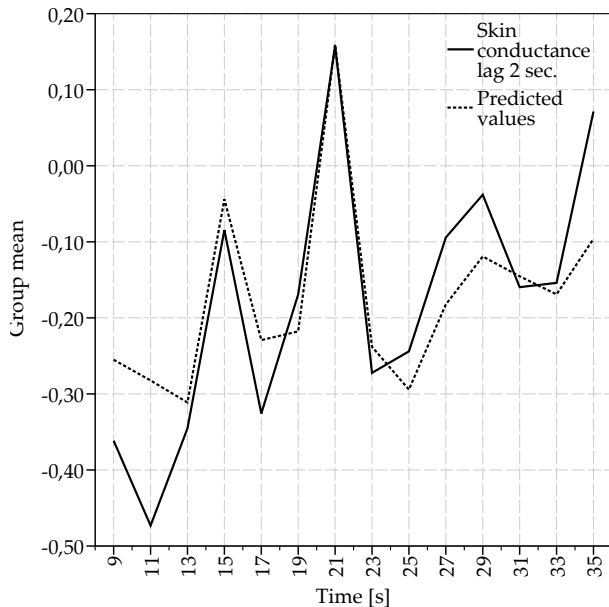


Abbildung 2: Plot der SCR Messungen des Stücks *Acht Stücke für Flöte Allein: VI. Lied, leicht bewegt* von Paul Hindemith (Time Lag 2 Sekunden) mit vorhergesagten Werten des HLMs über die Zeit.

den drei abhängigen Variablen liess sich SCR mit 42 signifikanten Koeffizienten quantitativ am Besten voraussagen (Arousal: 28 signifikante Koeffizienten, Valence: 11 signifikante Koeffizienten). Auf Ebene der Stimuli wies das Stück *Density 21.5* von Edgar Varèse aus dem Jahr 1936 mit 26 die meisten signifikanten Koeffizienten auf (neben weiterem Repertoire von J.S. Bach, Francis Poulenc, Paul Hindemith und Claude Debussy). Von den in

Tabelle 1: Anzahl der signifikanten Koeffizienten in den Ergebnissen der HLMs.

Stimulus	\sum signifikante Koeffizienten ($p < .01$)		
	Arousal	Valence	SCR
S1	4	4	5
S2	5	0	8
S3	14	1	11
S4	1	3	12
S5	3	3	0
S6	1	0	6

Tabelle 1 aufgeführten Summen der signifikanten Koeffizienten aller Modelle entfielen insgesamt 38 auf die Low und Mid Level Features sowie 43 auf die Higher Level Features. In der Gruppe der Higher Level Features wiederum entfielen 20 signifikante Koeffizienten auf IC und 23 auf UE. Dies bestätigt die Vermutung, dass die beiden Feature-Gruppen getrennt voneinander zu Vorhersagen

geeignet sind und dass die oben aufgeführten Mechanismen unabhängig voneinander wirken (beispielsweise *Musical Expectancy* mit IC/UE, *Rhythmic Entrainment* mit Tempo und *Brain Stem Reflex* mit RMS).

Literatur

- [1] Egermann, Hauke; M. T. Pearce; G. A. Wiggins; and Stephen McAdams (2013): “Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music.” In: *Cognitive, Affective, & Behavioral Neuroscience*. doi: 10.3758/s13415-013-0161-y.
- [2] Juslin, Patrik N.; Simon Liljeström; Daniel Västfjäll; and Lars-Olov Lundqvist (2010): “How Does Music Evoke Emotions? Exploring The Underlying Mechanisms.” In: Patrik N. Juslin and John A. Sloboda (Eds.) *Handbook of music and emotion (reprinted 2012)*, Series in affective science. Oxford: Oxford University Press, pp. 605–642.
- [3] Lartillot, Olivier and Petri Toiviainen (2007): “A MATLAB TOOLBOX FOR MUSICAL FEATURE EXTRACTION FROM AUDIO.” In: *Proceedings of the 10th International Conference on Digital Audio Effects (DAFX)*.
- [4] Pearce, M. T. (2005): “The Construction and Evaluation of statistical Models of melodic Structure in Music Perception and Composition: PhD Thesis.” URL <http://doc.gold.ac.uk/~textasciitildemas01mtp/papers/Pearce2005.pdf>.
- [5] Scherer, Klaus R. (2004): “Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?” In: *Journal of New Music Research*, **33**(3), pp. 239–251. doi:10.1080/0929821042000317822.
- [6] Scherer, Klaus R. (2005): “What are emotions? And how can they be measured?” In: *Social Science Information*, **44**(4), pp. 695–729. doi:10.1177/0539018405058216.
- [7] Schubert, Emery (2004): “Modeling Perceived Emotion With Continuous Musical Features.” In: *Music Perception*, **21**(4), pp. 561–585.