

CODEX FLORES

10.01.2006

Wie aus Zahlen Stilmerkmale werden



Gerhard Widmer ist Leiter des Institutes for Computational Perception der Johannes Kepler Universität Linz. (Bilder: Codex flores)

Der an der Johannes Kepler Universität in Linz tätige Computerwissenschaftler Gerhard Widmer hat mit empirischen Untersuchungen zu Interpretationen von Pianisten eine Brücke zwischen statistischen Verfahren der Künstlichen Intelligenz und der Musikanalyse geschlagen. Codex flores hat sich mit dem originellen Forscher über die Methoden unterhalten, die solchen Experimenten zugrundeliegen.

Codex flores: *Bei Ihren Untersuchungen zum individuellen Stil von Pianisten greifen sie auf eine Technik zurück, die Informatiker unter dem Schlagwort «Datamining» vertraut ist. Können Sie in ein paar Worten erklären, was man sich darunter vorzustellen hat?*

Gerhard Widmer: Die Technik geht auf Forschungen im Bereich des maschinellen Lernens zurück. Man fragte sich, wie man Computern beibringen könnte, aus vielen Einzelbeobachtungen allgemeine Schlüsse zu ziehen. Lernen lässt sich in vielen Fällen darauf reduzieren, dass in einer bestimmten Menge von Daten allgemeine Zusammenhänge und Gesetzmässigkeiten entdeckt werden. Später hat man gesehen, dass man damit verbundene Methoden auch auf sehr grosse Datenmengen anwenden kann.

Das heisst, man findet Dinge, an die man gar nicht gedacht hätte?

Das ist ein Aspekt: Man kann den Computer quasi selbständig auf Entdeckungsreise schicken. Man geht nicht mit vorgefassten Theorien hinter die Daten und versucht, darin das zu entdecken, was man sowieso schon vermutet. Der Rechner soll vielmehr selber Zusammenhänge zwischen Variablen finden, die keiner sonst aufgespürt hätte.

Kann man solche Schlüsse nicht ohne Hilfe des Rechners ziehen?

Eben nicht. Die Anzahl möglicher statistischer Zusammenhänge in einer grossen Datenmenge ist so exorbitant gross, dass es selbst für den Rechner schlicht und ergreifend unmöglich ist, alle möglichen Variablenkombinationen durchzuprüfen. Er würde dazu schnell einmal mehr Zeit brauchen, als das Universum alt ist. Der Computer stellt

gewisse Ausgangsvermutungen an und schaut sich nur einen Bruchteil der möglichen Kombinationen an.

Wie nutzen Sie solche Techniken für die Musikforschung?

Wir haben uns gefragt wie das, was ein klassischer Musiker spielt und tut, mit der Struktur der Musik selber zusammenhängt. Gibt es da Regeln, die sich natürlicherweise ergeben? Wir haben den Rechner nach Regeln suchen lassen, welche die Unterschiede zwischen dem tatsächlichen Spiel eines Pianisten und dem, was in den Noten steht, erfassen.

Kann man das an einem Beispiel erläutern?

Wir haben etwa versucht, in Aufnahmen von Mozartsonaten, die wir sehr präzise gemessen haben, Regelmässigkeiten zu entdecken, aufgrund deren sich vorherzusagen lässt, an welcher Stelle in einem Stück ein Interpret langsamer oder schneller wird oder eine Note betont wird oder nicht.

Angenommen der Computer soll versuchen vorherzusagen, an welcher Stelle der Pianist lokal langsamer werden wird. Der Rechner sucht in der Interpretation alle Stellen heraus, an denen der Pianist sein Spiel verzögert. Zum Vergleich sucht er nach allen Situationen, in denen keine Verzögerung im Spiel festzustellen ist. Aufgrund dieser beiden Ereignisklassen versucht der Computer herauszufinden, welche Eigenschaften der Musik die beiden Situationen charakterisieren.

Sie stellen die Tempo- und Lautstärkeschwankungen der Interpretation in einem zweidimensionalen Koordinatensystem dar. Daraus lässt sich auf einen Blick grafisch überblicken, wie ein Pianist eine Passage gestaltet: Die Korrelation der beiden Parameter bewegt sich wie ein Würmchen über die Fläche.

