

BE OPEN-FESTIVAL

BE OPEN-FESTIVAL

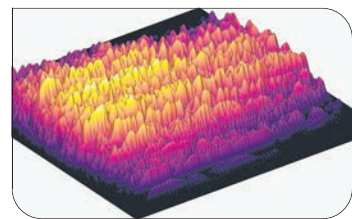
Grundlagenforschung in Österreich: Beispiele für FWF-geförderte Spitzenforschung

Kollaborative Kreativität



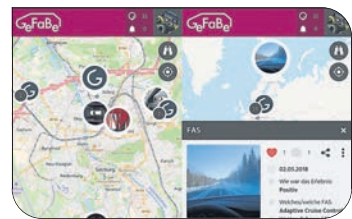
Musizierende müssen ihr Spiel mit höchster Präzision miteinander koordinieren, um harmonieren zu können. Laura Bishop untersucht mit ihrem Team, wie Musizierende die Tendenz zur Einzigartigkeit mit der Notwendigkeit der gegenseitigen Vorhersagbarkeit balancieren.

Akustische Modelle



Österreich ist berühmt für zwei Dinge: Musik und Berge. Peter Balazs und sein Team untersuchen, wie Audiosignale in der Zeit-Frequenz-Dimension optimal als „Berge“ abgebildet werden können und wie eine solche „Topographie“ verwendet werden kann, um akustische Signale und Systeme besser zu verstehen.

Gemeinsam Fahren Wir Besser (GeFaBe)



Manfred Tscheligi, Professor für Human-Computer Interaction, leitet ein Citizen-Science-Projekt mit Autofahrer/innen, die in einer App ihre Erfahrungen mit Fahrassistenzsystemen festhalten. Unter wissenschaftlicher Anleitung werden Verkehrsszenarien ausgearbeitet, die die Grundlage für die erste auf Fahrassistenzsysteme fokussierte Verkehrsunterrichtseinheit in Österreich bilden.

Formale Methoden für komplexer Systeme



Der Informatiker Thomas Henzinger vom IST Austria beschäftigt sich mit mathematischen Methoden und Computeralgorithmen für die Entwicklung verlässlicher Softwaresysteme.

Können Computer Musik wahrnehmen?

Künstliche Intelligenz. Uns fällt es ziemlich leicht, in Schallwellen Muster zu erkennen. Für Maschinen ist das eine harte Nuss.

Wenn ein Mensch nur zehn Sekunden eines Musikstückes hört, kann er eine Menge daraus ableiten. Man erkennt eine Melodie, einen Rhythmus, man erkennt, ob ein Sänger oder eine Sängerinnen am Werk ist, in welcher Sprache gesprochen wird, welche Instrumente beteiligt sind, in welcher Stilrichtung sich das Stück bewegt, welche Stimmung es rüberbringt und so weiter. „Unser Gehirn leistet da unheimlich viel, ohne dass wir es merken“, sagt Gerhard Widmer. Für einen Computer hingegen sind das sehr schwierige Aufgaben – von denen der Informatiker bereits so einige lösen konnte. Am Institut für Computational Perception der Universität Linz und am Österreichischen Forschungsinstitut für Artificial Intelligence (OFAI) in Wien widmet er sich seit 20 Jahren intensiv dem Fachgebiet der Computerwahrnehmung, im Speziellen der Wahrnehmung von Audiosignalen.

Enorm viele Daten

„Wahrnehmung ist ein zentraler Teil der Intelligenz: Ein Wesen, das sich in der Welt intelligent verhalten soll, muss zuerst einmal in der Lage sein, wahrzunehmen, wo es ist und was um es herum passiert“, so Widmer. Beim Menschen sei es relativ klar, was Wahrnehmung ist: „Wir haben Sinnesorgane – z. B. Augen, Ohren, Nase oder Tastsinn –, mit denen wir Information über die Welt um uns bekommen und die im Gehirn irgendwie interpretiert werden.“ Diese Prozesse stellen eine große kognitive Leistung dar: „Aus Millionen von Low-level-Informationen baut unser Gehirn ein abstraktes Bild der Welt zusammen.“ Er meint damit, dass die Si-



Mit künstlicher Intelligenz kann ein Computer in den Noten mitlesen, wenn er die Musik „hört“.

