

KI und Musik: SaxEx

Jörg Leuser

Ausarbeitung im Rahmen des Proseminars „Künstliche Intelligenz“ im Sommersemester 2004 an der Universität Ulm

Diese Ausarbeitung befasst sich mit dem Thema KI und Musik. Zu Beginn werden die drei Hauptbereiche kurz vorgestellt, in denen KI-Techniken in der Musik angewendet werden. Begonnen wird mit den komponierenden Systemen, gefolgt von improvisierenden Systemen und schließlich werden vortragende Systeme vorgestellt. Dabei wird die geschichtliche Entwicklung an Hand von Beispielen aufgezeigt. Danach konzentriert sich der Artikel auf das vortragende System „SaxEx“ von Ramon López de Mántaras, Josep Lluís Arcos und Xavier Serra. Dieses System fügt einer ausdruckslosen musikalischen Eingabe Ausdruck hinzu wie es ein menschlicher Musiker tun würde. Es werden die drei Schritte dargestellt, in denen SaxEx arbeitet: Analyse (spektrales Modellersystem), Schlussfolgerung (Case-Based Reasoning), Synthese (spektrales Modellersystem). Besonderes Augenmerk gilt dem Case-Based Reasoning-Subsystem, da einerseits dieses Subsystem hauptsächlich auf KI-Techniken basiert und es andererseits einen neuen Ansatz gegenüber früheren vortragenden Systemen darstellt. Die Vorteile dieser Herangehensweise sowie ihre Grenzen werden ebenfalls kurz behandelt.

1. Die Geschichte

Techniken aus der künstlichen Intelligenz wurden schon früh eingesetzt um musikalische Probleme anzugehen. In der Anfangszeit der Computer-Ära, in der die Rechenleistung der Computer noch sehr beschränkt war, wurde an wenig zeitkritischen Systemen geforscht: Den komponierenden Systemen. Wie der Name schon sagt, versuchen komponierende Systeme Musikstücke zu komponieren.

Erst einige Zeit später wurde an improvisierenden Systemen geforscht. Diese interpretieren ein vorgegebenes Stück und ändern es – begrenzt durch Regeln – ab.

Der letzte große Bereich, in dem KI-Techniken in der Musik angewendet werden, ist der Bereich der vortragenden Systeme. Vortragende Systeme versuchen, möglichst ausdrucksstark zu spielen. Dabei wird das gestalterische Repertoire verwendet, das einem (menschlichen) Musiker zur Verfügung steht, wie z.B. Dynamik oder Artikulation. Der Bereich der vortragenden Systeme war lange Zeit nicht so sehr im Blickpunkt der Wissenschaftler. Das Hauptaugenmerk galt den komponierenden und improvisierenden Systemen. So stammen die meisten Arbeiten aus dem Bereich der vortragenden Systemen aus den letzten Jahren. Das System SaxEx, mit dem sich dieser Artikel hauptsächlich befasst, ist so ein vortragendes System.

Komponierende Systeme

Bereits 1958 stellten Hiller und Isaacson eine wegweisende Arbeit in der Computer-Musik vor¹. Ihre Arbeit führten sie am Illiac-Computer durch. Ihr Haupterzeugnis bestand in der sogenannten „Illiac Suite“, einem Stück für ein Streicherquartett. Für die Komposition verwendete ihr System (mit Hilfe von Markov-Ketten) pseudozufällig erzeugte Noten. Die erzeugten Noten wurden darauf geprüft, ob sie vorgegebenen Kompositions-Regeln der klassischen Harmonie genügten. Nur passende Noten wurden behalten, die anderen verworfen und in einem neuen Durchgang ersetzt. Die Arbeit von Hiller und Isaacson klammerte die Bereiche Ausdruck und emotionalen Inhalt aus. Sie gaben zu bedenken, dass vor dem Angehen des Problems des Ausdrucks erst einmal die einfachen Probleme zu lösen galt. Es gab in dieser Zeit noch weitere Arbeiten,

die mit Markov-Ketten, gearbeitet haben. Es zeigte sich, dass man mit Markov-Ketten nicht einfach hochqualitative Kompositionen erzeugen kann.

So ist es auch wenig verwunderlich, dass andere Herangehensweisen gewählt wurden. Moorer entwickelte 1972 ein System², das auf einem heuristischen Vorgehen beruht. Es war in der Lage, einfache Melodien zu generieren. Sein Vorgehen imitiert dabei die Vorgehensweise eines menschlichen Komponisten. Es gab auch regelbasierte Systeme wie das von Rader von 1974³. Die Regeln beschreiben dabei, wie Noten oder Akkorde zusammengesetzt werden können bzw. dürfen. Das Besondere daran war, dass die Regeln nicht starr angewendet wurden, sondern bewertet wurde, ob eine bestimmte Regeln im Kontext sinnvoll ist. Des weiteren gab es gewichtende Regeln, die besagten, wie wahrscheinlich eine passende Regel in dem aktuellen Kontext auch Anwendung findet.

Unter den komponierenden Systemen gibt es auch eine Reihe von Systemen, die sich damit befassen, eine vorhandene Melodie automatisch zu harmonisieren.

Eine der vollständigsten Arbeiten in diesem Gebiet ist die von Ebcioglu (1993)⁴. Sie trägt den Namen „Choral“ und ist in der Lage, Choräle im Stile von J.S.Bach zu harmonisieren. Dazu verwendete es heuristische Regeln sowie Constraints.

Das Imitieren von Komponisten ist auch Ziel des Projekts „Emi“ von David Cope (1990, 1987)⁵. Das System analysiert dabei Musikbeispiele von Komponisten und sucht nach den Regeln und Einschränkungen, die der Komponist verwendet hat. Aus den extrahierten Informationen erstellt das System ein statistisches Modell, nach dem es dann neue Stücke im Stile des analysierten Komponisten erzeugt. Mit dem System gelang es schon, den Stil von Mozart, Brahms, Debussy, Bach, Chopin und einigen anderen zu imitieren.

