Der Einfluss von Audiofeatures auf das Blickverhalten

Jörg Mühlhans[1], Felix Klooss[1], Christoph Reuter[2]

[1] *Medialab, Universität Wien, 1090 Wien, E-Mail: joerg.muehlhans@univie.ac.at, felix.klooss@univie.ac.at*
[2] *Institut für Musikwissenschaft, Universität Wien, 1090 Wien, E-Mail: christoph.reuter@univie.ac.at*

Einleitung

(Hintergrund-)Musik kann das Blickverhalten beim Betrachten von Bildern und Videos verändern. Dabei haben vor allem Tempo und Ereignisdichte Einfluss auf die Dauer von Fixationen und Häufigkeit sowie Sprungweite der Sakkaden [1][2][3]. In Studien zum Tempo von Musik wird meist der angegebene Wert in beats per minute (BPM) herangezogen. Dieser Wert deckt sich nicht zwingend mit dem subjektiv wahrgenommenen Tempo, da es neben Verdopplungen oder Halbierungen auch zu Einflüssen anderer Eigenschaften wie der Notendichte, Anzahl harmonischer Wechsel oder melodischer Bewegung kommen kann. In einem 2020 am Medialab der Universität Wien durchgeführten bildschirmbasierten Experiment wurde der Einfluss von Musik auf Dauer und Verteilung von Fixationen untersucht [4]. Die Hypothese, schnelle Musik (160 BPM) führe zu kürzeren Fixationsdauern als langsame (90 BPM), konnte nur teilweise bestätigt werden. Allgemein zeigten sich jedoch Variationen in Dauer der Fixationen, Sakkaden- und Pupillenaktivität zwischen den Stimuli, die sich nicht einfach erklären ließen. Um musikalische Parameter für statistische Auswertungen zu quantifizieren, wurden in der vorliegenden Studie über 130 einzelne Audiofeatures des verwendeten Tonmaterials berechnet. Ziel dabei ist es, einen tieferen Einblick in die Einflussnahme einzelner Audiofeatures auf die unterschiedlichen messbaren Parameter des menschlichen Blickverhaltens zu erhalten.

Versuchsanordnung

In einem Laborsetting wurden die Versuchspersonen (n=44, f=25, A=37,1, SD=16,6) instruiert insgesamt 11 Videos mit einer durchschnittlichen Länge von 15 Sekunden anzusehen, während ihr Blickverhalten von einer Tobii Pro 2 Eyetracking-Brille mit einer Messfrequenz von 100 Hz aufgezeichnet wurde. Die Stimuli setzten sich aus 7 Videos musikalischer Performances aus den Bereichen Klassik/Jazz sowie 4 unbewegten Bildern mit Hintergrundmusik aus den Bereichen Jazz/Swing zusammen. Nach jedem Video folgte ein ausführlicher Fragebogen über die subjektive Wahrnehmung von Tempo, Expressivität, Professionalität, Lautheit, optischer Farbwahrnehmung und vielem mehr. Zusätzlich wurden soziodemografische Daten, allgemeine Genrepräferenzen und Persönlichkeitseigenschaften für weitere Analysen erhoben.

Die Versuchspersonen wurden zufällig einer von zwei Gruppen zugewiesen und bekamen unterschiedliche visuelle Manipulationen der Performance-Videos präsentiert sowie jeweils 2 der unbewegten Bilder mit Hintergrundmusik bei 90 BPM und zwei bei 160 BPM. Das musikalische Material war hierbei zwischen den Varianten jeweils exakt gleich, lediglich das Tempo wurde variiert.

Methode

Da sich die Stimulustypen (Videos vs. unbewegte Bilder) stark unterschieden, wurden ihre Tonspuren getrennt ausgewertet. Die Hintergrundmusik der unbewegten Bilder differieren nochmal im Tempo (90/160 BPM), sodass diese insgesamt 8 Tonspuren getrennt von den 7 der Videos untersucht wurden.

Um menschliche Interpretationsschwankungen innerhalb der Stimuli möglichst zu vermeiden, wurde die Hintergrundmusik der unbewegten Bilder mit virtuellen Instrumenten erstellt. Die mittels iReal Pro [5] erzeugten Kompositionen im Midi-Format wurden hierfür virtuellen Instrumenten zugeführt und das Tempo entsprechend variiert.

