

Spot

Chancen und Risiken des Computereinsatzes in der Musikpsychologie

Klaus Frieler & Elke B. Lange

Der Siegeszug des Computers ist nicht mehr aufzuhalten, auch in den Musikwissenschaften nicht. Wir warten nur noch auf den Moment, an dem die künstlichen Intelligenzen die Weltherrschaft übernehmen und vollkommen automatisch das letzte Konzert in der Elbphilharmonie diskutieren. Bevor es allerdings soweit ist, muss man beim Computereinsatz in der Wissenschaft immer noch viel manuelle Arbeit und Gehirnschmalz einsetzen. Man sollte besser genau wissen, was man tut, will man nicht gehobenen Unsinn produzieren. Auch heutzutage ist es schon so, dass kein musikpsychologisches Experiment ohne Computereinsatz auskommt. Sei es als Erhebungstool (z. B. Onlineumfragen oder Experimentaltoolboxen [PsychoPy]) oder einfach als Statistikknecht. Niemand rechnet heutzutage noch eine ANOVA mit der Hand (was aber als Übung dringend empfohlen sei!). Man verlässt sich auf die Korrektheit der Algorithmen. Als aber kürzlich ein Fehler in vielen bekannten MRI-Analyse-Tools entdeckt wurde (Eklund, Nichols & Knutsson, 2016), hat das viele veröffentlichte fMRI-Studien in Gefahr gebracht (die anfängliche geschätzte Zahl von 40.000 wurde mittlerweile deutlich nach unten korrigiert). Das zeigt aber ein grundlegendes Problem: Als Fachwissenschaftler muss man sich oft auf externe Werkzeuge wie Software verlassen, ist aber aufgrund deren Komplexität nicht mehr in der Lage, diese noch vollständig zu durchschauen. Allein schon die Anwendung ist von zahlreichen Fallstricken, Undurchschaubarkeiten und Gut-Glück-Heuristiken geprägt. Als Fachwissenschaftler ist man dennoch auf die Verfügbarkeit zuverlässiger und getesteter Tools angewiesen, die letztlich in Zusammenarbeit mit Informatikern, Technikern, Statistikern u. a. entstehen sollten, am besten mithilfe von Experten, die beide Domänen adäquat beherrschen. Die anderen müssen allerdings immer zumindest ein Grundverständnis der benutzten Prinzipien haben, um die Tools vernünftig einsetzen zu können. Da die Werkzeuge beständig erweitert werden und immer neue Methoden hinzukommen, bedeutet dies einen permanenten Fortbildungsdruck auf die in der Regel sowieso schon überarbeiteten Fachwissenschaftler.

Ein Beispiel für den Bereich der Musikpsychologie sind die Methoden, die in den letzten 10 bis 15 Jahren im Rahmen des Music Information Retrievals (MIR) entwickelt wurden und weiter werden und die vielversprechende Ansätze und spannende Möglichkeiten für die Musikwissenschaft bieten. So hat die Software *Tony* (oder auch das ältere *Praat*) die Möglichkeit eröffnet, relativ

bequem und sehr präzise Intonation, Tonart und Mikrotiming beim Singen zu untersuchen (Frieler et al., 2013; Mauch, Frieler & Dixon, 2014; Fischinger, Frieler & Louhivuori, 2015). Gerade für Produktionsparadigmen, in denen Versuchspersonen aktiv etwas Musikalisches hervorbringen, was immer einen sehr hohen Analyseaufwand bedeutet, sind die neuen Tools Gold wert. Aber auch in klassischen Hörexperimenten erlauben diese Werkzeuge neue Zugänge und präzisere Messungen. So wurde z. B. die MIRtoolbox (Lartillot & Toivainen, 2007) genutzt, um musikalische Überraschungen in der Audioaufnahme eines Tango Nuevo zu finden und diese dann zeitlich mit elektrischen Potenzialen im EEG in Beziehung zu setzen (Poikonen et al., 2016). Dies stellt eine innovative Erweiterung des herkömmlichen neurowissenschaftlichen Paradigmas von Erwartungsverletzung bei einfachen Tonsequenzen („mismatch negativity“; vgl. Näätänen et al., 2007) oder Akkordfolgen (Koelsch et al., 2001) auf real existierende Musik dar.