Das basiert auf einer Idee und Grundmethode des deutschen Musikwissenschaftlers Jörg Langner. Diese simultane Darstellung von Tempo und Dynamik finden wir ausgesprochen intuitiv und übersichtlich, und sie eignet sich auch gut für didaktische Zwecke.

Andererseits muss man schon auch ein paar kritische Worte dazu machen. Diese Würmer («Performance Worms», wie wir sie nennen), sehen schön aus, es handelt sich aber nur um eine visuelle Metapher, die mehr oder weniger überzeugen kann, und die man auch überinterpretieren kann. Die Darstellung ist nur so gut wie die zugrundeliegenden Daten. Und das sind in den meisten Fällen – vor allem dort, wo wir Tempo und Dynamik aus Tonaufnahmen berühmter Pianisten herausgemessen haben – nur sehr ungenaue und abstrakte Daten.

Aus Audiodaten können wir Tempo und Dynamik nur in bestimmten Intervallen bestimmen, das sind keine kontinuierlichen Aufzeichnungen. Wir messen nur gerade bei den Taktschlägen, oder vielleicht eine metrische Ebene darunter, wenn es geht. Diese werden in der Aufnahme markiert.



«Im Endeffekt ist jede Form des Lernens und Datenanalysierens eine Frage der Statistik.»

Und Sie erstellen daraus eine simple Statistik?

Im Endeffekt ist jede Form des Lernens und Datenanalysierens eine Frage der Statistik. Es geht darum zu zählen, wie oft bestimmte Dinge mit andern Dingen vorkommen, um herauszufinden, welche Aspekte einer Situation zu einer bestimmten Konsequenz führen und welche Aspekte zu einer andern Konsequenz führen.

Die Ausgangsvermutungen, die man zugrundelegt, sind natürlich auch statistisch orientiert. Der Computer hat ja kein wirkliches Verständnis für die Musik. Für ihn sind alle Variablen gleich wichtig. Er weiss nicht, welche musikalisch mehr Sinn machen oder weniger, welche eher einen Einfluss auf eine Tempoentscheidung eines Pianisten haben.

Man muss sich also auch gut überlegen, ob das, was der Computer gefunden hat und vielleicht statistisch durchaus signifikant ist, auch musikalisch Sinn ergibt. Es kann ja durchaus sein, dass der Rechner in der Interpretation von Horowitz statistisch signifikante Eigenheiten ausfindig macht, die musikalisch völlig uninteressant sind, nebensächliche Artefakte, die so minim sind, dass man sie beim Hören gar nicht wahrnimmt.

Möglicherweise liesse sich ja auch aus musikalisch wenig signifikanten Korrelationen etwas auslesen? Zum Beispiel etwas wie einen unverwechselbaren «Fingerabdruck» eines Pianisten?

Das haben wir auch getan. Wir haben den Computer mit Daten zu Aufnahmen von Mozartsonaten von sechs verschiedenen Pianisten gefüttert. Dann haben wir ihm Aufnahmen anderer Stücke – auch von Mozart – eingegeben und geschaut, ob er entscheiden kann, welcher Pianist spielt. Da haben wir Resultate erhalten, die weit über dem blinden Raten sind.

Das Verfahren erinnert ein wenig an die Pionierarbeiten des Pianisten und Mathematikers Manfred Clynes. Der hat ja schon relativ früh behauptet, er könne mit ähnlichen Methoden etwas wie einen charakteristischen «Puls» ermitteln, der als eine Art Signatur für einen Komponisten gesehen werden kann.

Das mit den «ähnlichen Methoden» würde ich so nicht stehen lassen. Wir machen Aussagen über konkret gemessene Abweichungen eines Pianisten von einem fixen Tempo oder einer konstanten Lautstärke – Dinge, die direkt und relativ objektiv in der Aufnahme messbar sind.

Clynes postuliert zum Teil Dinge, die eben nicht messbar sind, Dinge, für die man nur indirekte Evidenz vorbringen kann, und die auch in zahlreichen Experimenten anderer Forscher nicht in der Form reproduziert werden konnten. Einige von Clynes' Behauptungen und Theorien werden daher in der Literatur äusserst kontrovers gesehen.