Es gibt einige Arbeiten, die (zum Teil) auf neuronalen Netzen basieren. „Harmonet“ von Feulner (1993)⁶ verwendet neuronale Netze in Verbindung mit Constraint-Satisfaction Techniken um Akkorde zu vervollständigen. Diese Arbeit wurde von Hörnel und Dagenhardt 1997⁷ bzw. Hörnel und Menzel 1998⁸ zum „Melonet“-System erweitert. „Melonet“ ist in der Lage, höhere musikalische Strukturen in Melodien zu erzeugen. Nach Aussage der Autoren sind beide Systeme zusammen in der

Lage, Variationen von Chorälen zu erzeugen, die qualitativ vergleichbar mit der Variation eines erfahrenen menschlichen Organisten sind.

Ein eher ausgefallenes Beispiel stellt das Projekt „Sicib“ dar.⁹ Es komponiert Musik nach Körper-Bewegungen. Dabei werden Daten von Sensoren verwendet, die an einem Tänzer befestigt sind. Das System wurde von Morales-Manzanares 2001 vorgestellt und ist echtzeitfähig.

Als letztes Beispiel sei noch die Arbeit von Sabater, Arcos, López de Mántaras (1998)¹⁰ genannt, die wie „SaxEx“ unter anderem auf Case-Based-Reasoning (CBR) basiert. Diese Herangehensweise kommt aus der Beobachtung, dass im Allgemeinen die Regeln nicht die Musik machen, sondern die Musik die Regeln (dies wird durch diese Herangehensweise besonders gut abgebildet, da die Regeln aus Beispielen (Fällen) entnommen werden).

Improvisierende Systeme

Eine der ersten Entwicklungen war das „Flavors Band“ System von Frey (1984)¹¹. „Flavors Band“ ist eine prozedurale Sprache um Jazz und populäre Musikstile zu spezifizieren. Das System kann dann aus gegebenen Partituren nach Wahl des gewünschten Stils eine neue Partitur generieren. Dabei kommen Zufallsfunktionen zur Anwendung, die durch die spezifizierten musikalischen Einschränkungen „in die richtige Bahn“ gelenkt wurden. Das bedeutendste Resultat war ein interessantes Arrangement der Bass-Linie und einem improvisierten Solo von John Coltranes Komposition „Giant Steps“.

Einen anderen Ansatz verfolgte Biles 1994¹² mit seinem System „GenJam“. Es generiert ein Modell eines Jazz-Musikers, das durch einen genetischen Algorithmus zu improvisieren lernt. Dabei wird der Lernerfolg des Systems von einem menschlichen Zuhörer bewertet. Die Arbeit von Papadopoulos und Wiggins (1998)¹³ geht einen ähnlichen Weg, allerdings wird die Qualität der Improvisation nicht von einem Menschen beurteilt sondern durch das System selbst. Das System bewertet die Improvisation nach acht Kriterien wie z.B. Notendauer oder Intervalle zwischen Noten. Allen bisher genannten improvisierenden Systemen fehlt allerdings die Interaktivität, die eigentlich die Improvisation von Jazz so interessant macht. Gerade das spontane Kreieren

von Melodien macht den eigentlichen Charme des Improvisierens aus.

Eine Lösung dieses Kritikpunktes ist das Projekt „Band-out-of-a-box“ (BoB) von Thom (2001)¹⁴. Es ermöglicht die Echtzeit-Interaktion zwischen BoB und einem menschlichen Spieler. Man kann deshalb BoB als eine Art „Musik-Begleiter“ bezeichnen, mit dem man in Echtzeit improvisieren kann. Dabei passt sich das System an den menschlichen Musiker an und lernt dabei die Beschränkungen, die der Musiker in seiner Improvisation beachtet. BoB wurde erfolgreich in zwei unterschiedlichen Stilen getestet. Einerseits in dem des Saxophonisten Charlie Parker und andererseits in dem des Geigers Stephane Grapelli.

Vortragende Systeme

Viele der bisher genannten Systeme basieren auf MIDI und betrachten den Aspekt „Ausdruck“ überhaupt nicht. Dabei sind die mit dem Hören befassten Neuronen des Menschen gerade dann besonders angeregt, wenn sich ein Zustand, also die Musik, verändert. Diese Veränderung ist aber bei der synthetisierten Musik nicht gegeben. Das könnte auch erklären, weshalb „menschlich“ gespielte Musik viel interessanter für uns Menschen ist als synthetisierte. Doch die vortragenden Systeme versuchen, genau dieses Manko zu beheben.

Erst in letzter Zeit haben sich die Forscher mit vortragenden Systemen befasst, doch sind inzwischen schon einige interessante Projekte entstanden. Eines ist „SaxEx“, über das später mehr berichtet wird. „SaxEx“ gehört dabei nicht zu den allerersten Systemen dieser Art. 1992 versuchte sich Johnson¹⁵ in einem System, das musikalische High-Level Transformationen regelbasiert durchführt. Ihr Expertensystem war in der Lage, das Tempo und die Artikulation herauszufinden, die beim Spielen von Bachs Fugen aus „Das wohltemperierte Klavier“ verwendet werden sollten. Die Regeln erlangte das System dabei von zwei Profimusikern. Die Ausgabe des Systems ist dabei keine Audio-Datei sondern vielmehr eine mit Spielanweisungen versehene Partitur für menschliche Musiker. Es hat sich gezeigt, dass die vom System generierten Spielanweisungen sehr ähnlich zu denen sind, die in einer gut bekannten kommentierten Ausgabe von „Das wohltemperierte Klavier“ niedergelegt sind. Doch das System ist in seiner Anwendung sehr beschränkt. Das Hauptproblem ist, dass es nur

für Fugen im 4/4-Takt geeignet ist. Für andere Taktarten wären andere Regeln von Nöten. Eine weitere Einschränkung ist die, dass nur Fugen von J.S.Bach verarbeitet werden können.