Grundsätzliche Aufgabe des MIR ist die automatische Generierung inhaltlicher Beschreibungen (Metadaten, Deskriptoren) aus musikalischem Material (z. B. Audio- und MIDI-Dateien, aber auch digitalisierte Notentexte). Der grundlegende Vorteil gegenüber der herkömmlichen Musikanalyse liegt neben der Automatisierung und Objektivierung der Prozesse und den viel größeren Korpora, die so untersucht werden können, vor allem darin, dass nun auch die erklingende Musik selbst in vollem Umfang der Analyse zugänglich wird, seien es verschiedene Interpretationen einer in Noten aufgezeichneten Komposition oder Stücke, denen nur ein rudimentärer oder gar kein Notentext zugrunde liegt (wie Popsongs oder Stücke aus oralen Musikkulturen). Damit rücken primär akustische und klangliche Aspekte in den Blick, die sich der textlichen Fixierung entziehen und daher bislang kaum je ein Gegenstand von Musikanalyse waren. Die ursprüngliche Motivation des MIR war, Musik im Internet such- und auffindbar zu machen. Eine klassische und vielbeforschte Aufgabe ist demzufolge auch die Bestimmung des musikalischen Stils direkt aus der Audioaufnahme (im musikwissenschaftlichen Bereich siehe z. B. Eerola, 2011). Hat man ein solches System, kann man etwa beim Internetverkauf von Musik den Kunden eine Stilbeschreibung an die Hand geben, ohne dass Menschen aus Fleisch und Blut diese hinzufügen müssen. Mittlerweile hat sich aber gezeigt, dass die herkömmlichen Stilbezeichnungen so unscharf und uneinheitlich sind, dass die eingesetzten Algorithmen, zumeist statistische Klassifikationsverfahren, mit dieser sogenannten „Ground Truth“ nur bedingt weit kommen, und nur bei begrenzten Datensätzen eine Klassifikationsgenauigkeit von max. um die 80 % erzielen (Sturm, 2014).

Eine weitere klassische Aufgabe des MIR, die auch direkt in das wichtige Feld der musikpsychologischen Emotionsforschung reinspielt, ist die automatische Bestimmung des emotionalen Gehalts, Ausdrucks oder Stimmung („Mood“) einer Musikaufnahme (z. B. Coutinho & Cangelosi, 2011; Friberg et al., 2014; Leman et al., 2005). Dabei geht es darum, mithilfe einer großen Menge beschreibender Eigenschaften (sog. „Features“) diejenigen herauszufiltern, die den wahrgenommenen Emotionsgehalt (oder auch tatsächlich beim Hörer evozierte Emotionen) möglichst gut und generalisierbar modellie-

