

Von Andreas Aichinger

Zwei Seelen schlagen in Gerhard Widmers Brust. »Völliger Quatsch« ist nämlich für den Experten für Künstliche Intelligenz und Maschinelles Lernen, was das amerikanische Kultmagazin »Wired« einmal über ihn geschrieben hat. Nämlich dass der Österreicher im Rahmen seiner Forschung nicht weniger als den mythologisierten »Sound of Music« messbar machen wolle. „Die volle Wirkung und Schönheit von Musik quantitativ fassen zu wollen, das wäre überhaupt die falsche Zielsetzung,“ beteuert hingegen Widmer, der einst im zarten Alter von 13 Jahren einen landesweiten Klavier-Wettbewerb für sich entschieden hatte. Dass der international für seine Musikforschung bekannte Computewissenschaftler dennoch alles daran setzt, seine Maschinen mit immer neuen Methoden gleichermaßen auf Mozart und »Metallica« loszulassen, ist da die andere Seite der Medaille.

# ZUKUNFT

## Künstliche Intelligenz und musikalische Konsequenz



*Forschung: »The Wheel« bietet eine intelligente Shuffle-Funktion, ordnet Musik selbstständig nach musikalischer Ähnlichkeit und organisiert sie als Playlist entlang eines schlichten Navigationsrades*



Begonnen hat alles eher spielerisch. Obwohl das Interesse für Musik immer da war, dienten musikalische Problemstellungen anfangs nur als Testdomänen für Widmers Lernalgorithmen. Und obwohl Musik zusehends in den Mittelpunkt seiner Arbeit rückte, wagte Widmer anfangs nicht, sie als zentrales Thema seiner Forschung anzugeben. 1998 dann der Durchbruch: Widmer erhält den wichtigen österreichischen START-Forschungsförderungspreis, 872.000 Euro machen es möglich, sich ab sofort im größeren Stil musikalischen Problemstellungen zu widmen. Als Widmer in der Folge eine Methode entwickelt, die individuelle Interpretation verschiedener Klavier-Virtuosen zu charakterisieren und in so genannten »Performance Worms« anschaulich darzustellen, bringt ihm das weltweite Bekanntheit und breites Medien-echo ein. Als Ausgangsmaterial dienten Widmer damals übrigens 13 exakt vermessene Mozart-Sonaten.

Heute beschäftigt sich Widmer als Vorstand des Instituts für Computerwahrnehmung (Computational Perception) an der Johannes Kepler Universität in Linz mit Forschungsfeldern, die neben rein akademischem auch großes ökonomisches Potential haben. Etwa mit neuartigen Music Interfaces. So wurde im Rahmen des Projekts »The Wheel« ein Java-Applet entwickelt, das eine Musiksammlung in ähnlich klingende Musikstücke entlang eines Navigationsrades gliedert und so eine intuitive Anwahl der Titel erlaubt. Wird das Rad nur um wenige Grad gedreht, bekommt der Musikfan ähnliche Musik zu hören, dreht er es um 180 Grad, wird völlig anderer Musikgeschmack bedient.

Ein zweiter Forschungsschwerpunkt Widmers, der auch die Abteilung für »Intelligent Music Processing and Machine Learning« am Österreichischen Forschungsinstitut für Künstliche Intelligenz in Wien leitet, ist der immer wichtigere Bereich »Music Information Retrieval« (MIR), speziell im Bereich Klassischer Musik haben die Österreicher hier die Nase vorne. Vor allem zwei Quellen lassen sich dabei anzapfen: Einerseits das Musiksinal selbst, andererseits kontextbasierte Informationen aus dem Internet – Stichwort Music Web Mining.

2006 wurde Widmers hervorragendes Standing in der Fachwelt gleich doppelt augenscheinlich. So fungierte er kürzlich als verantwortlicher Guest Editor einer Spezialausgabe des renommierten Journals Machine Learning zum Thema Musik („28 Beiträge wurden eingereicht, 21 davon habe ich abgelehnt“). Gleichzeitig wurden die Österreicher damit betraut, 2007 die wichtigste Fachkonferenz ihrer Disziplin (ISMIR, »International Conference on Music Information Retrieval«) in Wien zu veranstalten. Technology Review hat Gerhard Widmer in Linz besucht.

**Technology Review:** Musik wurde in der Informatik lange Zeit lediglich zum Testen von Algorithmen herangezogen und ist erst seit einigen Jahren als eigenständiger Forschungsgegenstand ins Zentrum gerückt?