Eines der bekanntesten Langzeitprojekte ist das „Director Musices“-System der KTH-Gruppe aus Stockholm (Bresin 2001; Friberg 1995; Friberg, Sunberg und Fryden 2000; Friberg 1998). Momentan arbeitet das System mit Regeln für Tempo-, Dynamik- und Artikulations-Veränderungen auf MIDI beschränkt. Die Regeln werden dabei einerseits aus theoretischem Musikwissen und andererseits experimentell aus Trainingsdurchläufen gewonnen – insbesondere durch die sog. Analyse-durch-Synthese-Methode. Die Regeln werden dabei in drei Hauptklassen eingeteilt: Differenzierungs-, Gruppierungs- und Ensembleregeln. Ein anderes Langzeitprojekt ist das von Widmer (2001, 1996)¹⁶. Das System ist ebenfalls Beispiel für die Verwendung von Regeln um beim Vortrag Tempo und Dynamik zu verändern. Der Ansatz des Projekts war, möglichst viel hochqualitative Musikvorführungen in das System zu geben und vorzuverarbeiten. Das Musikmaterial umfasste mehr als 400.000 Noten. Aus diesem großen Archiv werden Regeln abgeleitet – das System „lernt“. Das Projekt hat schon sehr erfolgsversprechende Resultate geliefert.

Die Arbeit von Danneberg und Derenyi (1998)¹⁷ ist ein gutes Beispiel für Artikulations-Veränderungen. Sie entwickelten einen Trompeten-Synthesizer, der ein physikalisches Modell mit einem Vortrags-Modell vereint. Das Ziel des Vortrags-Modells ist es, Steuer-Informationen für das physikalische Modell zu generieren, welche wiederum aus per Hand erstellten Regeln gewonnen werden. Diese Regeln werden aus einer Sammlung kontrolliert aufgenommener menschlichen Vorführungen abgeleitet.

Probleme bei vortragenden Systemen

Mit Ausnahme der Arbeit der KTH-Gruppe, welche drei Ausdrucks-Ressourcen verwendet, alle Projekte auf maximal zwei Ausdrucks-Ausprägungen beschränkt sind – seien es nun Rubato und Dynamik oder Rubato und Artikulation. Diese Einschränkung ist auf die Verwendung von Regeln zurückzuführen. Denn in der Tat ist das Problem der regelbasierten Herangehensweise, Regeln zu finden, die allge-

mein genug sind, alle Ausprägungen verschiedener Vorträge abzudecken – ja sogar die Ausprägungen innerhalb des selben Vortrages des selben Musikers im selben Stück. Die Komplexität wird auch dadurch erhöht, dass die Ausdrucks-Ressourcen nicht unabhängig voneinander eingesetzt werden, sondern voneinander abhängen. Daraus kann man schließen, dass mit jeder weiteren Ressource, die man modelliert, die Schwierigkeit steigt, passende Regeln zu finden.¹⁸

2. SaxEx

Während die bisher vorgestellten Systeme das Wissen um einen ausdrucksstarken Vortrag in explizite Regeln zu verpacken versucht haben, geht SaxEx anders an das Problem heran. Es verwendet direkt das Interpretations-Wissen, das implizit in aufgenommenen Beispielen monophoner menschlicher Vorträge vorhanden ist. Diese Herangehensweise ermöglicht es, mit den fünf bedeutendsten Ausdrucks-Ressourcen zu arbeiten: Dynamik, Rubato, Vibrato, Artikulation und Anstoß der Noten. Des Weiteren basiert das System nicht auf ausdruckslosem/ausdrucksarmem MIDI sondern auf einem ausdrucksreichem Instrument: Dem Tenor-Saxophon. Neu ist auch die Herangehensweise mittels Case-Based Reasoning sowie der Versuch, ein System zu entwickeln, das von einer ausdruckslosen Eingabe (nicht im MIDI-Format) bis zu einer ausdrucksstarken Audio-Ausgabedatei alle Schritte in sich vereint. Der Verzicht auf MIDI-Eingabe stellt dabei eine weitere Schwierigkeit dar, da aus der Eingabe erst einmal die High-Level-Parameter der Musik extrahiert werden müssen, die für die Ausdrucksstärke relevant sind – zu beachten ist jedoch, dass die Noten (insbesondere die notierten Notenwerte) durchaus per MIDI in das System gegeben werden. Durch diese Vorgehensweise bekommt man allerdings eine Ausgabe, die viel näher an einem menschlichen Vortrag liegt, als die Ausgabe anderer Systeme. In den folgenden Abschnitten wird die Vorgehensweise von SaxEx genauer dargestellt. Dabei steht weniger die genaue technische Umsetzung als viel mehr die zugrundeliegenden Konzepte im Vordergrund.