ren. Kann man diese Modelle auf Basis der Deskriptoren auch inhaltlich interpretieren, so hat man damit im Grunde ein psychologisches Modell geschaffen, auch wenn dies nicht die Zielsetzung war. Der Ansatz, aus einer Menge extrahierter Features mithilfe von statistischen Modellen Perzepte und andere Hörerverhaltensweisen zu Musik zu modellieren (z.B. motorische Reaktionen, Beatinduktion, Präferenz, therapeutische Wirkung), ist die natürliche Schnittmenge zwischen MIR und Musikpsychologie, auch wenn sie unter sehr verschiedenen fachdisziplinären Überschriften laufen, was die Kommunikation zwischen beteiligten Forschern oft erschwert. Dies führt manchmal auch zu blinden Flecken und „naiven“ Ansätzen in den jeweiligen Zugängen (Aucouturier & Bigand, 2013), was die Notwendigkeit einer Zusammenarbeit nahelegt. Wie die Erfolge und Misserfolge des MIR und der Musikpsychologie in der Vorhersagekraft ihrer Modelle gezeigt haben, sind die Aufgaben nicht trivial. Das fängt bereits bei der Auswahl und Konstruktion aussagekräftiger und sinnvoller Features an und hört bei der komplexen Struktur der zu modellierenden Verhaltensweisen auf. Letztere sind oft sehr breit streuend, da persönliche Hör- und Musikerfahrung, Kontextsituationen und momentane Verfasstheiten immer auch eine Rolle spielen. Diese müssten günstigenfalls mit modelliert werden, was aber die Aufgabe erheblich kompliziert. Will man eine große Zahl von Personen jenseits von Kontext und Persönlichkeit modellieren, kann man zwar im Mittel die richtige Vorhersage treffen, was aber in der Regel über den Einzelfall oder auch eine große Zahl von Fällen wenig aussagt, wenn die Streuung hoch ist, wie etwa bei emotionalen Einschätzungen. Die Populationsmittelwerte sagen dann vergleichsweise wenig über den Einzelfall aus, was je nach Experimentaldesign und zugrunde liegender Forschungsfrage unterschiedliche Konsequenzen hat.

Wie wir kürzlich in einer Studie gezeigt haben (Lange & Frieler, 2017), werden bereits die einfachsten emotionalen Ausdrucksgehalte von Musik sehr unterschiedlich eingeschätzt, d. h. mit hoher Varianz und oft sogar mit bimodalen Verteilungen (vgl. auch Schedl et al., 2016). Das stellt große Herausforderungen an die statistische Modellierung dar. Darüber hinaus zeigte sich, dass sich die Mittelwerte der emotionalen Einschätzungen zwar recht gut mit linearen Modellen von Audiofeatures aus der MIRtoolbox (Lartillot & Toiviainen, 2007) vorhersagen ließen, doch der Satz der Features in gewisser Hinsicht beliebig war. Der Grund dafür war, dass die Mittelwerte der emotionalen Variablen eine viel geringere Streuung über die Stücke aufweisen als die Streuung der Bewertung der Einzelstücke. In der Konsequenz lieferten Nullmodelle (d. h. die Mittelwerte der Mittelwerte) vielfach eine recht gute Vorhersage (interessante Ausnahme: „happy“). Die Hinzunahme einzelner Features brachte dann zwar Verbesserungen, aber aufgrund der zahlreichen sehr hohen Korrelationen der Features untereinander, war es in gewisser Hinsicht beliebig, welches Feature man dafür nahm. Das untergräbt die Interpretierbarkeit, auch wenn es durchaus starke Einzelkorrelationen zwischen Features und emotionalen Variablen gibt. Die Streuung der Bewertungen ließ sich hingegen nicht durch Features erklären, was auf persönliche und messtechnisch bedingte Unterschiede bei der Einschätzung hindeutet. So ist der epistemologische Status derartiger Modelle eher un-

klar. Sie haben aber einen praktischen Wert, denn die Modelle machen brauchbare Vorhersagen – man weiß nur nicht genau, wieso. Wenn man zeigen kann, dass derlei Modelle gut generalisieren, kann man sie aber zur Auswahl und Beschreibung von Stimuli in weiteren Experimenten gut einsetzen.

Es gilt also auch hier wie überall: Wenn man weiß, was man tut, ist alles gut. Dann können die modernen MIR-Verfahren spannende neue Möglichkeiten zur Analyse und Modellierung bieten und vor allem helfen, musikalische Stimuli objektiver und präziser zu beschreiben (und auszuwählen), was ein oft vernachlässigter Punkt im musikpsychologischen Experiment ist. Auch für Produktionsparadigmen tun sich erhebliche Erleichterungen in der Analyse auf. Man sollte die sich bietenden Chancen für die musikpsychologische Forschung weiter evaluieren und nutzen, etwa in Form von vertiefenden Studien über den Zusammenhang von Audiofeatures und Hörerwahrnehmungen. Im besten Fall wartet man als Musikwissenschaftler oder Musikpsychologin nicht einfach darauf, dass Computerwissenschaftler aufgrund ihrer technischen Interessen neue Feature oder Tools entwickeln, um dann zu schauen, wie man sie anwenden kann, sondern entwickelt mit ihnen gemeinsam etwas, geleitet von genuinen musikbezogenen Forschungsfragen.

