



„Können Maschinen spielen wie Glenn Gould?“

Der Informatiker Gerhard Widmer spricht über den Einsatz von Künstlicher Intelligenz in der Musik und erklärt, was mit Hilfe von Computern beim Spielen, Hören und Analysieren musikalischer Werke möglich ist – und was nicht.

Von Peter Payer

„Wiener Zeitung“: Herr Widmer, Sie beschäftigen sich seit mehr als zwei Jahrzehnten mit dem Einsatz von Künstlicher Intelligenz (KI) in der Musik. Was genau ist Ihr wissenschaftliches Interesse?

Gerhard Widmer: Ich will versuchen, es auf zwei Hauptebenen zu reduzieren. Einerseits geht es um akustische Wahrnehmung: Kann man einer Maschine beibringen, Musik zu „hören“? Kann also eine Maschine in akustischen Signalen, Schallwellen und Audiostreams abstraktere musikalische Muster erkennen, so wie wir Menschen es beim Musikhören tun? Wir erkennen quasi automatisch Beat, Takt, Rhythmus, Tempo, Melodien und Akkorde, wir können Musikstücke Stilrichtungen und Genres zuordnen. All diese Dinge versuchen wir Computern beizubringen. Das zugehörige wissenschaftliche Forschungsgebiet wird üblicherweise *Music Information Retrieval* (MIR) genannt, und das ist inzwischen ein sehr großes, weltweites Forschungsgebiet. Eine zweite Hauptrichtung unserer Forschung, die auch eng mit meinem Namen verbunden ist,

begreift den Computer samt KI-Methoden als Hilfsmittel zum Studium musikalischer Fragen. Hier geht es um die ausdrucksvolle Musikinterpretation und ihre Analyse und Modellierung im Kontext der sogenannten klassischen Musik. Wie „muss“ man eine Komposition spielen, damit sie lebendig wird, einen bestimmten musikalischen Charakter annimmt, bestimmte Affekte zum Ausdruck bringt? Wie spiegelt sich das in technischen Parametern und Entscheidungen der Interpreten wider?

Spannende Fragen. Wie gehen Sie da genau vor?

Unser Zugang ist ein computer- und datenbasierter. Wir vermessen Interpretationen im Detail – also die genauen Verläufe von Tempo, Timing, Dynamik, Artikulation in Aufnahmen vor allem von Pianistinnen und Pianisten – und suchen in den Daten nach Mustern, nach systematischen Zusammenhängen zwischen bestimmten Aspekten der geschriebenen Musik und der Art, wie sie gespielt wird. Dazu trainieren wir zum Beispiel *Machine-Learning-*

Modelle, die lernen sollen, Interpretationsmuster für ein gegebenes Stück vorherzusagen und damit auch neue Stücke selbst „musikalisch“ zu spielen – „ausdrucksvoll“ wäre wohl ein zu großes Wort. Diese gelernten Computermodelle kann man dann analysieren, und in kontrollierten Experimenten kann man bestimmte Hypothesen empirisch testen. Analyse via Modellierung würde ich diesen Zugang nennen.

Sie haben selbst eine musikalische Ausbildung, haben Klavier gelernt. Wie hat sich biografisch gesehen Ihre Faszination für die Musik herausgebildet?

Ich glaube, vieles, was einem und mit einem passiert und was man schlussendlich zu tun entscheidet, sind kontinuierliche Prozesse, gelenkt von vielen kleinen Zufällen, aber halt auch zusammengehalten von einer zugrunde liegenden Grundkonstitution – Interessen, Fähigkeiten, Persönlichkeitseigenschaften. Also in meinem Fall: Ich liebe Musik, für mich ist das etwas sehr Wichtiges im Leben. Das war schon immer so, solange ich zurückdenken

kann, auch wenn ich schlussendlich zu faul war, konsequent Klavier zu üben. Dass daraus für mich ein Forschungsthema geworden ist, das mich dann auch zu allerlei Ehren und wissenschaftlichem Erfolg geführt hat, hat mit diesem Grundinteresse, mit einer vielleicht doch ein bisschen analytischen Ader – und mit vielen glücklichen Zufällen zu tun.