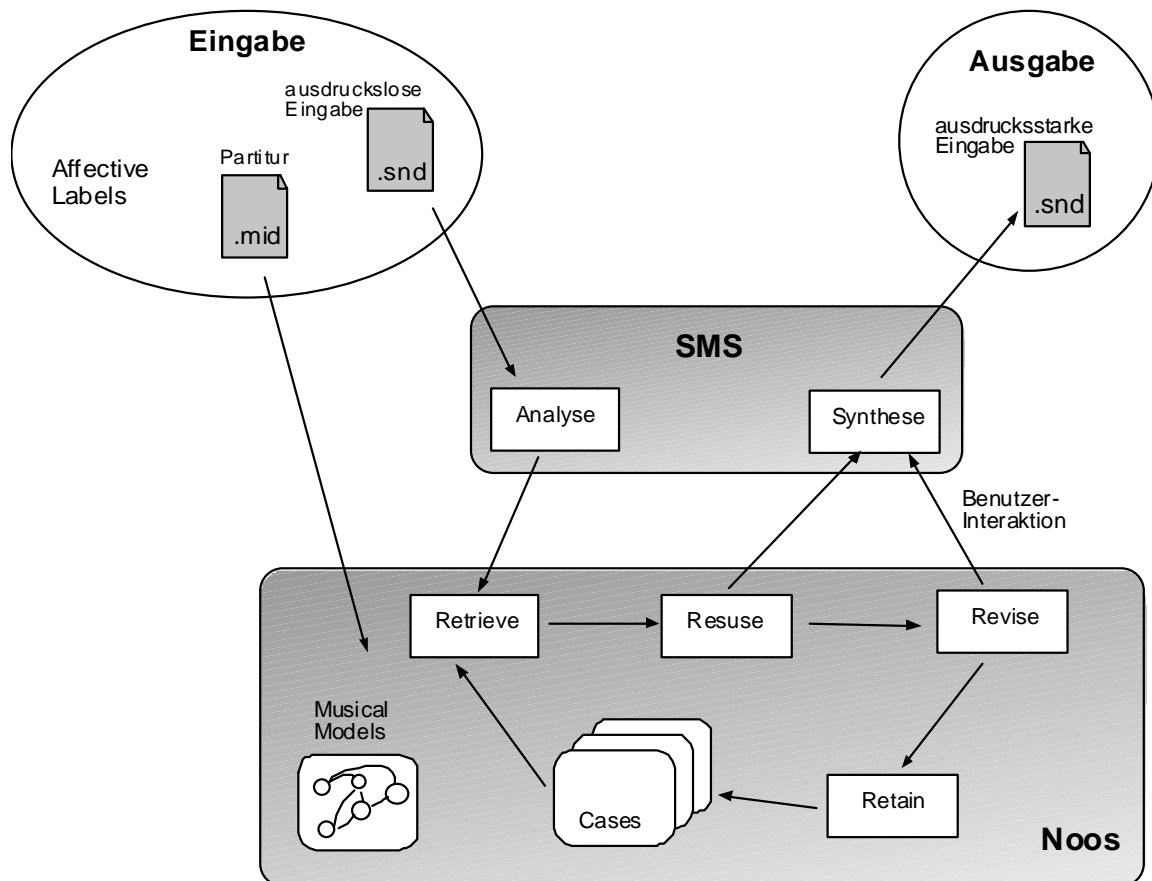


Abbildung 1 - Das SaxEx Schema

2.1. Die Arbeitsweise von SaxEx

Die Aufgabe von SaxEx ist es, eine Reihe von Ausdrucks-Transformationen zusammenzustellen, die auf jede Note aus dem ausdruckslosen Eingabematerial angewendet werden kann, sodass diese ausdrucksstark klingt. Das heißt, SaxEx fügt einer ausdruckslosen Quelldatei Ausdruck hinzu. Um dieses Ziel zu erreichen hat SaxEx einen Fall-Speicher, der Beispiele von ausdrucksstarken menschlichen Vorträgen enthält. Dieses Material wird mittels spektralem Modelliersystem sowie Musikmodellen analysiert. Bei dieser Methode geht es hauptsächlich darum, die „Rolle“ jeder einzelnen Note innerhalb einer Phrase herauszufinden um deren Eigenschaften dann später auf eine Note gleicher/ähnlicher „Rolle“ anzuwenden. Wie in Abbildung 1 dargestellt erfolgt die Verarbeitung in drei Schritten: Analyse (spektrales Modelliersystem), Schlussfolgerung (Case-Based Reasoning), Synthese (spektrales Modelliersystem, SMS).

2.2. Das spektrale Modelliersystem (SMS)

Im spektralen Modelliersystem (SMS) wird die ausdrucksstarke, menschlich erzeugte Musik analysiert. Dabei werden jeder Note Werte für jede der fünf Ausdrucks-Ressourcen zugeordnet. Das heißt, ihr wird zugeordnet, wie der menschliche Musiker die Note gespielt hat. Dabei wird jedoch keine freie Skala sondern vielmehr fünf verschiedene diskrete Stufen verwendet. Sie reichen von „sehr gering“ bis „sehr stark“. So kann Dynamik einer Note als „sehr stark dynamisch“ klassifiziert werden. Das „sehr stark dynamisch“ käme dadurch zu Stande, dass die Note sehr stark von der durchschnittlichen Dynamik in dem Stück nach oben abweicht. Für die anderen vier Ressourcen wird analog verfahren.

Im letzten Schritt ist das spektrale Modelliersystem dann in der Lage, die gewonnenen Parameter auf eine ausdruckslose Eingabedatei anzuwenden, ohne dass dadurch ein qualitativer Verlust entstände.

Kurz gefasst kümmert sich das SMS darum, aus der Musik High-Level Eigenschaften eben dieser zu gewinnen um diese dann später auf andere Musik anwenden zu können.

2.3. Generative theory of tonal music

In diesem Teil-Schritt wird versucht, die „Rolle“ jeder einzelnen Note mit Hilfe der Theorie der tonalen Musik von Lerdahl und Jackendoff (1983), sowie dem kognitiven Musik-Modell von Narmour (1990) zu bestimmen.

Als Ergebnis des ersten Schritts hat man eine Beschreibung von Noten mit ihrer Rolle und den qualitativen Ausprägungen der fünf Ausdrucks-Ressourcen. Diese Informationen liefern wichtiges Kontext-Wissen für die spätere Anwendung.

Um einen ausdrucksstarken Vortrag zu erzeugen wendet SaxEx den Analyse-Schritt auch auf das ausdruckslose Eingangs-Material an. Nachdem die Rolle jeder Note des Eingangs-Materials bestimmt ist, kann SaxEx in seinem Fall-Speicher nach Noten suche, die ähnliche Rollen haben. Wenn SaxEx eine passende Note gefunden hat, wendet es deren Ausdrucks-Eigenschaften auf die bisher ausdruckslose Note an, sodass diese ausdrucksstark klingt. Der Ablauf dieser Suche wird im Abschnitt 2.4 noch genauer ausgeführt.

Der Grund dieser Vorgehensweise mit den Vergleichen mit bereits vorhandenen Informationen liegt darin, dass ähnliche Ausdrucks-Nuancen auf ähnliche musikalische Phrasen angewendet werden sollten. Sind die nötigen Informationen gefunden, gibt SaxEx die Eingabedatei sowie die nötigen Informationen an sein spektrales Modelliersystem weiter, welches dann die Parameter auf die ausdruckslose Eingabe anwendet und am Ende die ausdrucksstarke Ausgabe erstellt.