**Gerhard Widmer:** Das ist richtig. Dennoch wurde im Bereich der Künstlichen Intelligenz immer davon geträumt, eines Tages auch Maschinen bauen zu können, die Musik machen, die Musik komponieren, Musik war immer etwas Faszinierendes. Ja, der Stellenwert von Musik ist in der Informatik viel größer geworden. Man sieht jetzt, dass mit Computern und Musik durchaus interessante, komplexe und



Gerhard Widmer bei der Arbeit: Interaktives Steuern einer Musikinterpretation mittels eines MIDI-Theremins

vor allem auch sinnvolle Dinge getan werden können, und dass es sich nicht nur um esoterische Spielereien handelt. Durch die Digitalisierung des Musikmarktes – Stichwort MP3, Stichwort »iPod« – wird Musik jetzt zu einem sehr wichtigen und auch ökonomisch relevanten praktischen Anwendungsgebiet. Über das Internet werden ja gigantische Mengen an Musik bezogen, und plötzlich besteht auch ein akuter Bedarf an intelligenten Methoden, um mit dieser Fülle an Musikstücken umzugehen.

**TR:** Ihr Projekt »The Wheel« etwa zielt genau auf diesen Bedarf ab. Erobert Künstliche Intelligenz jetzt auch mobile Music Player?

**Widmer:** Ich würde sagen ja. »The Wheel« ist prototypisch für viele verschiedene Ideen, die derzeit verfolgt werden. Music Player werden in Zukunft auch eine rudimentäre musikalische Intelligenz haben, weil der Benutzer sonst mit der riesigen Menge verfügbarer Musik auf seinem Gerät nicht mehr zurecht kommt. Wir wollen, dass der Benutzer einfach jene Art von Musik finden kann, die gerade seiner Stimmung entspricht.

**TR:** Hat »The Wheel« das Potential, eine Art »iPod plus« zu inspirieren?

**Widmer:** Das Spannende daran ist, dass dieses Gerät eine beliebige Musiksammlung selbstständig nach musikalischen Kriterien ordnet. Auf MP3-Playern müssen Sie sich ja heute noch selbst um eine hierarchische Strukturierung in Verzeichnissen kümmern. Viele Musikfans wollen sich das aber nicht mehr antun, was etwa auch den Erfolg des »iPod Shuffle« erklärt. »The Wheel« steht nun mit einer intelligenten Shuffle-Funktion eine Stufe darüber, dieses Gerät ordnet Musik selbstständig nach musikalischer Ähnlichkeit an und organisiert sie als Playlist entlang eines schlichten

*(Anm. der Red.: Dieses Interview ist am 14. 9. in der Zeitschrift »Technology Review« erschienen – wir empfanden es als gelungenen Kontrast zu den an dieser Stelle sonst abgedruckten »Geschichten« und bedanken uns beim Heise-Verlag für das Recht zur Veröffentlichung)*

„Musik in all ihrer Reichhaltigkeit auf ein paar Zahlen zu reduzieren, halte ich für völligen Quatsch“

Das Projekt »nepTune« (ein grafisches 3D-Interface samt Immersive Surround Sound zum interaktiven Explorieren und Navigieren in Musiksammlungen mittels eines Game Pads

Navigationsrades. Beim Abspielen folgen somit stets zueinander passende Musikstücke aufeinander. Die dahinterstehende Frage musikalischer Ähnlichkeit ist ja vielleicht überhaupt das zentralste Thema in unserem Forschungsfeld Music Information Retrieval (MIR). Der Computer muss quantifizieren, wie ähnlich zwei Musikstücke sind – das ist aus der Sicht der Audiosignalanalyse ein schwieriges Problem.