Für die Featuregewinnung und -analyse kamen Toolboxen für Matlab und Python zum Einsatz, darunter MIRToolbox [6], Essentia [7] oder LibROSA [8], bzw. daraus kombinierte Parameter [9]. Die über 130 Einzelfeatures wurden zur besseren Überschaubarkeit in drei Kategorien geteilt: spektrale, rhythmisch/temporale und harmonische Parameter.

Ergebnisse

Da es aufgrund der vielen Kombinationsmöglichkeiten und der verhältnismäßig geringen Anzahl an Datenpunkten zu einer großen Zahl signifikanter Ergebnisse kam, werden im Folgenden nur jene Ergebnisse angeführt, die mit einem Signifikanzniveau $p<.01$ belegt werden konnten.

Performance-Videos

In den 7 Performance-Videos finden sich ausschließlich Zusammenhänge der Features mit der durchschnittlichen Dauer von Blickfixationen. Es zeigen sich besonders Zusammenhänge mit Parametern, die die spektrale Form beschreiben, wie *MFCC2* (r=0,89, p<,01) und *MFCC4* (r=0,85, p<,01) oder *spectral_spread* (r=0,86, p<,01).

Überraschend ist die positive Korrelation mit der *event_density* (r=0,85, p<,01), was der initialen Hypothese widerspricht, da dieser Parameter eher in Zusammenhang mit hohem Tempo in der Musik steht. Andere Parameter, die bei Musikstücken mit höherem Tempo ebenfalls hohe Werte zeigen, lieferten jedoch keine signifikanten Ergebnisse (z.B. *onset_rate*, *BPM*, usw.). Nicht ausschließlich tempobezogen ist die *beat_loudness* (r=0,86, p<,01), bei der wiederum der positive Zusammenhang dadurch erklärbar ist, dass Beats in der Musik üblicherweise langsamer sind (120 BPM entspricht einem reinen Inter-Onset Interval (IOI) von 500 ms) als die durchschnittliche Fixationdauer im Experiment (ca. 280-370 ms). Ein stärker ausgeprägter Beat kann also auch bei höherem Tempo (durchaus auch 160 BPM) dazu führen, dass sich die Sakkaden mit dem Beat teilweise synchronisieren und sich dadurch die Fixationsdauern erhöhen. Eine Synchronisation mit der passendsten Halbierung hätte

demnach zur Folge, dass bei 90 BPM mit einem IOI von 667 ms auf 333 ms synchronisiert wird, da dieser Wert innerhalb der durchschnittlichen Fixationsdauer (280-370 ms) liegt. Hingegen ergibt sich bei 160 BPM ein IOI von 375 ms – eine Halbierung in diesem Fall hätte eine wesentlich höhere Abweichung zum Normalbereich zur Folge. Die naheliegendste Synchronisation ist demnach bei 120 BPM bei 333 ms, bei 160 BPM jedoch bei 375 und ist somit langsamer bei schnellerem Tempo.

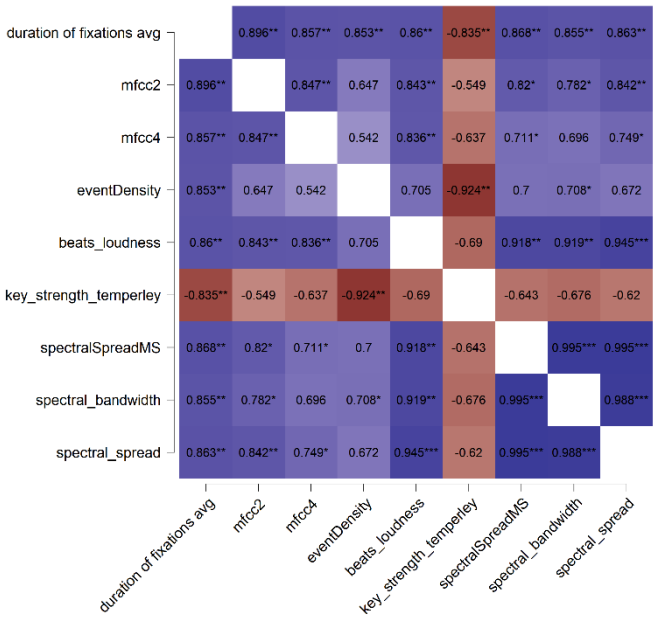

Abbildung 1: Heatmap für Parameter, die mit der Fixationsdauer korrelieren und auch untereinander einige Zusammenhänge aufweisen (* p<,05, ** p<,01, ***p<,001).