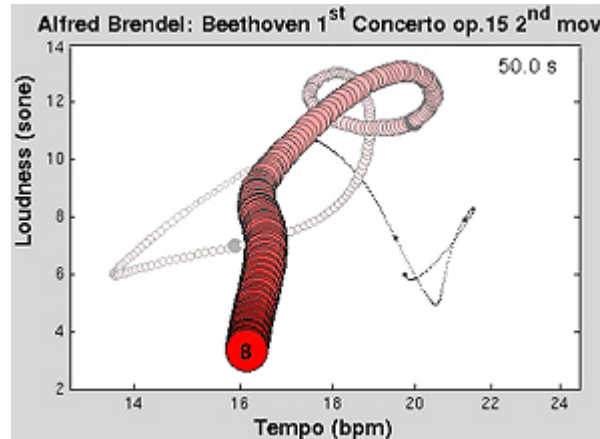
Er hat aber mit einem gewissen Recht auch darauf hingewiesen, dass man beim Hören in kürzester Zeit sagen kann, dass ein Interpret etwas zu sagen hat oder eben nicht. Das zeigt doch, dass musikalischer Ausdruck auf der einen Seite sehr komplex scheint, auf der andern wiederum sehr einfach und mechanisch produzierbar. Einfach ja, mechanisch produzierbar nein. Damit hat man es in der Künstlichen Intelligenz jeden Tag zu tun. Genau die Probleme, die Menschen intuitiv und schnell lösen, sind in der Regel äusserst schwer zu formalisieren. Das hat damit zu tun, dass sich der Mensch ein Leben lang in einer realen Welt bewegt und enorm komplexe Interpretationen von Wahrnehmungen mit grösster Geschwindigkeit vollzieht. Unser Gehirn verbindet alles mit allem in unglaublichem Tempo. Es macht deshalb ständig Vorhersagen, was als nächstes plausiblerweise zu erwarten ist. Das geht so weit, dass wir meinen, Dinge zu sehen, die gar nicht da sind, weil wir überzeugt sind, dass sie nun kommen müssten.

Was heisst das wiederum für die Musik?

Nehmen wir zum Beispiel ein Programm, das die Melodie aus einem polyphonen Geschehen extrahieren soll, etwas was ein Mensch mit grösster Leichtigkeit erledigt. Der klassische Ansatz besteht darin, dass man aus dem Audiosignal physikalische Ereignisse herausmisst und mit Hilfe dieser einen vom Einzelereignis zur globalen Sicht

fortschreitenden Prozess, einen sogenannten «Bottom-Up-Algorithmus» baut. Dieser soll dann entscheiden, wo die Melodie ist.

Experimente zeigen aber, dass derartige Methoden derzeit noch weit von der Leistungsfähigkeit von Menschen entfernt sind. Und ein wesentlicher Grund dafür ist eben, dass sie nur von lokaler Information ausgehen und keine Vorstellung vom grösseren musikalischen Kontext haben - keine Erfahrung mit Musik, keine Erwartungshaltungen, die über jahrzehntelanges Hören von Musik geformt wurden. Genau das aber sind Dinge, die auch sehr schwer zu modellieren sind.



Die Resultate von Widmers Untersuchungen lassen sich als Würmer in einem Koordinatensystem mit den Achsen Lautstärke und Tempo darstellen. Spielt ein Pianist etwa immer schneller und zugleich leiser, bewegt es sich von links oben nach rechts unten. Je älter die Spur, umso mehr verblasst sie. Die Darstellungsmethode findet sich zuerst in: Langner J., Goebel W.: «Visualizing expressive performance in tempo-loudness space.» *Computer Music Journal*, 27(4), 69-83, 2003.

Die beiden Amerikaner Fred Lerdahl und Ray Jackendoff haben in ihrem epochalen Buch «Generative Theory of Tonal Music» in den achtziger Jahren gezeigt, dass automatische Melodieanalysen sehr schnell sehr komplex werden.

In dem Buch werden sowohl Bottom-Up- als auch ihre Umkehrung, die Top-Down-Aspekte betont. Nimmt man die Theorie etwas genauer unter die Lupe, stellt man fest, dass die Bottom-Up-Aspekte sehr gut formalisiert sind. Bei den Top-Down-Aspekten wird die ganze Sache hingegen viel vager. Da greifen die beiden auf sprachliche Umschreibungen zurück.

Können Sie ein Beispiel machen?