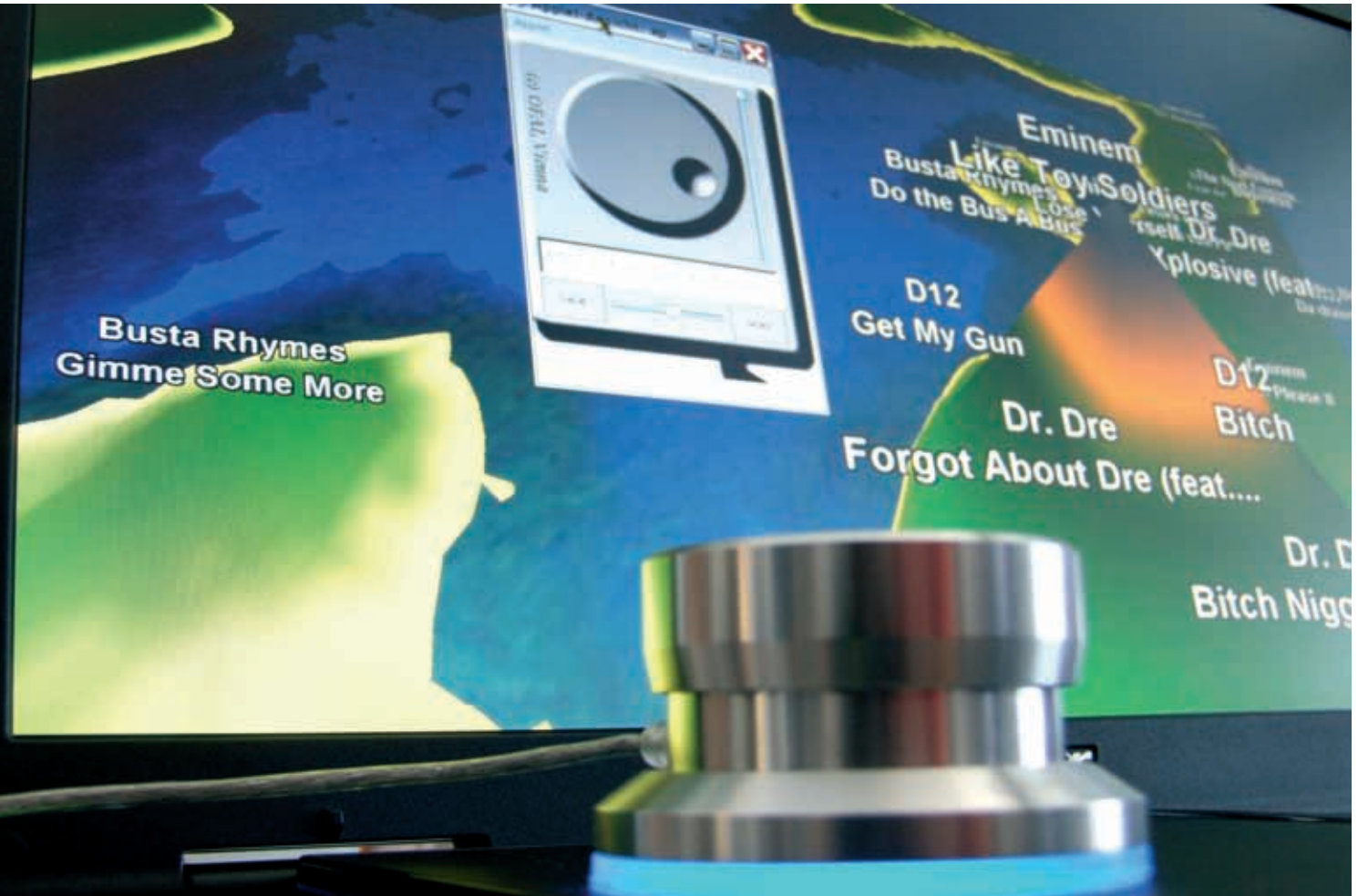
**TR:** Wie genau formen Sie aus einer beliebigen Musiksammlung, also aus einer Ansammlung verschiedener Audiosignale, eine Playlist, deren benachbarte Tracks Musikstücke derselben Stilrichtung sind?

**Widmer:** Zunächst macht es keinen Sinn, die beiden Signale direkt zu vergleichen, das wäre wie der Vergleich zweier Bilder auf der Pixel-Ebene. Uns interessieren abstraktere Informationen, die wir Features nennen. Also unterziehen wir die Signale einer Fourier-Transformation, um sie in verschiedene Frequenzbänder aufzuteilen. In diesen Frequenzbändern wird dann nach Mustern gesucht: Sind da Rhythmen auszumachen, die etwas darüber aussagen, ob das ein schnelles oder langsames, ein aggressives oder sanftes Stück ist? Aus dem fouriertransformierten Signal werden dazu Kennzahlen (Anm.: Mel Frequency Cepstral Coefficients,

MFCCs) berechnet, die etwas über die Klangfarbe des Audiosignals zu einem bestimmten Zeitpunkt aussagen. Ich kann nun versuchen, ein ganzes Musikstück durch die statistische Verteilung verschiedener Klangfarbenmischungen zu beschreiben. Für zwei Stücke lässt sich daraus ein Ähnlichkeitsmaß – eine Zahl zwischen Null und Eins – berechnen, bei tausend Stücken entsteht eine Ähnlichkeitsmatrix mit einer Million Einträgen. Zu guter Letzt wird auf dieser Matrix ein Travelling Salesman Problem optimiert, wobei die Ähnlichkeiten als Distanzen interpretiert werden. Die Playlist resultiert dann aus der optimalen Anordnung der Musikstücke.