Bilder mit Hintergrundmusik

Bei den unbewegten Bildern zeigen sich kaum Ergebnisse für die Fixationsdauern, dafür umso mehr für die durchschnittliche Pupillengröße. Bei Fixationsdauern konnten lediglich schwächere negative Zusammenhänge (p<,05) für *event_density*, *onset_rate* und einige der höheren *MFCCs* beobachtet werden (r=0,71 bis 0,82).

Für die durchschnittliche Pupillengröße konnten die stärksten Korrelationen wiederum mit den helligkeitsbezogenen spektralen Parametern gefunden werden, z.B. *spectralCentroid* (r=-0,92, p<,01), *relativeEnergy0to100Hz* (r=0,84, p<,01), *brightnessHighFrequency* (r=-0,95, p<,001) oder *brightness_combination* (r=-0,94, p<,001). Hier liegt jedoch eine Konfundierung durch die visuellen Parameter des Bildmaterials vor, die weiter unten genauer beschrieben wird.

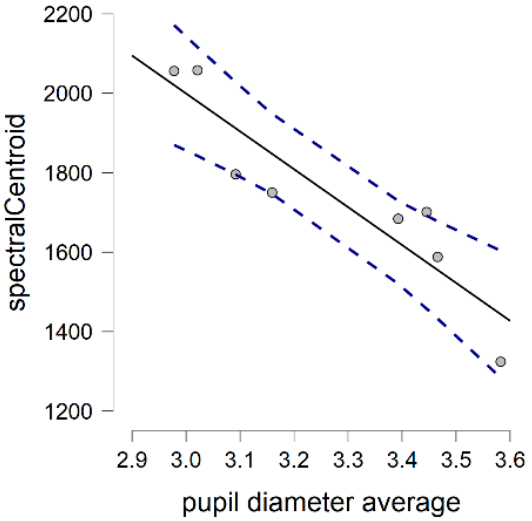

Abbildung 2: Scatterplot für die Korrelation zwischen Pupillengröße und *spectralCentroid* der 8 Audiotracks unbewegter Bilder mit 95%-Konfidenzintervall.

Ein weiterer interessanter Aspekt ist der augenscheinliche Zusammenhang mit der Angenehmheit von Klängen. Bei starkem Amplitudenanteil (mit ausgeprägter Tonhöhe) zwischen 2-4 kHz werden Klänge eher als unangenehm empfunden [10][11]. Dieser Frequenzbereich wird gut durch das Feature *2-4kHz_energy* (r=-0,875, p<,01) ausgedrückt.

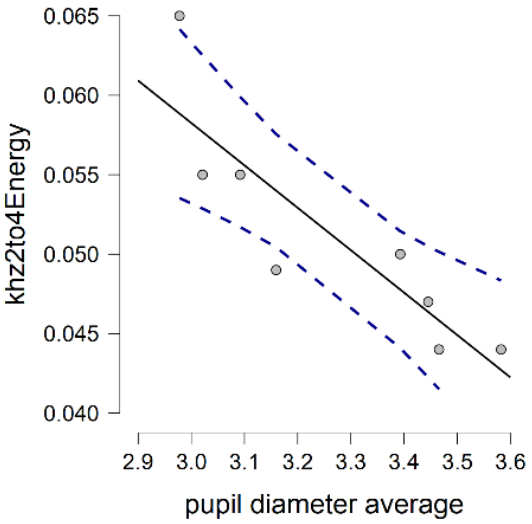

Abbildung 3: Scatterplot für die Korrelation zwischen Pupillengröße und spektraler Energie im Bereich zwischen 2-4 kHz der 8 Audiotracks unbewegter Bilder mit 95%-Konfidenzintervall.