Literatur

- Aucouturier, J.-J. & Bigand, E. (2013). Seven problems that keep MIR from attracting the interest of cognition and neuroscience. *Journal of Intelligent Information Systems*, 41, 483–497. <http://doi.org/10.1007/s10844-013-0251-x>
- Coutinho, E. & Cangelosi, A. (2011). Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion*, 11 (4), 921–937. <http://doi.org/10.1037/a0024700>
- Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research*, 40, 349–366. <http://doi.org/10.1080/09298215.2011.602195>
- Eklund, A., Nichols, T. E. & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113 (28), 7900–7905. <http://doi.org/10.1073/pnas.1602413113>
- Fischinger, T., Frieler, K. & Louhivuori, J. (2015). Influence of virtual room acoustics on choir singing. *Psychomusicology: Music, Mind, and Brain*, 25 (3), 208–218. <http://doi.org/10.1037/pmu0000117>
- Friberg, A., Schoonderwaldt, E., Hedblad, A., Fabiani, M. & Elowsson, A. (2014). Using listener-based perceptual features as intermediate representations in music information retrieval. *Journal of the Acoustical Society of America*, 136 (4), 1951–1963. <http://doi.org/10.1121/1.4892767>
- Frieler, K., Fischinger, T., Schlemmer, K., Jakubowski, K., Müllensiefen, D. & Lothwesen, K. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae, Special issue: Replication in music psychology*, 7 (3), 334–349.
- Koelsch, S., Gunter, T. C., Schröger, E., Tervaniemi, M., Sammler, D. & Friederici, A. (2001). Differentiating ERAN and MMN: An ERP study. *Neuroreport*, 12 (7), 1385–1389. <http://doi.org/10.1097/00001756-200105250-00019>

- Lange, E. B. & Frieler, K. (2017). *Challenges of analyzing music by subjective evaluations and automatized feature extraction: An application for emotion research*. Accepted for publication in *Music Perception*.
- Lartillot, O. & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. *Proceedings of the 10th International Conference on Digital Audio Effects* (DAFx-07). Bordeaux, FRA. Retrieved from <http://dafx.labri.fr/main/papers/p237.pdf>
- Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D. & Lesaffre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, 34 (1), 39–67. <http://doi.org/10.1080/09298210500123978>
- Mauch, M., Frieler, K. & Dixon, S. (2014). Intonation in unaccompanied singing: Accuracy, drift and a model of reference pitch memory. *Journal of the Acoustical Society of America*, 136 (1), 401–411. <http://doi.org/10.1121/1.4881915>
- Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544–2590.
- Poikonen, H., Alluri, V., Brattico, E., Lartillot, O., Tervaniemi, M. & Huotilainen, M. (2016). Event-related brain responses while listening to entire pieces of music. *Neuroscience*, 312, 58–73. <http://doi.org/10.1016/j.neuroscience.2015.10.061>
- Schedl, M., Eghbal-Zadeh, H., Gomez, E. & Tkalcic, M. (2016). An analysis of agreement in classical music perception and its relationship to listener characteristics. *Proceedings of the 17th International Society for Music Information Retrieval Conference*, New York, 578–583. Retrieved from http://m.mr-pc.org/ismir16/website/articles/260_Paper.pdf
- Sturm, B. L. (2014). A simple method to determine if a music information retrieval system is a „horse“. *IEEE Transactions on Multimedia*, 16 (6), 1636–1644. <http://doi.org/10.1109/TMM.2014.2330697>