In den 1980er Jahren haben Sie in den USA studiert und sind dann gewissermaßen in das Computerzeitalter hineingewachsen. Inwiefern hat die zunehmende Digitalisierung Ihr Denken und Ihre wissenschaftliche Arbeit verändert?

Das ist interessant, dass Sie diesen Aspekt ansprechen. Genau damals in den USA, Mitte der 1980er Jahre, habe ich als Student meine erste E-Mail geschrieben, auf einem Monitor ohne Fenster und Maus, Zeile für Zeile. Ich erinnere mich noch gut daran, das war irgendwie schon aufregend. Ich hoffe aber, dass die Digitalisierung mein grundsätzliches Denken nicht oder möglichst wenig

verändert hat. Was sich natürlich verändert hat, ist die praktische Art, wie wissenschaftliche Arbeit im Alltag vor sich geht. Allein schon die Tatsache, dass man heute große Mengen von Musik in digitaler Form codieren, speichern, bearbeiten und weltweit verfügbar machen kann, hat unsere Arbeit massiv verändert bzw. macht vieles technisch erst möglich.

Wenn Sie musikalische Strukturen mit Hilfe von Algorithmen analysieren: Was kann man daraus lernen, das ohne KI nicht möglich wäre?

Wahrscheinlich nichts – also zumindest rein theoretisch betrachtet. Alles, was wir quantitativ,

„Wir trainieren Machine-Learning-Modelle, die lernen sollen, Interpretationsmuster für ein gegebenes Stück vorherzusagen und damit auch neue Stücke selbst ‚musikalisch‘ zu spielen.“

mit Computerhilfe analysieren, ist ja da, in den Daten, in der Musik. Und das müsste man daher auch ohne Computer finden, wenn man genug Zeit und Konzentration für die Analyse der Daten hätte. Was Computer tun können, ist, uns beim Messen helfen, bei der Suche nach Mustern in großen Mengen von Messdaten. Und sie erlauben uns auch, alternative Szenarien zu simulieren, um verschiedene Annahmen zu testen.

Arbeiten Sie dabei hauptsächlich mit Klavierwerken?

Ja, fast ausschließlich. Und zwar nicht nur weil mir das Klavier als Instrument persönlich nahesteht, sondern auch aus ganz praktischen Gründen. Einerseits ist es eine Frage der Komplexität. Die Freiheitsgrade für Interpreten auf dem Klavier sind relativ eng eingegrenzt und definiert: zeitliche Faktoren (Tempo, Timing), Anschlag (Lautstärke, Artikulation), Pedal. Dazu kommt das Messproblem: Um an solche Daten zu kommen, brauchen wir spezielle Instrumente, die diese Spieldetails tatsächlich messen und aufzeichnen. Und da gibt es im Bereich des Klaviers diese wunderbaren Computerflügel wie den Bösendorfer CEUS, die uns solche Daten liefern können. Bei anderen Instrumenten ist das nicht so einfach möglich.

Wenn Sie den Computer ein Klavierstück spielen lassen, beispielsweise nach Vladimir Horowitz oder Glenn Gould: Wie sehr können Sie sich hier den Originalinterpreten annähern? Ist das nur mehr eine Frage der Technik und der Rechenleistung?

Also, der Computer ist noch ganz weit davon entfernt, wie Horowitz oder Gould spielen zu können. Technisch natürlich schon: Ein Computerflügel kann schneller und präziser spielen als sogar ein Horowitz. Aber musikalisch... Wir könnten vielleicht sogar, wenn wir alle Gould-Aufnahmen präzisiert vermessen und maschinelle Lernprogramme mit diesen Daten trainieren würden, Modelle lernen, die bestimmte Aspekte des Gould-"Stils" reproduzieren könnten, etwa im Bereich der Artikulation. Aber wir könnten nicht vorhersagen, wie Gould ein Stück gespielt hätte, das er nie gespielt hat. Und dann: Ein Gould hätte ein Stück auch sehr verschieden spielen können, je nach Raum, nach geistiger Disposition, nach affektiver und intellektueller Absicht, je nachdem, was er mit einem Stück und einer Interpretation sagen wollte. Das ist also keine Frage der Technik oder Rechenleistung, sondern schlussendlich eine Frage der Grenzen der Modellierbarkeit. Musik zu verstehen und zu interpretieren ist eine intellektuelle Leistung des ganzen Menschen, mit all seinen persönlichen Erfahrungen, seinem Wissen, seinen emotionalen und künstlerischen Ausdrucksintentionen. Kein mathematisches Modell der Welt kann das alles einbeziehen. Und das ist auch gut so.