gnale der Sinnesorgane gigantische Mengen an Daten darstellen, die jede für sich gesehen kaum Informationen liefern. „Unsere Augen kann man mit einer Pixelkamera vergleichen: Es liefert Millionen von Pixeln, die für sich fast null Information haben. Wenn man eines weglässt, ändert das überhaupt nichts.“ Ähnlich ist es bei einem Audio-Signal: „Das Eingangssignal ist eine Abfolge von Schwankungen des Luftdruckes über die Zeit. Dies Schallwellen treffen auf unser Ohr und bringen das Trommelfell zum Zittern. Aus diesen Messwerten und Kurven kann man gar nichts aussagen. Aber der Mensch nimmt

dennoch ohne Anstrengung tausende Dinge wahr.“ Und zwar in Echtzeit, wie Widmer bewundernd hinzufügt. Das einem Computer beizubringen ist aus mehreren Gründen schwierig. Zum einen geht es um riesige Datenmengen, die innerhalb kürzester Zeit verarbeitet werden müssen. „Ein typischer Popsong mit drei Minuten Länge liefert nach der Digitalisierung rund 16 Millionen Zahlen.“ Zum anderen sind viele der Prozesse, die im menschlichen Gehirn ablaufen, noch zu wenig verstanden, sodass man diese Fähigkeiten nicht in einem Computersystem programmieren kann. „Niemand kann

einen Algorithmus schreiben, der aus der Folge von Zahlen erschließt, ob es sich um Musik, um Sprache oder um Lärm handelt.“ Die einzige Weise, wie man das lösen kann, ist durch Methoden der Künstlichen Intelligenz – insbesondere durch maschinelles Lernen. „Der Computer muss es selbst lernen, indem man ihm tausende Beispiele vorspielt, bei denen ein Mensch zuvor annotiert hat, was zu hören ist.“ Aus solchen Trainingsdaten muss der Computer relevante Muster und Gesetzmäßigkeiten erlernen – und er muss diese so verallgemeinern, dass er auch neue, unbekannte Situationen richtig beurteilen kann. „Ler-“

nen heißt nicht nur klassifizieren und zusammenfassen, was in den Trainingsdaten enthalten ist, sondern auch verallgemeinern. Das ist das absolut schwierigste beim Lernen.“ Auf diese Weise könne man grundsätzlich alle Wahrnehmungsprobleme lösen, so Widmer: So beruhen zu Beispiel alle Sprach- oder Gesichtserkennungsprogramme auf diesem Prinzip. Theoretisch gesprochen handelt es sich dabei um eine Art Suchproblem: Der Rechner muss in der sehr großen Menge von Möglichkeiten Kombinationen finden, die statistisch signifikant sind. Zum Einsatz kommen dabei viele unterschiedliche Verfahren, unter anderem neuronale Netze. „Deep learning ist derzeit der Renner. Das sind neuronale Netze, die mehrere Schichten haben. Das konnte man früher nicht machen, weil es sehr rechenintensiv ist.“

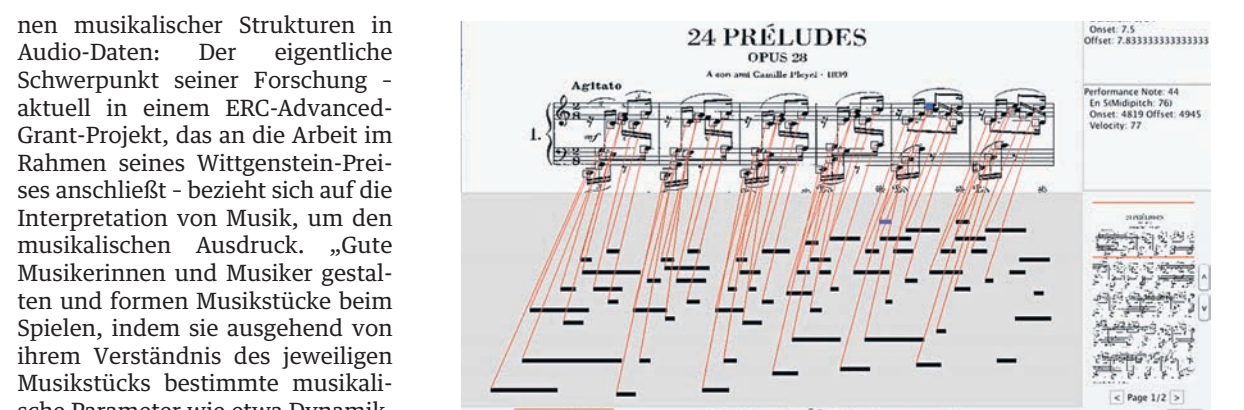
Musiker live zuhört und ihn begleitet: Er erkennt mit hoher Zuverlässigkeit den Rhythmus z. B. eines Gitarrensongs, spielt dazu Percussion-Instrumente und passt sich selbstständig Tempoänderungen an. Im Einsatz sind auch schon Systeme, die automatisch Ähnlichkeiten zwischen Musikstücken ermitteln und auf dieser Basis Benutzern andere Stücke vorschlagen (realisiert wurde das beispielsweise im FM4-Soundpark). Oder Systeme, die selbstständig den Refrain in einem Popsong finden, sodass man beim schnellen Durchhören einer Musiksammlung zielsicher jene paar Sekunden zu hören bekommt, die man hören will. Mit solchen Entwicklungen gewinnen Widmer und seine Mitforscherinnen und -forscher regelmäßig Preise bei internationalen Wett-