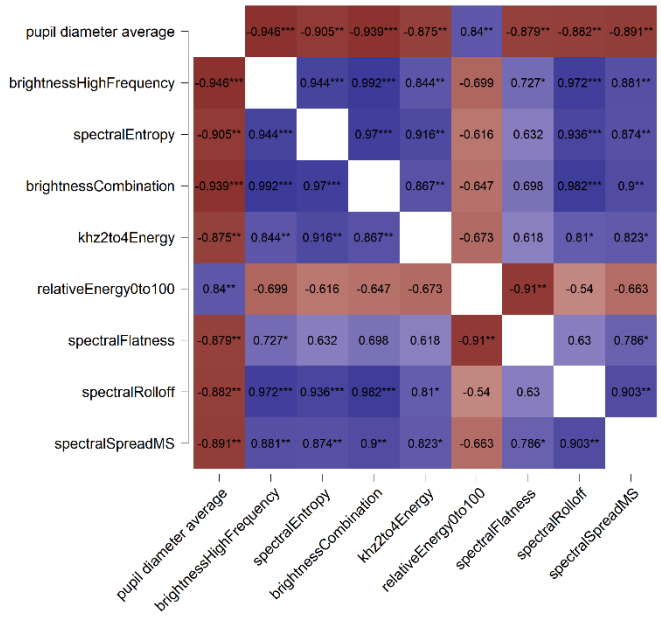

Abbildung 4: Heatmap für Parameter, die mit dem Pupillendurchmesser korrelieren und auch untereinander einige Zusammenhänge aufweisen (* p<,05, ** p<,01, ***p<,001).

Einschränkungen

Bewegung in Videos

Beim Videomaterial wurde auf eine möglichst unbewegte Szenerie geachtet. Es handelt sich ausschließlich um Stativaufnahmen ohne Zooms oder Schwenks, jedoch bewegen sich selbstverständlich die Musiker*innen beim Spielen ihrer Instrumente. Auch die Informationsdichte ist zwischen den Videos aufgrund der Anzahl sichtbarer Personen und Objekte sehr unterschiedlich. Diesen Einfluss zu messen ist schwierig, da bewegte Objekte, kontrastierende Farben und Formen sowie unerwartete Informationen das Blickverhalten unbewusst lenken [12]. Die Gewichtung einzelner visueller Faktoren ist dabei nicht bekannt, somit ist auch schwer zu beurteilen, wie stark der Einfluss akustischer Komponenten auf das Blickverhalten ist.

Helligkeit der Stimuli

Da der Bildschirm eine aktive Lichtquelle darstellt und die physiologische Anpassung des Lichteinfalls auf die Netzhaut über die Regelung der Pupillenweite (biologischer Regelkreis) gesteuert wird, existiert ein starker Zusammenhang zwischen der optischen Helligkeit des dargebotenen Stimulus und der Pupillengröße.

Bei gleichbleibendem Umgebungslicht emittiert der Bildschirm bei höherem Weißanteil mehr Licht. Die Helligkeit der Stimuli wurde über den durchschnittlichen Grauwert mit ImageJ berechnet [13]. Dadurch konnte dieser Zusammenhang auch im Versuch global über alle Stimuli nachgewiesen werden (r=-0,977, p<,001). Diese Tatsache konfundiert besonders den Zusammenhang zwischen der Pupillenweite und den helligkeitsbezogenen Audiofeatures. Dennoch kann festgehalten werden, dass im Paarvergleich zwischen den beiden Stimulusvarianten (90/160 BPM), bei denen jeweils der visuelle Stimulus identisch ist, sich trotzdem die auditiven Helligkeitsunterschiede konsistent in der Pupillengröße niederschlagen. Um die Stärke dieses Effekts näher zu untersuchen sind jedoch weitere Versuche nötig.

Diskussion

Die Schwierigkeit in der Untersuchung von Performance-Videos liegt im Auseinanderhalten der Einflüsse visueller und auditiver Komponenten auf die durch Eyetracking gemessenen Parameter. Es ist wohl unumstritten, dass der visuelle Einfluss dem auditiven übergeordnet ist, jedoch sollte die Rolle von (Hintergrund-)Musik nicht unterschätzt werden. Die Ergebnisse dieser Untersuchung stellen im Wesentlichen explorative Befunde dar, da die Erstellung der Stimuli primär im Hinblick auf die visuellen Manipulationen erfolgte und nicht auf die akustischen Parameter. Besonders bei den Videos ist die Rolle der Manipulationen innerhalb der vorliegenden Fragestellungen unklar. Dies gilt jedoch nicht für die unbewegten Bilder, da hier ausschließlich die Tonspur verändert wurde. Hier ist wiederum die Anzahl an Datenpunkten im Paarvergleich (schnell/langsam) gering, sie bieten jedoch einen guten Ansatz für Folgestudien.