Sie sagen etwa, dass beim Erkennen von Gliederungen einer Melodie die Wahrnehmung von Ähnlichkeiten eine Entscheidung verstärken kann, dass zwei Bereiche Segmente sind. Wie das Erkennen der angesprochenen Ähnlichkeiten auf dieser hohen Ebene funktioniert, das können sie dann auch nicht mehr beschreiben. Da bricht auch ihre Theorie zusammen oder muss sich auf allgemeine Formulierungen zurückziehen.

Die ersten Kritiken an der Theorie wiesen auch darauf hin, dass zirkuläre Strukturen beschrieben werden: gewisse Regeln hängen gegenseitig voneinander ab. Eine weitere Schwierigkeit bietet für die exakte Analyse die Einführung von Präferenzregeln. Sie geben keine eindeutigen Anweisungen, sondern bloss Hinweise, welche Strukturierung vorzuziehen ist, wenn gewisse Sachverhalte in einem bestimmten Mass beobachtet werden. Dazu müssen also noch Schwellwerte in Form quantitativer Gewichtungen angegeben werden, damit man sie überhaupt anwenden kann.

Es muss allerdings nicht die Aufgabe des Theoretikers sein, konkrete

Schwellwerte anzugeben. Jeder Mensch hört halt doch wieder ein bisschen anders. Wie die Grundwerte gegeneinander gewichtet sind, kann man auch offen lassen. Das ist legitim.

Oder man versucht, die Werte experimentell zu ermitteln.

Das wird dann aber schon wieder sehr schwierig. Wenn Experimente da gültige Resultate liefern sollen, dann muss man sie sehr kontrolliert durchführen. Das wiederum heisst, dass man mit sehr künstlichen musikalischen Stimuli arbeiten muss. Wenn man zuviel Kontext zulässt, dann weiss man nicht mehr, was wovon beeinflusst wird.

Die amerikanische Musikpsychologin Carol Krumhansl und andere haben gewisse Befunde der Theorie experimentell nachvollzogen.

In ihren Experimenten sind isolierte Faktoren so verwendet worden, dass unbeobachtete Effekte ausgeschlossen werden konnten. Nur kann man sich dann fragen, ob das noch etwas mit Musik zu tun hat. Das ist wahrscheinlich ein fundamentales Dilemma: Wenn man präzise messen will, muss man ein Experiment so einschränken, dass man zerstört, was eigentlich studiert werden sollte. Es wird dann nicht mehr viel Relevantes über Musik ausgesagt.



«Auch wenn ich in einem empirisch orientierten Fach tätig bin, habe ich durchaus Verständnis dafür, dass man den empirischen Methoden gegenüber eine gewisse Skepsis entgegenbringen kann.»

Viele derartige experimentelle Resultate muten zugegebenermassen trivial an. Da kann man sich fragen, ob die Idee dahinter tatsächlich ist, einen echten Erkenntnisgewinn zu erreichen, oder ob es da nicht im Sinn von Vorarbeiten einfach darum geht, die Möglichkeiten und Grenzen des Experimentalapparates für die Musikpsychologie auszuleuchten.

Auch banale Resultate haben ihren Sinn. Man kann sogenannt banale Erkenntnisse auch einfach einmal abhaken, indem man sie experimentell überprüft. Vielleicht steckt da nicht mehr dahinter. Man weiss dann, dass die Methode an einem konkreten Problem funktioniert. Ob sie dann auch ausreicht, um komplexere musikpsychologische Fragen anzugehen, bleibt aber immer noch offen.

Geisteswissenschaftler entschuldigen ihre Ignoranz gegenüber experimentellen Methoden häufig mit der angesprochenen scheinbaren Banalität der gegenwärtigen Ergebnisse. Man hat aber häufig den Eindruck, dass sie nicht wirklich verstehen, was Experimentalpsychologen überhaupt machen, was es bedeutet, ein Experiment zu designen und durchzuführen, und welche Leistungen da erbracht werden.

Wie schwierig es ist, Experimente korrekt zu designen und gleichzeitig aussagekräftig zu halten, weiss wohl nur, wer sich wirklich intensiv damit befasst hat. Es mag schon sein, dass manche Geisteswissenschaftler sich Experimentieren zu einfach vorstellen.