bewerben. „Musik ist ein Milliardenmarkt, weltweit gibt es hunderte Forschungsgruppen, die an solchen Methoden arbeiten. Auch die großen Medienkonzerne investieren viel Geld in diese Forschung“, so Widmer. Interpretationsfragen Für großes Aufsehen sorgte auch ein System aus dem Team Widmers, das bei Live-Konzerten automatisch dem Notentext folgt. „Unabhängig von Temposchwankungen und etwaigen Fehlern erkennt das Programm, um welches Musikstück es sich handelt und wo der Interpret gerade ist“, so Widmer. Dadurch wird zum Beispiel ein intelligenter Automat zum Umblättern von Noten möglich. Oder – wie es im Koninklijk Concertgebouw in Amsterdam live präsentiert wurde – ein System, das interessierten Zuhörern zu jedem Zeitpunkt exakt synchron zur Musik den aktuellen Notentext per WLAN auf ein Tablet zum Mitlesen des Gehörten schicken. Widmers Forschung bezieht sich aber nicht nur auf das Erken-

„Lernen bedeutet nicht nur klassifizieren und zusammenfassen, sondern auch verallgemeinern.“

Viele Anwendungen denkbar

Die Forschungsziele seines Teams umreißt Widmer so: „Unsere Forschung zielt darauf ab, Computern musikalisches „Hören“ und ein grundlegendes musikalisches „Verständnis“ beizubringen, um einerseits selbst mehr über Musik und Musikwahrnehmung zu lernen und andererseits mittelfristig neue Anwendungen mit Mehrwert für die digitale Musikwelt zu ermöglichen.“ So wurde beispielsweise ein System entwickelt, mit dem in einem Audiostream eines Radioprogramms unterschieden werden kann, wo gesprochen wird und wo Musik zu hören ist. Das ist unter anderem für die Abrechnung von Lizenzen für die Musik wichtig. Das System arbeitet zu mehr als 96 Prozent zuverlässig – das ist ähnlich gut wie bei Menschen, die die Aufgabe händisch vollführen. Konstruiert wurde weiters ein kleiner „Musikroboter“, der einem



Analyse von Interpretationsdetails in einer Chopin-Aufnahme.

Künstliche Intelligenz verändert Wissenschaft

Forschung. Kaum ein Bereich bleibt von den digitalen Umwälzungen ausgespart.

Künstliche Intelligenz (KI) ist ein Thema, das viele Leute berührt: „Jeder hat eine Vorstellung davon – ob richtig oder falsch“, meint Gerhard Widmer, Professor für Computational Perception an der Universität Linz. Entsprechend wird an diesem Thema geforscht, auch in Österreich. Schon im Jahr 1984 wurde das Österreichische Forschungsinstitut für Artificial Intelligence (OFAI) in Wien gegründet, das unverändert ein wesentlicher Kristallisationskeim für jegliche Forschung in diesem Bereich ist – sowohl in der Grundlagenforschung als auch in der angewandten Forschung. In Wien haben sich insbesondere die Technische Universität (etwa um Georg Gottlob) oder das Austrian Institute of Technology (etwa in Sachen Mustererkennung) als Hot-Spots etabliert.

Einen bemerkenswerten Schwerpunkt in Sachen KI gibt es an der Universität Linz: Dort sind neben Widmer unter anderem auch die Informatiker Armin Biere oder Sepp Hochreiter tätig. Letzterer war 1997 federführend an der Entwicklung einer Technologie namens LSTM („long short-term memory“) beteiligt, die im Zentrum praktisch aller Maschinenlern-Systeme steht. Der Schwerpunkt wird nun durch einen Ausbau des Linz Institute of Technology (LIT) stark erweitert, etwa durch ein neues LIT Artificial Intelligence Lab und Lehrstühle für Roboter-Psychologie, Digitale Soziologie oder Cyber-physical Systems.