Ist es nach wie vor einfach zu erkennen, ob ein Musikstück von einer Maschine gespielt wird oder von einem Menschen?

Ja, der Unterschied ist immer noch groß, vor allem bei komplexer Musik. Die offensichtlichsten Unterschiede liegen in der Art der Fehler, die Maschinen machen, nämlich Fehler einer unmusikalischen Art, die Menschen in dieser Form nie machen würden, wie sie eine bestimmte Passage nie spielen würden. Und diese Fehler weisen auf tieferliegende Probleme hin, die Maschinen mit Musik haben: Es fehlt an einem Verständnis von abstrakten Konzepten wie Struktur, musikalischem Fluss, dramatischem Bogen, und der Unterscheidung, was wichtig und was unwichtig ist. Und einem Verständnis dafür, wie Zuhörer das



Glenn Gould bei einem Konzert mit den Berliner Philharmonikern, Berlin 1957. Foto: picturedesk.com / Erich Lessing

alles wahr- und aufnehmen. Das sind Aspekte, die noch schwer formal zu fassen sind.

Es gibt ja auch den berühmten „Turing-Test“.

Ja, mit „Turing-Test“ (benannt nach dem britischen Logiker und Informatiker Alan Turing, 1912-1954, Anm.) bezeichnet man in der KI-Forschung Experimente, bei denen getestet wird, ob Menschen eine Maschine oder eine maschinelle Leistung noch von einer menschlichen unterscheiden können. In diesem konkreten Fall hat ein australisch-italienisches Forscherteam verschiedene Aufnahmen ein und desselben Musikstücks – einer einfachen Kuhlau-Sonatine – einem Testpanel von 170 Personen vorgespielt. Alle Aufnahmen wurden auf demselben Flügel gemacht; vier davon waren von Computermodellen produzierte Interpretationen (von Forschungsgruppen aus Italien, Schweden, Japan – und unserer eigenen), eine von einem echten Konzertpianisten. Dann wurden die Testpersonen gefragt, welche Aufnahmen sie für von Menschen gespielt hielten. Und da hat dann offenbar tatsächlich unser Modell einen leicht höheren Score erreicht als sogar der echte Pianist. Aber man darf das nicht überbewerten. Das Experiment war sehr eingeschränkt, mit nur einem Teststück, das noch dazu

Weniger als 0,01 Prozent... Ist Musik überhaupt (be)rechenbar?

Die Frage ist in dieser Form zu allgemein, als dass ich sie beantworten könnte. Was könnte man mit „Berechenbarkeit“ in der Musik meinen? Können wir berechnen, wie ein Stück an einer bestimmten Stelle weitergehen „muss“? Nein; aber wir können abschätzen, welche Fortsetzungen mehr oder weniger wahrscheinlich sind (und daher auch für Hörer aus unserem Kulturkreis mehr oder weniger vorhersehbar oder überraschend sind). Können

„Unsere automatischen Klavierbegleiter hoffen wir in ein paar Monaten so weit zu haben, dass wir ihn in einem Konzert präsentieren können.“

sehr einfach ist und Interpreten wenig Möglichkeiten bietet.

Wenn Sie mittels KI das Rätsel der Musik ergründen, ihrer komplexen Wirkung auf uns Menschen nachgehen: Wie viel Prozent würden Sie sagen, haben Sie gelöst?

Weniger als 0,01 Prozent...

Ist Musik überhaupt (be)rechenbar?