Ausblick

In dieser Studie konnten bei männlichen VPN eine durchschnittlich 2 ms höhere Fixationsdauer festgestellt werden, was sich auch mit anderen Studien deckt [14]. Da diese jedoch innerhalb der zeitlichen Messungenauigkeit liegt, sollte für zukünftige Studien auf eine höhere Samplingrate zurückgegriffen werden. Diese ermöglicht auch die Betrachtung mikrosakkadischer Aktivitäten, welche durch Musikhören beeinflusst werden können [15].

In einer Folgestudie werden die Manipulationen auf akustische Parameter eingeschränkt und die Anzahl an Stimuli erhöht. Durch die Verwendung des stationären, bildschirmbasierten Eye-Trackers SR Research EyeLink 1000 können die genannten Einschränkungen durch eine sehr hohe Samplingfrequenz (1000 Hz) vermieden werden.

Besonderes Augenmerk wird auf den Zusammenhang zwischen Pupillengröße und akustischer Helligkeit unter Kontrolle der visuellen Helligkeit gelegt. Zudem wird der mögliche Zusammenhang zwischen Pupillen- (Durchmesser, Blinzelrate) und Blickdaten (Fixationen, (Mikro-)Sakkaden) und der Angenehmheit von Klängen untersucht.

Literatur

[1] Schäfer, T., Fachner, J.: Listening to music reduces eye movements. Atten. Percept. Psychophys., 77, 2015, 551–559

[2] Ross, V. et al.: Persuite Position Gain, Fixation Duration and Saccadic Gain with Music Intervention in Eye Motion Tracking. Science and Informatics Conference, 2015, 818–821

[3] Franek, M. et al.: Eye movement in scene perception while listening to slow and fast music. J. of Eye Movement Research, 11(2):8, 2018, 1–13

[4] Mühlhans, J., Klooss, F., Stowasser, C.: Der Einfluss von Musik auf Dauer und Verteilung von Blickfixationen. 36. Jahrestagung der DGM, 4.–6. September, 2020, Online

[5] iReal Pro Software, URL: https://www.irealpro.com/

[6] Lartillot, O., Toiviainen, P.: A Matlab Toolbox for Musical Feature Extraction from Audio. International Conference on Digital Audio Effects, Bordeaux, 2007

[7] Bogdanov, D. et al.: ESSENTIA: an Audio Analysis Library for Music Information Retrieval. ISMIR'13, 2013, 493–498

[8] McFee, B., et al.: librosa: Audio and Musik Signal Analysis in Python. Proc. 14. Python in Science Conf., 2015, 18–24

[9] Czedik-Eysenberg, I.: Music Information Retrieval und Klangfarbe. Universität Wien, 2016

[10] Reuter, C., Oehler, M.: Psychoacoustics of chalkboard squeaking. JASA 130(4), 2011, 2545

[11] Reuter, C., Oehler, M., Mühlhans, J.: Physiological and acoustical correlates of unpleasant sounds. Proc. ICMPC13-APSCOM5 Seoul, 2014, 97

[12] Bojko, A.: Eye Tracking the User Experience. Rosenfeld Media, New York, 2013

[13] Schneider, C. A., Rasband, W. S., & Eliceiri, K. W.: NIH Image to ImageJ: 25 years of image analysis. Nature Methods, 9(7), 2012, 671–675

[14] Sargezeh, B. A., Tavakoli, N., Daliri, M. R.: Gender-based eye movement differences in passive indoor picture viewing: An eye-tracking study. Physiology & Behavior 206, 2019, 43–50

[15] Lange, E. B., Zweck, F., Sinn, P.: Microsaccade-rate indicates absorption by music listening. Consciousness and Cognition 55, 2017, 59–78