**TR:** Ein aktueller MP3-Player wäre aber damit wohl technisch überfordert?

**Widmer:** Das Hauptproblem beim derzeitigen Stand der Technik ist, dass die Feature-Extraktion, das Berechnen dieser MFCCs, sehr aufwändig ist und auf einem externen Rechner geschehen muss. Während eines Ladevorgangs könnten dann aber auch gleich die nötigen Berechnungen durchgeführt und die Playlist übergeben werden. Zudem wächst die Rechen- und auch die Speicherkapazität dieser kleinen Geräte ja gigantisch schnell, das wird irgendwann kein Thema mehr sein.





**TR:** Ihre Arbeit stützt sich aber keineswegs nur auf die Analyse der Musik selbst?

**Widmer:** Wir verwenden immer mehr das Internet als Medium, aus dem wir automatisch Informationen gewinnen können, die unser Ähnlichkeitsmaß für Musikstücke verbessern. Die Ähnlichkeitsabschätzungen für unsere neuen Musik-Interfaces beruhen dann nicht nur auf dem Audiosignal selbst, sondern auch darauf, wie Musik von Menschen eingeordnet wird. Mit dem »Genre Crawler« haben wir etwa ein kleines Programm entwickelt, das versucht, die Stilrichtung von Künstlern nur auf Grund der Informationen aus dem Internet zu bestimmen. Es kommt ja auch auf den Kontext an, in dem Ähnlichkeit gesehen wird. Wenn Sie etwa Musik für Ihre Hochzeit suchen, dann haben Sie andere Kriterien, als wenn Sie Musik zum Joggen zusammenstellen. Bei Hochzeitsmusik kommt es auf »Tanzbarkeit« an, bei Laufmusik auf den Rhythmus. Die Kriterien ändern sich je nach Zweck und Situation total, es gibt keine eindeutige Definition.

**TR:** Welche neuen Features sind noch für tragbare Music Player der neuen Art denkbar? Woran forschen Sie in diesem Bereich?

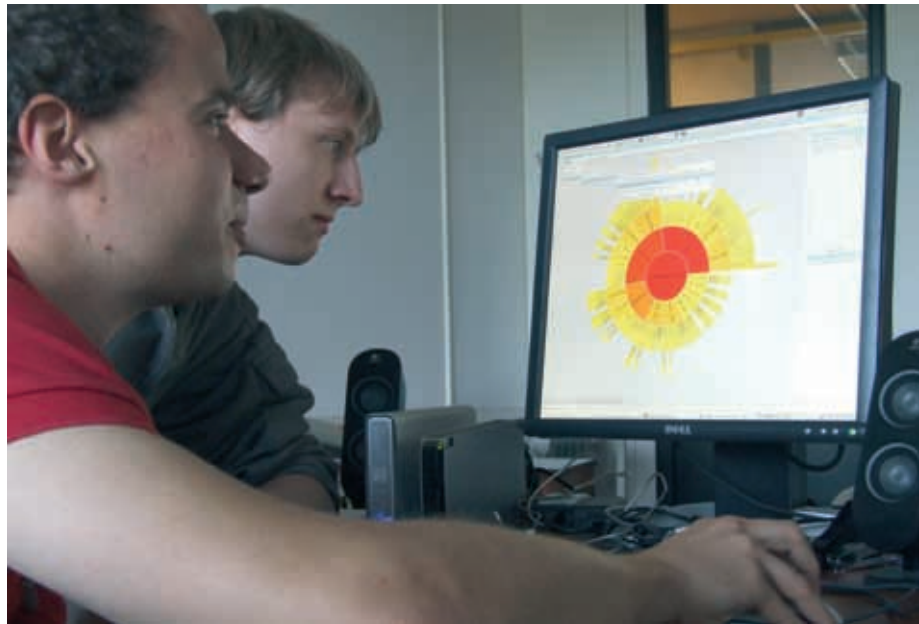
**Widmer:** Es ist sicher ein Trend, dass diese Geräte mehr in ihre Umgebung und in die Internetumgebung eingebunden sein werden, sich permanent zusätzlich zur Musik Informationen aus dem Internet holen werden. Also Meta-Informationen wie Lyrics, die Bilder der Band, CD-Covers. Wir arbeiten an einem Programm, das automatisch aus dem Web die zugehörigen Songtexte sammelt, und zwar nicht aus einschlägigen Portalen. Diese Geräte werden also immer mehr vernetzt werden, und die Musikindustrie wird nicht nur Musik verkaufen, sondern auch Meta-Informationen. Bilder, Videos, die Biografien der Künstler, Konzerttermine. In Sachen automatisches Finden relevanter Meta-Informationen im Internet ist aber noch viel Forschung notwendig.