Die Frage ist in dieser Form zu allgemein, als dass ich sie beantworten könnte. Was könnte man mit „Berechenbarkeit“ in der Musik meinen? Können wir berechnen, wie ein Stück an einer bestimmten Stelle weitergehen „muss“? Nein; aber wir können abschätzen, welche Fortsetzungen mehr oder weniger wahrscheinlich sind (und daher auch für Hörer aus unserem Kulturkreis mehr oder weniger vorhersehbar oder überraschend sind). Können

wir vorhersagen, welche Stücke jemand mögen wird oder nicht? Nein (trotz gelegentlicher Artikel in den Medien über angeblich „Hit Prediction“-Algorithmen); aber wir können natürlich Präferenzen abschätzen (und werden manchmal danebenliegen), wenn wir Information darüber haben, was eine Person oft gehört hat. Können wir harmonische, rhythmische, melodische Eigenschaften eines Stücks mathematisch beschreiben? Ja, natürlich. Musik ist ja ein hochstrukturiertes Artefakt, nach bestimmten stilistischen Konventionen und Grundregeln komponiert; und die Beziehungen zwischen Tönen, Intervallen, Akkorden in unserer tonalen Musik lassen sich sehr elegant in mathematischen Strukturen abbilden.

Wo sehen Sie die praktische Anwendung Ihrer Forschungen im Alltagsgebrauch?

Wir haben in den letzten 15 Jahren Technologien für eine ganze Menge von Anwendungen entwickelt, in Kooperation mit Playern in der Musik- und Audiowelt. Unsere Algorithmen zum Abschätzen musikalischer Ähnlichkeit finden Sie zum Beispiel in der BeoSound 5 von Bang & Olufsen, einer digitalen „Stereoanlage“, die selbst entscheidet, was sie als Nächstes spielt. Oder im automatischen Music Recommender im FM4 Soundpark, der Benutzern auf Basis musikalischer Ähnlich-

keit neue Songs vorschlägt. Unsere Algorithmen für die Detektion von Musik und Sprache in Audiostreams helfen der Schweizer Swisssperform, die Menge von Musik, die von Radiosendern in einem Jahr ausgestrahlt wird, zu schätzen, zwecks Tantiemenverrechnung. Im Kontext der Musikinterpretation beschäftigen wir uns derzeit mit interaktiven musikalischen „Partnern“ – „Companions“, wie wir sie nennen –, die mit Menschen gemeinsam Musik spielen. Nicht auf einem wirklich künstlerischen Level natürlich, aber so, dass sie als Partner beim Üben dienen könnten. So ein automatischer Klavierbegleiter muss in der Lage sein, mir beim Spielen zuzuhören, mir zu folgen, meine Ausdrucksintentionen zu erkennen und sich in seinem Spiel daran anzupassen, damit wir wirklich zusammen spielen. Unseren automatischen Klavierbegleiter namens „ACCompanion“ hoffen wir in ein paar Monaten so weit zu haben, dass wir ihn in einem Konzert präsentieren können.

Was würden Sie sagen, waren bisher Ihre größten Erfolge?

Worauf ich am stolzesten bin, und was sicher das wichtigste Resultat meiner Arbeit ist, sind all die jungen Forscherinnen und Forscher, die sich bei uns entwickelt und ihre Doktorarbeiten gemacht haben und die jetzt in vielen Ländern und spannenden Kontexten tätig sind. Dass sie bei mir und uns gelernt haben, was Wissenschaft ist, wie man gute, saubere und elegante Arbeit macht. Da geht im Übrigen auch das ganze Forschungsgeld hin: in die Finanzierung junger, kreativer Menschen, die sich vor keinem schwierigen Forschungsproblem fürchten. Mit einem solchen Team arbeiten zu können, ist der schönste Aspekt meiner Arbeit.

Peter Payer ist Historiker und Stadtforscher sowie Kurator im Technischen Museum Wien. Ebendort ist derzeit eine große interaktive Ausstellung mit dem Titel „Künstliche Intelligenz?“ zu sehen: www.technischesmuseum.at



Foto: Ethan Vincent

Gerhard Widmer, geboren 1961, war in seiner Jugend erfolgreicher Pianist, studierte Informatik an der TU Wien – und gehört heute zu den international profiliertesten Forschern im Bereich der computergestützten Musikanalyse. Er ist Leiter des Instituts für Computational Perception der Johannes Kepler Universität Linz und stellvertretender Leiter des Artificial Intelligence Labs am Linz Institute of Technology (LIT). 2009 erhielt er den Wittgenstein-Preis, und 2021 (nach 2015) bereits den zweiten ERC Advanced Grant des Europäischen Forschungsrates.