2.4. Case-Based Reasoning (CBR)

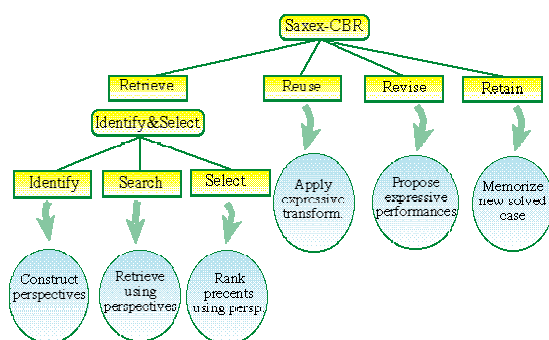


Abbildung 2 - Schema des SaxEx-CBR

Das Case-Based-Reasoning-Subsystem (CBR) ist dafür verantwortlich, zu jeder Note aus der

Eingabe eine korrespondierende Note aus dem Fall-Speicher zu finden. Dieses Subsystem ist es auch, das hauptsächlich auf KI-Techniken basiert. Um eine Vorauswahl treffen zu können ist es möglich, einer Sequenz bzw. einem kompletten Musikstück ein sog. Affective Label (emotionelle Beschriftung) zuzuweisen. Solche Affective Labels drücken aus, wie die Musik interpretiert werden soll (weich – energisch, traurig – lustig, ruhig – ruhelos). Diese Eigenschaften können in verschiedenen Abstufungen zugeordnet werden. In einem späteren Abschnitt wird genauer auf Affective Labels eingegangen.

Abbildung 2 zeigt, in welchem hierarchischen Verhältnis die einzelnen Verarbeitungsschritte im CBR-Subsystem stehen. In den nächsten vier Abschnitten wird auf die folgenden Schritte näher eingegangen: Retrieval-Schritt (Aufspürschritt), Reuse-Schritt (Wiederverwendungsschritt), Revise-Schritt (Überprüfungsschritt) und schließlich dem Retain-Schritt (Lernschritt).

2.4.1. Retrieval-Schritt

Das Hauptproblem im Retrieval-Schritt ist es, eine passende Note für die Imitation zu finden. Anders ausgedrückt ist es das Problem, herauszufinden, dass zwei Noten aus zwei unterschiedlichen Musikstücken ähnliche Rollen in ihrer jeweiligen Phrase haben. Musiker verwenden beim Vergleichen ihr musikalisches Wissen um das Problem zu lösen. Primär werden sie wahrscheinlich auf folgende fünf Dinge achten: Die metrische Stärke der Noten, die harmonische Stabilität, die Notendauer, die hierarchische Relation der Noten und schließlich die Position der Noten.

Dabei bedeutet metrische Stärke, dass Noten, die auf betonten Schlägen im Takt gespielt werden (z.B. in einem $\frac{3}{4}$ -Takt das erste Viertel), eine höhere metrische Stärke haben als die, die auf den unbetonten Schlägen gespielt werden.

Die harmonische Stabilität der Noten ist in Bezug auf die Jazz-Harmonie zu sehen. Dabei ist die Position im Vergleich zum zu Grunde liegenden Akkord zu sehen sowie ihre Rolle im weiteren Verlauf des Akkords.

Die hierarchischen Relation ist im Sinne der „Generative theory of tonal music“ zu sehen. Dabei werden Noten zum Beispiel in die Kategorien „wichtig“ und „verzierend“ eingeordnet. Auch ihre Funktion im Stück wird be-

stimmt: Manche Noten erzeugen Spannung, andere lösen sie etc.

Als letztes Kriterium wird die Position der Noten im Kontext von Namours „Basic musical grouping structures“ herangezogen. Dieses beinhaltet Dinge wie z.B. Wiederholungen oder melodische Entwicklung.

SaxEx speichert all diese Informationen in seinem Fall-Speicher. Dazu wird die objekt-orientierte Programmiersprache namens „Noos“ verwendet, die 1997 von Arcos entworfen wurde. In dem Retrieval-Schritt, der hier gerade beschrieben wird, wird nun dieser Fall-Speicher durchsucht. Gehen wir einmal davon aus, dass in dem ausdruckslosen Eingangsmaterial eine Note gefunden wird, die die letzte Note einer melodischen Entwicklung darstellt, auf einem betonten Schlag gespielt wird und harmonisch stabil ist. SaxEx sucht dann in seinem Fall-Speicher nach Noten mit ähnlichen Eigenschaften. Im einfachsten Fall wird nur eine passende Note gefunden, deren Eigenschaften in dem einfachen Szenario auf die ausdruckslose Note angewendet wird.

Manche mögen jetzt einwenden, dass nur zwei der vorher fünf genannten Kriterien beachtet werden. Aber wie erwähnt, sollte dies ja einen einfachen Fall illustrieren. Sollte die Suche nur mit diesen zwei Kriterien mehrere mögliche Noten gefunden haben, dann werden die anderen Kriterien ebenfalls benötigt um die Zahl der in Frage kommenden Noten weiter einzuzugrenzen. Sollten mehrere Möglichkeiten bestehen bleiben, kann verschieden verfahren werden, doch dazu in einem der nächsten Schritte mehr.

Um die Vorgehensweise noch zu verdeutlichen, hier ein Beispiel in musikalischer Notation. Nehmen wir an, dass wir zu der eingekreisten Note eine passende Note in unserem Fall-Speicher finden wollen.



Zuerst muss herausgefunden werden, welche Rolle die Note hat. Als Ergebnis könnte ungefähr folgende Aussage gemacht werden: „Die Note ist die erste einer P-Phrasierung.“



Nun müsste nach passenden Noten im Fall-Speicher gesucht werden. Gehen wir davon aus, dass folgende zwei Noten (P1 und P2) gefunden werden.



Jetzt muss eine der beiden Noten ausgewählt werden. Nimmt man nun die melodische „Richtung“ als Auswahlkriterium, so wird P1 gewählt (sie gehört zu einer absteigenden Phrase, genauso wie die Note, zu der wir eine ähnliche suchen). Da nur eine Note als Ergebnis übrig geblieben ist, werden deren Eigenschaften wie im folgenden Abschnitt beschrieben wiederverwendet.