Andererseits: Auch wenn ich in einem empirisch orientierten Fach tätig bin, habe ich durchaus Verständnis dafür, dass man den empirischen Methoden gegenüber eine gewisse Skepsis entgegenbringen kann. Sie bringen Einschränkungen mit sich, die vielleicht nicht immer zu einem

vollen Verständnis eines Phänomens beitragen.

Auf der andern Seite werden Experimentalpsychologen halt auch Opfer ihrer eigenen Marketinganstrengungen. Sie trivialisieren ihre Ergebnisse häufig selber. Das Paradebeispiel ist der Mozart-Effekt, bei dem aufgrund eines simplen Experimentes medienwirksam irgendwelche globalen Behauptungen über den Intelligenzsteigernden Wert von Mozartmusik gemacht worden sind.

War da nicht schon das Experiment an und für sich verfehlt?

Man hat das Experiment auch mit Ratten durchgeführt und behauptet, diese fänden sich in einem Labyrinth besser zurecht, wenn sie zuvor Musik von Mozart gehört hätten. Ein findiger Kritiker hat aber gezeigt, dass die Ratten ja eigentlich taub sind und die Musik gar nicht hören können. Da wird das Ganze zur Realsatire.

Wenn man das Experiment mit zwei Gruppen macht, und die eine hört Mozart, bevor sie die Aufgaben lösen muss, dann wundert es mich auch nicht, dass die das besser hinbringt. Vielleicht sind ihre Mitglieder einfach konzentrierter. Die andern sind vielleicht einfach eingeschlafen. Es gibt für so etwas ganz banale Erklärungen. Die Journalisten interessieren sich häufig auch nicht für die tatsächlichen Verhältnisse. «Mozart-Effekt» ist ein wunderbares Schlagwort.

Da wird in der Wissenschaftsberichterstattung sicher viel gesündigt. Seit Jahren geistert in der deutschsprachigen Presse ein Gerücht umher, wonach israelische Forscher ein Programm entwickelt hätten, das Sprache lernt. Es wird behauptet, die Software könne die Welt der Intelligenz revolutionieren. Jeder Mensch, der sich mit der Materie beschäftigt, weiss, wie komplex Sprache ist. Er weiss, dass da etwas nicht stimmen kann. Wenn das Programm wirklich existieren würde, dann wäre es in der Wissenschaftswelt Thema Nummer eins. Solche Behauptungen wandern durch alle Zeitungen. Der eine übernimmt sie vom andern.

Wissenschaftliche Arbeiten sind aber auch nicht unbedingt ein Hort der Redlichkeit.

Jedes Jahr werden Abermillionen von wissenschaftlichen Artikeln produziert, darunter findet sich auch viel Unfug, das muss man auch sagen.

Das Problem hat auch eine politische Dimension. Da wird auch immer wieder auf der Ebene einzelner Experimente argumentiert, zum Beispiel in der Gentechnik-Diskussion.

Auf der Basis von einzelnen Experimenten sollte man gar nicht argumentieren.

Man sieht auch im Skandal um den koreanischen Forscher Hwang, der seine «bahnbrechenden Resultate» zur Stammzellenforschung gefälscht hat, welche Konsequenzen die mediale Überhitzung mit sich bringen kann. Aber irgendwie muss man das Publikum doch auch an seinen Arbeiten Teil haben lassen.

Dessen sind wir uns durchaus bewusst. Wir haben deshalb ein Projekt lanciert, bei dem es darum geht, Methoden zu entwickeln, wie der Computer die Komplexität von Musik und ihrer Konstruktion erlebbar machen kann.

Können sie mehr dazu sagen?

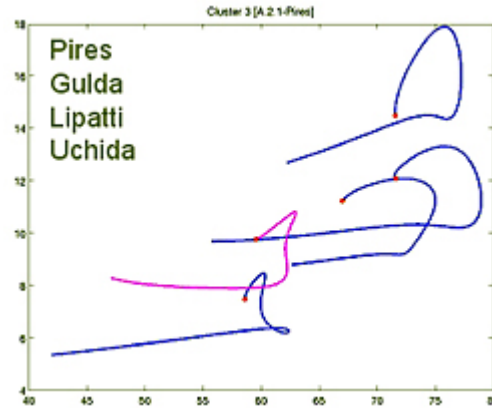
Da geht es unter anderem darum, im Konzertsaal live und in Echtzeit eine visuelle Animation der gespielten Musik zu präsentieren, die Struktur zu entwickeln. Dazu braucht man einen Computer, der in Echtzeit umsetzen kann, was gerade gespielt wird.