Methoden der KI sind aus praktisch keinem Forschungsfeld mehr wegzudenken – nicht aus der Medizin, nicht aus der Linguistik, nicht aus der Geologie. Robert Trapp, Leiter des OFAI, untersucht beispielsweise archetypische Charaktere (wie etwa Helden und Bösewichte) in Film und Drama. Eingesetzt werden dabei KI-Methoden aus den Gebieten der automatischen Text- und Affektanalyse, der Sprachverarbeitung und des maschinellen Lernens, um in digital verfügbaren Dialogen nach Merkmalen der fiktionalen Gestalten zu suchen. Im Endeffekt soll dadurch auch ein Beitrag zur Generierung von facettenreicheren und interessanteren virtuellen Figuren geleistet werden. Ein weiteres bemerkenswertes Beispiel ist das Projekt „akustisches Tunnelmonitoring“ bei Joanneum Research, wo Methoden aus der Computerwahrnehmung im Dienste der Verkehrssicherheit angewendet werden: Konkret werden aus dem Verkehrslärm in Straßentunnels verdächtige Geräusche herausgefiltert, um im Fall eines Unfalls möglichst frühzeitig Maßnahmen einzuleiten.

Wie groß das Interesse an diesem Thema auch aus der Wirtschaft ist, zeigt die Einrichtung einer gemeinsamen Stiftungsprofessur an der Universität Klagenfurt und der Technischen Universität Graz. Dort sollen insbesondere neue Methoden für adaptive und vernetzte Produktionssysteme (Industrie 4.0) entwickelt werden.

Tapping Friend: Wie Musiker kommunizieren

Kooperation. Gemeinsam zu musizieren erfordert einen hohen Koordinationsaufwand zwischen allen Beteiligten. Was dabei in der Praxis entscheidend ist, wurde nun im Detail geklärt.

Zwischenmenschliche Kommunikation und das Koordinieren und Synchronisieren von Handlungen gehören zu den grundlegendsten menschlichen Fähigkeiten – nicht nur z. B. im Straßenverkehr oder im Mannschaftssport, sondern auch in der Musik.

Die Form der Kooperation und Synchronisation kann sehr unterschiedlich sein: Das reicht von diktatorischen Systemen wie in großen Orchestern mit starken Dirigenten bis hin zu demokratischen

Gepflogenheiten in kleinen Ensembles. Welche Prozesse dabei im Einzelnen ablaufen, um die eigenen Handlungen mit jenen eines anderen präzise zu koordinieren und auch zu antizipieren, hat Werner Goebel, Wissenschaftler am Österreichischen Forschungsinstitut für Artificial Intelligence und an der Universität für Musik und darstellende Kunst Wien, untersucht. Dabei wurden Wahrnehmungs- und Aufführungsexperimente mit klassischen und Jazz-Ensembles sowie Computermodellierung von

einzelnen Aspekten der Synchronisation durchgeführt. Dabei zeigte sich, dass neben akustischen Signalen – nämlich der Musik selbst – insbesondere visuelle Signale extrem wichtig sind, also etwa Kopf- und Handbewegungen an bestimmten Stellen eines Musikstückes. Die grundlegende Synchronisationsmechanismen wurden in ein künstliches Begleitsystem implementiert, das schließlich auf seine Einsatzfähigkeit in der musikalischen Praxis getestet wurde.

Ein Ergebnis der Forschung ist auch ein Demonstrator namens „Tapping Friend“, der zeigt, was es im musikalischen Kontext heißt, dass menschliche Synchronisation immer kooperativ ist. Bei diesem Gerät können Freiwillige erproben, wie gut sie dem Takt eines Dirigenten folgen können – und zwar abhängig davon, ob dieser nur seine Vorstellungen verfolgt oder ob er sich kooperativ zeigt, sich an die Umgebung anpasst und zum Beispiel Rhythmusfehler von Musikern ausgleicht.



Pavillon kuratiert von Pavillon curated by

Gerhard Widmer ist Leiter des Institutes für Computational Perception an der Johannes Kepler Universität Linz. Er wurde vom Wissenschaftsfonds (FWF) sowohl mit dem STARTI- (1998) als auch mit

dem Wittgenstein-Preis (2009) ausgezeichnet. 2015 erhielt er den vom Europäischen Forschungsrat einen ERC Advanced Grant.