2.4.2. Der Reuse-Schritt

In dem Reuse- bzw. Wiederverwendungsschritt wird – wie der Name schon sagt – etwas wieder verwendet. In diesem Fall werden die Eigenschaften wiederverwendet, die die gefundenen Noten haben und auf die Eingangsnoten angewendet. Im einfachsten Fall (der oben geschildert wurde), wenn nur eine passende Note gefunden wurde, werden deren Eigenschaften direkt auf die Eingangs-Note angewendet. Sind aber aus dem letzten Schritt mehrere in gleicher Weise passende Noten gefunden worden, dann wird ein weiterer Mechanismus in Gang gesetzt. Ein Fuzzy-Anhäufungsoperator, der ähnlich wie der gewichtete Mittelwert funktioniert (Arcos, López de Mántaras, 2002) wird angewendet um Eigenschaften für die Note auszurechnen, die die Eigenschaften aller gefundenen passenden Noten vereint. Alternativ hat der Benutzer mehrere Möglichkeiten, diesen Schritt im Voraus zu konfigurieren. Konfigurieren heißt in diesem Falle, den Anhäufungsoperator auszu-schalten und eine der folgenden Alternativen zu wählen:

Der Benutzer kann eine Mehrheitsregel wählen. Dabei wählt SaxEx für jede der vorhandenen Ausdrucks-Ressourcen für die zweifelhafteste Note den Wert aus, den die Mehrheit der gefundenen ausdrucksreichen Noten haben. Als zweite Alternative kann der Benutzer auch die strikte Mehrheitsregel wählen. Bei dieser Konfiguration wird der Wert verwendet, der auf mindestens 50% der gewählten ausdrucksreichen Noten angewendet wurde.

Als weitere Möglichkeit kann er die Minderheitenregel wählen. Ihre Funktionsweise ist der ersten Alternative genau entgegengesetzt. Die strikte Minderheitenregel dagegen wendet den Wert an, der maximal auf eine der gefundenen ausdrucksstarken Noten angewendet wurde.

An fünfter Stelle steht die Möglichkeit, nach Kontinuität zu entscheiden. Dabei werden auf die aktuelle Note die Werte angewendet, die auf die vorhergehende Note in der selben musikalischen Phrase angewendet wurde.

Als letzte Möglichkeit kann noch die Diskontinuität gewählt werden, die das Gegenteil der fünften Möglichkeit wählt.

Sollte nach Anwendung einer der Möglichkeiten immer noch mehr als eine mögliche Wertekombination zur Auswahl stehen, findet eine zufällige Auswahl statt.

Aus den geschilderten Alternativen sollte hervorgehen, dass die Standard-Strategie von SaxEx ist, mit dem Fuzzy-Anhäufungsoperator zu arbeiten. Da aber die musikalische Interpretation ein kreativer Prozess ist, sind diese Alternativen für den Benutzer vorhanden, der dadurch das Ergebnis nach seinen Vorstellungen beeinflussen kann. Besonders die Kriterien für die Minderheiten sollten dies deutlich machen, denn sie generieren eine ausdrucksstarke Ausgabe, die die am wenigsten gebräuchliche Kombination von Ausdrucks-Möglichkeiten verwendet.

2.4.3. Der Revise-Schritt

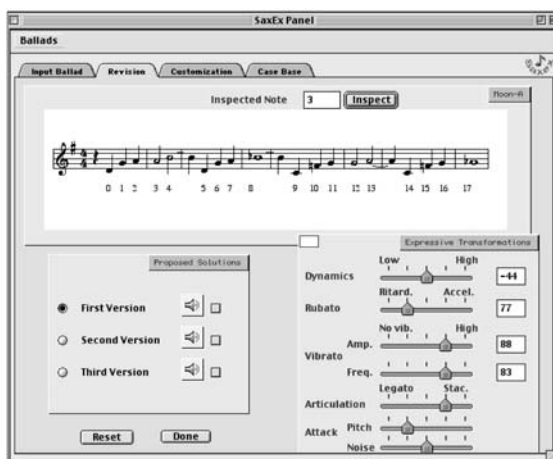


Abbildung 3 - Das interaktive Überprüfungs-Fenster

Im Revise- bzw. Überprüfungs-Schritt kann sich der Benutzer die erlangten Ergebnisse anhören. Durch das interaktive Überprüfungs-Fenster (siehe Abbildung 3) kann er für jede

Note getrennt die Eigenschaften von Hand beeinflussen (genauer gesagt: festlegen), falls er der Meinung ist, die Lösung von SaxEx sollte korrigiert werden. Sollten mehrere Versionen erstellt worden sein, kann er diese kombinieren oder die auswählen, die ihm am besten gefällt. So könnte er je Phrase eine andere Lösung wählen: Für die erste Phrase die zweite Lösung, für die zweite Phrase die dritte Lösung...

2.4.4. Der Retain-Schritt

Der Vortrag, der aus dem Revise-Schritt letztendlich hervorgeht, wird automatisch in den Fall-Speicher von SaxEx aufgenommen und gespeichert. Zu beachten ist auch, dass nur positives Feedback an das System zurückgegeben wird. Die im vorhergegangenen Schritt verworfenen Vorträge werden nicht zurück ans System gegeben. Dies ist der Schritt, in dem das System lernt. Alles, was es gelernt hat, steht dann in jedem neuen Durchlauf zur Verfügung.

2.5. Affective Labels Emotionale Beschriftungen

Bei den sogenannten Affective Labels handelt es sich um vom Benutzer festgelegte Markierungen, in welcher Stimmung eine musikalische Phrase interpretiert werden soll. Affective Labels sind auch den Fällen im Fall-Speicher zugeordnet, sodass diese bei der Auswahl passender Noten berücksichtigt werden können. In sogenannten emotionalen Regionen werden dann mehrere Noten zusammengefasst, die gemeinsame emotionale Eigenschaften haben. Diese Eigenschaften können sich in folgenden drei Achsen bewegen: (1) Weich – energisch, (2) traurig – lustig, (3) ruhig – ruhelos. Diese Achsen sind ähnlich wie die Ausdrucks-Eigenschaften in fünf Abschnitte gegliedert. Diese Stufen sind verbal beschrieben und passen sprachlich zwischen die oben genannten Endpunkte der Achsen.

Legt der Benutzer fest, dass ein Bereich ruhig und sehr weich sein soll, dann sucht SaxEx nach passenden Noten. Findet es keine, so sucht es nach Noten, die zu ruhig und weich passen (SaxEx sucht dann immer in der Abstufung, die der gewünschten am Nächsten ist).

2.6. Experimente und Einschränkungen

Die Macher von SaxEx haben sich bei dem Entwurf des Systems auf ein Instrument sowie einstimmige Musik eingeschränkt. Des Weiteren wurde nur ein Musikstil und daraus nur ein verschwindend geringer Teil des Repertoires bei den Experimenten verwendet. SaxEx beschränkt sich auf Tenor-Saxophon sowie Standard Jazz-Balladen (z.B. „All of Me“, „Autumn Leaves“ und „Misty“).