Man kann da skeptisch sein. Am Versuch, Zeitprozesse in visuelle Darstellungen zu übersetzen, sind schon viele gescheitert, angefangen bei Skrjabin. Das Auge bietet immer eine globale und simultane Sicht, das Ohr eine serielle und momentane.

Aber das ist ja gerade die Chance der Visualisierung. Diese «Interpretationswürmer» setzen ja zum Beispiel interessante Eigenschaften einer Interpretation bildhaft um.

Aber sie sehen nicht so aus wie Musik klingt, wenn man das mal so paradox formulieren will.

Das ist richtig. Andererseits habe ich eben gerade die Möglichkeit, etwas Komplementäres herauszulesen. Es geht nicht darum, etwas zu zeigen, das akustisch genauso gut wahrgenommen werden kann, sondern etwas, das akustisch eben schwer zu sehen ist. Mit der Spur des Interpreteten im Koordinatensystem sieht man etwa die letzten dreissig Sekunden eines Vorganges und hat nicht nur die letzten fünf präsent wie beim Hören.



Mit Hilfe typischer Bewegungsformen der «Würmer» lassen sich Interpretationen verschiedener Pianisten zu Gruppen zusammenstellen. Pires, Gulda, Lipatti und Uchida haben in Bezug auf die untersuchten Stücke zum Beispiel ähnliche Interpretationsmuster.

Das Lesen einer solchen Darstellung ist dann aber kein intuitiver Prozess des Erfassens mehr.

Richtig, man muss die visuelle Sprache lernen. Mit der Visualisierung kann man eben gerade Dinge sichtbar machen, die man nicht so intuitiv erlebt, und einen Aha-Effekt auslösen. Das funktioniert auch durchaus: Wir haben vor Publikum schon Präsentationen gemacht - im Wiener Konzerthaus. Da haben wir sehr anschaulich gezeigt, worin unterschiedliche Interpretationsstrategien bestehen. Wenn man das anhand der Aufnahmen vergleichen will, dann hat man schon vergessen, wie es der eine gemacht hat, wenn man endlich die Aufnahme des andern abspielen kann. Für gewisse Aspekte ist so eine Visualisierung also durchaus sinnvoll.

Ihre Resultate sind in «Musicae Scientiae», einer Zeitschrift erschienen, die sich den Kognitionswissenschaften verschrieben hat. Da liegt die Frage nahe, ob man die Resultate für Konzepte im Bereich der Kognition fruchtbar machen kann.

Das glaube ich nicht, nicht direkt. Wenn wir empirisch bleiben und uns nicht in Spekulationen versteigen wollen, dann müssten wir auch irgendetwas in den Messdaten gemessen haben, damit wir etwas über Kognition aussagen können. Entweder wie ein Zuhörer die Musik wahrnimmt, oder welche Zuordnung von Pianist und Stil ein Zuhörer machen würde, das müsste man dann auch experimentell messen. Dann könnte man versuchen, die Daten zu korrelieren und zu schauen, welche Muster, die wir gemessen haben, die Entscheidungen der Zuhörer am ehesten erklären. Solange wir keine Messdaten von Hörern haben, die uns irgendeinen Rückschluss auf kognitive Entscheidungen oder Dinge zulassen, können wir auch keine solchen Schlüsse ziehen.

In der traditionellen Musikästhetik wären da aber schnell einmal Formulierungen zur Hand wie: «Man spürt da die depressive Anlage des Interpreteten in diesen schleppenden Achteln».

Da kommt das Problem der sprachlichen Metapher hinzu. Die «Würmer», die wir analysieren, wären ein wunderbares Medium für solche Spekulationen. Aus empirischer Sicht kann man aus diesen

Daten aber keinerlei Rückschlüsse auf kognitive Vorgänge oder Prinzipien ableiten. (*wb*)

Gerhard Widmer: «Studying a creative act with computers: Music performance studies with automated discovery methods», *Musicae Scientiae – The Journal of the European Society for the Cognitive Sciences of Music*, Volume IX Number 1, Spring 2005, Pages 11-28.

Copyright <http://www.codexflores.ch>