**TR:** Informationen aus dem Internet – wir sprechen jetzt also sozusagen über Music Web Mining?

**Widmer:** Ja, das ist ein weites Feld, aber dazu gibt es noch nicht allzu viel Forschung. Was wir aktuell machen, klingt trivial, ist es im Detail aber nicht. Etwa die Frage: Wie kombiniert man Ähnlichkeitsinformationen aus dem Audiosignal in sinnvoller Weise mit jenen, die wir aus dem Internet beziehen? Ein Beispiel: Ich habe zwei Zahlen. Die eine sagt mir, dass zwei bestimmte Künstler eine Ähnlichkeit von 0,13 haben, die andere, dass zwei Lieder eine Ähnlichkeit von 0,73 haben. Ich kann daraus im einfachsten Fall den Mittelwert ausrechnen, das Problem aber auch auf tausend andere Arten formulieren. Das ist dann eine Frage intensiver Experimentierung, das ist Knochenarbeit. Ein Trend im Bereich Web Mining ist, dass mehr und mehr versucht wird, auch das User-Verhalten aus dem Internet zu rekonstruieren. Also herauszufinden, was sich Menschen in welchen Situationen wirklich anhören, in welchem Kontext, in welchen Kombinationen.

**TR:** Mensch und Musik versus Maschine und Musik – gibt es aus Ihrer Sicht Grenzen der Machbarkeit?

**Widmer:** Wenn wir den Menschen einmal auch als Maschine betrachten wollen, dann glaube ich daran, dass es



im Prinzip keine Machbarkeitsgrenzen gibt. Allerdings lässt sich das nicht einfach nur durch die Verbesserung von Computermodellen erreichen. Wenn der Computer in puncto Musikwahrnehmung jemals so gut werden soll wie der Mensch, dann muss er auch dieselbe Erfahrung haben, und er muss selbstständig lernen. Lernen ist ja ein großes Forschungsgebiet in der Künstlichen Intelligenz.

**TR:** Die Schwierigkeiten beginnen mit Sicherheit schon mit der Modellbildung?

**Widmer:** Ja, das wird völlig unterschätzt. Musikverständnis – etwas, das Menschen ein Leben lang erfahren und intuitiv erleben – ist in der Maschine enorm schwierig nachzubilden. Die meisten in MIR-Anwendungen verwendeten Ähnlichkeitsmaße basieren eigentlich nicht auf Features, die denen entsprechen, die wir als Menschen wahrnehmen. Nicht auf dem Rhythmus, nicht auf der Melodie, nicht auf der Harmonie. Es gibt eben nach wie vor noch kein Computerprogramm, das aus einer beliebigen Aufnahme verlässlich die Melodie extrahieren kann. Wir Menschen hören hingegen fast automatisch, welche Stimme die Melodie trägt.

**TR:** Wired hat einmal über Sie geschrieben: „His goal, simply put, is to quantify the elusive, often rapturously mythologized, sound of music.“ Einverstanden?

**Widmer:** Nein. Das würde implizieren, dass man Musik in all ihrer Reichhaltigkeit auf ein paar Zahlen reduzieren kann, und das halte ich für völligen Quatsch. Musik ist – auch in der Wirkung auf den Zuhörer – so reichhaltig, dass man diese Wirkung nicht in einer Menge von hundert Zahlen festmachen kann. Man kann bestimmte Aspekte festhalten, und das ist für Anwendungen wie neue Interfaces oder automatische Musikempfehlung vielleicht ausreichend. Aber die volle Wirkung und Schönheit von Musik quantitativ fassen zu wollen, das wäre überhaupt die falsche Zielsetzung. Ich glaube nicht, dass das möglich ist. ■

*Hier wird geforscht:  
»CoMIRVA« ist ein Projekt, in dem isualisierungsmöglichkeiten für Musiksammlungen entwickelt werden*

*„Wenn der Computer in puncto Musikwahrnehmung jemals so gut werden soll wie der Mensch, dann muss er auch dieselbe Erfahrung haben, und er muss selbstständig lernen“*