Es wurden zwei verschiedene Arten von Experimenten durchgeführt, die jeweils auf den oben genannten Balladen basierten. Bei der ersten Versuchsreihe wurde SaxEx die erste Phrase der jeweiligen Balladen in den Fall-Speicher gegeben (also ausdrucksstark vorgespielt) und SaxEx musste dann die restlichen Phrasen der Balladen, die es ausdruckslos erhielt, in eine ausdrucksstarke Version bringen. Diese Versuche haben gezeigt, dass SaxEx in der Lage ist, die relevanten Informationen zu extrahieren und passend anzuwenden. Die entstandenen ausdrucksreichen Ausgaben klangen schlüssig.

In der zweiten Versuchsreihe musste SaxEx aus kompletten ausdrucksreichen Stücken Informationen extrahieren, mit Hilfe derer es dann andere Stücke mit Ausdruck versehen sollte. SaxEx gelang es, auch diese Aufgabe erfolgreich zu meistern.

So ist zu bemerken, dass SaxEx passend zu aufsteigenden Melodien den Phrasen auch Crescendi zuordnen kann und auch lange Noten in traurigen Stücken mit Hilfe von niederfrequentem Vibrato den passenden Ausdruck verleihen kann. Insbesondere ist dabei anzumerken, dass die selben musikalischen Stilmittel auch von menschlichen Spielern häufig an selber Stelle angewendet werden. Nach Aussage der Autoren wurden die von ihnen im Internet bereitgestellten Beispiele von vielen Menschen (Musikern und Nicht-Musikern) als menschlich, d.h. im Stil eines menschlichen Saxophonisten, eingeschätzt.

Nach Ansicht der Autoren von SaxEx haben sie mit ihrem System gezeigt, dass Case-Based-Reasoning eine sehr mächtige Technik darstellt, die direkt das implizit in den Beispielen vorhandene Wissen nutzen kann. Des Weiteren sind sie der festen Überzeugung, dass die Beschränkung auf Jazz-Balladen keine bedeutende Einschränkung darstellt, da es ohne größeren Aufwand möglich sei, SaxEx auch an andere Musikstile zu adaptieren. Man müsse

SaxEx eben anderen Beispiele liefern, aus denen es seinen Fall-Speicher aufbauen kann. Sie sind der Meinung, dass gerade langsame Jazz-Balladen viel von vortragenden Systemen verlangen, da falsche Verwendung von Ausdrucksmitteln eher ins Gewicht fallen als in anderen Stilen. Insbesondere die Richtung Jazz stelle dabei eine höhere Schwierigkeit dar, da Jazz meistens mit komplexeren Strukturen arbeitet als z.B. populäre Musik. Sie sind jedoch sicher, dass es äußerst schwierig wäre, Klassik mit Ausdruck zu versehen, denn dafür würden die passenden Beispiele nicht ausreichen. Man müsse dafür auch das Basiswissen der Harmonisierung des Systems austauschen bzw. anpassen.

Die Einschränkung auf Tenor-Saxophon ist nach der Meinung der Autoren auch nicht gravierend, da es nicht sonderlich schwer sei, andere Instrumente zu verwenden. Insbesondere spreche die Tatsache für das Saxophon, dass es sehr ausdrucksstark und damit auch schwieriger als andere Instrumente zu imitieren sei. Ein bisher ungelöstes Problem stellt die Polyphonie dar. Da das System mit „echter“ Musik arbeitet, können einzelne Stimmen nicht problemlos einzelnen Instrumenten zugeordnet werden. Es ist auch kaum möglich den Klang der einzelnen Instrumente klar zu trennen. Der einzig gangbare Weg wäre eine Mehrspuraufnahme, in der jedes Instrument bzw. jede Stimme ihre eigene Spur besitzt.

Das Ziel für die nähere Zukunft ist es, neben kleineren Verbesserungen, eine Echtzeit-Version des Systems zu erstellen. Des Weiteren soll SaxEx in die Lage versetzt werden, auch zu improvisieren, denn bis jetzt kann es nur die vorhandenen Noten in ihrem Ausdruck verändern, aber noch keine Noten weglassen oder hinzufügen, denn nach Meinung der Autoren muss ein menschenähnlicher Vortrag zumindest in gewissen Grenzen improvisierend sein.¹⁹

3. Abschließende Bewertung

Wie sich gezeigt hat, haben sich Wissenschaftler schon früh Gedanken gemacht, was Computer in der Musik bewirken bzw. erschaffen können. Doch wie sich meiner Meinung nach zeigt, ist die Musik noch um einiges zu kompliziert für Computer. Zwar lassen komponierende Systeme schon erahnen, dass viel mög-

lich sein wird, doch ist die Qualität der Ergebnisse noch nicht mit den Ergebnissen guter Komponisten vergleichbar – einmal ganz abgesehen von herausragenden Komponisten. Zu beachten sind auch immer die Beschränkungen, denen die Systeme unterliegen. Wenn sie denn menschenähnlich komponieren, hält sich die Zahl der eingesetzten Stimmen/Instrumente meist in Grenzen – und die Systeme sind dann auch nicht in der Lage, ein breites Spektrum an Komposition zu erlauben, sondern sie sind auf spezielle Arten zugeschnitten.

Interessant sind auch improvisierende Systeme. Obwohl noch nicht so lange an ihnen geforscht wird, zeigen Sie doch zum Teil erstaunliche Leistungen – wobei wieder gilt: Im Vergleich zu einem guten Musiker sind die Möglichkeiten eher bescheiden. Doch da die Systeme ja häufig selbstlernend sind, könnte Übung nicht schaden, denn es ist selten ein (menschlicher) Meister vom Himmel gefallen. Gerade darin sehe ich auch eines der Probleme der Bewertung solcher Systeme, denn von solchen Computer-Systemen wird oft erwartet, dass sie gute bis hervorragende Ergebnisse liefern – diese Ansprüche scheinen überzogen. Keiner würde von einem Musikschüler, der erst ein Jahr Musikunterricht hatte, so anspruchsvolle Musik erwarten.

Bei den vortragenden Systemen fällt auf, dass sie in ihren Möglichkeiten noch sehr beschränkt sind – beachtet man allerdings, dass an ihnen noch nicht sonderlich lange geforscht wird, sind die Ergebnisse schon beeindruckend. Gerade bei SaxEx fällt auf, dass der Mensch als eine Art letzte Instanz dem Computer doch noch immer sagen muss, ob er mit seiner Interpretation der Musik richtig liegt – und dort liegt auch eines der Probleme, die Computer mit Musik haben: Musik ist emotional und nicht von festen Regeln gesteuert. Es gibt zwar einige Grundregeln, doch gilt hier eher „gute Musik ist, was gefällt, und nicht, was den Regeln entspricht“. Aber gerade an SaxEx sieht man auch den Fortschritt: Während in der ersten Publikation¹⁸ 1997 noch einige Features als wünschenswert bezeichnet werden, sind sie in dem neuesten Artikel¹⁹ schon als realisiert beschrieben.

Um wirklich zu wissen, wie erfolgreich Computer musizieren können, sind meiner Meinung nach Systeme von Nöten, die ähnlich wie die Menschen das Musizieren Stück für Stück über mehrere Jahre lernen. Es ist zwar so, dass der Computer weniger vergesslich ist als Men-

schen dies sind, doch ist das menschliche Gehirn trotzdem nicht so schnell zu schlagen. Gerade in einem Langzeitexperiment würde sich zeigen, zu was die Computer trotzdem – oder gerade deswegen – fähig sind.

-
- ¹ Hiller, L. und Isaacson, L. 1993. Musical Composition with a High-Speed Digital Computer. In *Machine Models of Music*, eds. S. M. Schwanauer und D. A. Levitt, 9-21. Cambridge, Mass.: The MIT Press.
 - ² Moorer, J. A. 1993. Music and Computer Composition. In *Machine Models of Music*, eds. S. M. Schwanauer und D. A. Levitt, 167-186. Cambridge, Mass.: The MIT Press.
 - ³ Rader, G. M. 1993. A Method for Composing Simple Traditional Music by Computer. In *Machine Models of Music*, eds. S. M. Schwanauer und D. A. Levitt, 243-260. Cambridge, Mass.: The MIT Press.
 - ⁴ Ebcioğlu, K. 1993. An Expert System for Harmonizing Four-Part Chorales. In *Machine Models of Music*, eds. S. M. Schwanauer und D. A. Levitt, 385-401. Cambridge, Mass.: The MIT Press.
 - ⁵ Cope, D. 1990. Pattern Matching as an Engine for the Computer Simulation of Musical Style. In *Proceedings of the 1990 International Computer Music Conference*, 288-291. San Francisco, Calif.: International Computer Music Association
 - Cope, D. Experiments in Music Intelligence. In *Proceedings of the 1987 International Computer Music Conference*, 174-181. San Francisco, Calif.: International Computer Music Association.
 - ⁶ Feulner, J. 1993. Neural Networks That Learn an Reproduce Various Styles of Harmonization. In *Proceedings of the 1993 International Computer Music Conference*, 236-239. San Francisco, Calif.: International Computer Music Association.
 - ⁷ Hörnel, D. und Degenhardt, P. 1997. A Neural Organist Improvisation Baroque-Style Melodic Variations. In *Proceedings of the 1997 International Computer Music Conference*, 430-433. San Francisco: International Computer Music Association.
 - ⁸ Hörnel, D. und Menzel, W. 1998. Learning Musical Structures and Styles with Neural Networks. *Journal on New Music Research* 22(4): 44-62.
 - ⁹ Morales-Manzanares, R.; Morales, E. F.; Danenberg, R.; und Berger, J. 2001. Sicib: An Interactive Music Composition System Using Body Movements. *Journal of New Music Research* 25(2): 25-36

-
- ¹⁰ Sabater, J; Arcos, J. L.; und López de Mántaras, R. 1998. Using Rules to Support Case-Based Reasoning for Harmonizing Melodies. Paper presented at the AAAI Spring Symposium on Multimodal Reasoning, 27 July, Stanford, California.
- ¹¹ Fry, C. 1993. Flavors Band: A Language for Specifying Musical Style. In *Machine Models of Music*, eds. S. M. Schwanauer und D. A. Levitt, 427-451. Cambridge, Mass.: The MIT Press.
- ¹² Biles, J. A. 1994. GenJam: A Genetic Algorithm for Generating Jazz Solos. In *Proceedings of the 1994 International Computer Music Conference*, 131-137, San Francisco, Calif.: International Computer Music Association.
- ¹³ Papadopoulos, G., und Wiggins, G. 1998. A Genetic Algorithm for the Generation of Jazz Melodies. Paper presented at the Finnish Conference on Artificial Intelligence (SteP'98), 7-9 September, Jyväskylä, Finland.
- ¹⁴ Thom, B. 2001. BoB: An Improvisational Music Companion. Ph.D. dissertation, School of Computer Science, Carnegie-Melon University.
- ¹⁵ Johnson, M. L. 1992. An Expert System for the Articulation of Bach Fugue Melodies. In *Readings in Computer-Generated Music*, ed. D. L. Baggi, 41-51. Washington, D.C.: IEEE Computer Society.
- ¹⁶ Widmer, G. 1996. Learning Expressing Performance: The Structure-Level Approach. *Journal of New Music Research* 25(2). 179-205.
- Widmer, G. 2001. The Musical Expression Project: A Challenge for Machine Learning and Knowledge Discovery. In *Proceedings of the Twelfth European Conference on Machine Learning*, 603-614. *Lecture Notes in Artificial Intelligence* 2167. New York: Springer-Verlag.
- ¹⁷ Danneberg, R. B., und Derenyi, I. 1998. Combining Instrument and Performance Models for High-Quality Music Synthesis. *Journal of New Music Research* 27(3): 211-238.
- ¹⁸ Lopez, Ramon und Arcos, Josep Lluís, 2002. AI and Music – From Composition to Expressive Performance. *AI Magazine* Fall 2002, 43-57
- ¹⁹ Arcos, Josep Lluís; López de Mántaras, Ramon; Serra, Xavier, 1997. SaxEx: A case-based reasoning system for generating expressive musical performances.